

# Deep Learning in Dental Image Analysis: A Systematic Review of Datasets, Methodologies, and Emerging Challenges

Zhenhuan Zhou<sup>a,b</sup>, Jingbo Zhu<sup>c</sup>, Yuchen Zhang<sup>e</sup>, Xiaohang Guan<sup>f</sup>, Peng Wang<sup>a,b</sup>, Tao Li<sup>a,b,d,\*</sup>

<sup>a</sup>College of Computer Science, Nankai University, Tianjin, 300350, China

<sup>b</sup>Key Laboratory of Data and Intelligent System Security, Ministry of Education, China

<sup>c</sup>College of Software, Nankai University, Tianjin, China

<sup>d</sup>Department of stomatology, Tianjin Union Medical Center (The First Affiliated Hospital of Nankai University), Tianjin, China

<sup>e</sup>Tianjin Stomatological Hospital, Tianjin, China

<sup>f</sup>Haihe Lab of ITAI, Tianjin, 300459, China

---

## Abstract

Efficient analysis and processing of dental images are crucial for dentists to achieve accurate diagnosis and optimal treatment planning. However, dental imaging inherently poses several challenges, such as low contrast, metallic artifacts, and variations in projection angles. Combined with the subjectivity arising from differences in clinicians' expertise, manual interpretation often proves time-consuming and prone to inconsistency. Artificial intelligence (AI)-based automated dental image analysis (DIA) offers a promising solution to these issues and has become an integral part of computer-aided dental diagnosis and treatment. Among various AI technologies, deep learning (DL) stands out as the most widely applied and influential approach due to its superior feature extraction and representation capabilities. To comprehensively summarize recent progress in this field, we focus on the two fundamental aspects of DL research—datasets and models. In this paper, we systematically review 260 studies on DL applications in DIA, including 49 papers on publicly available dental datasets and 211 papers on DL-based algorithms. We first introduce the basic concepts of dental imaging and summarize the characteristics and acquisition methods of existing datasets. Then, we present the foundational techniques of DL and categorize relevant models and algorithms according to different DIA tasks, analyzing their network architectures, optimization strategies, training methods, and performance. Furthermore, we summarize commonly used training and evaluation metrics in the DIA domain. Finally, we discuss the current challenges of existing research and outline potential future directions. We hope that this work provides a valuable and systematic reference for researchers in this field. All supplementary materials and detailed comparison tables will be made publicly available on GitHub.

**Keywords:** Deep Learning, Dental Image Analysis, Dental Datasets, Systematic Review

---

## 1. Introduction

With the continuous improvement of global living standards and the steady advancement of medical science, oral health has increasingly become an important indicator of human well-being and social development, as well as an essential component of overall health for all humankind (Listl and Bärnighausen (2025)). However, over the past three decades, oral diseases have remained among the most prevalent health conditions worldwide, significantly affecting the health and quality of life of billions of people (Ben-zian and Beltrán-Aguilar (2025)). According to statistics from the World Health Organization, as of 2019, more than 3.5 billion people—nearly half of the global population—were affected by oral diseases of varying severity (Organization (2022, 2024)). By 2021, this number had risen to 3.69 billion Bernabe et al. (2025), exceeding the combined total cases of the world's five major non-communicable diseases (mental disorders, cardiovascular

disease, diabetes mellitus, chronic respiratory diseases, and cancers) by over 1 billion (Organization (2022)).

A significant factor contributing to this issue is the acute shortage of oral healthcare professionals on a global scale. Developing a competent junior dentist typically demands over ten years of training, leading to a global average of merely about 3.3 dentists per 10,000 individuals (Gal-lagher et al. (2024)). This shortage is especially severe in low-income and less developed regions (Organization et al. (2025)). With the rapid developments in machine learning and artificial intelligence (AI) technologies, AI-driven computer-aided diagnosis (CAD) systems are increasingly being adopted in the dental field (Cui et al. (2022c); Wang et al. (2025b)). These systems alleviate the scarcity of dental professionals, reduce their workload, and improve diagnostic efficiency and accuracy. By minimizing the subjectivity associated with variations in clinical experience, they provide robust support for precise diagnosis and treatment planning. The implementation of semi-automatic or fully automatic intelligent Dental Image

---

\*Corresponding Author, Email: litao@nankai.edu.cn

Analysis (DIA) represents a crucial step and prerequisite for advancing such CAD systems.

As one of the most widely applied AI technologies today, deep learning (DL) (LeCun et al. (2015)) has demonstrated remarkable potential in various domains of medical image analysis and processing, such as dermoscopy (Mirikharaji et al. (2023); Wu et al. (2025c)), fundus imaging (He et al. (2022); Li et al. (2023a)), pathology (Song et al. (2024a); Shen et al. (2025)), ultrasound (Ibtehaz and Kihara (2023); Lin et al. (2023a)), chest X-rays (Shao et al. (2022); Zhao et al. (2024a)), and abdominal CT scans (Rahman and Marculescu (2025); He et al. (2023)). Unlike other fields of medical imaging, dental imaging possesses unique characteristics that present diverse challenges for DIA. For instance, unfavorable projection angles can lead to ambiguous tooth topology or overlapping teeth in 2D images (Chaitanya et al. (2024)); noise interference and the inherently low contrast between soft tissues (Venkatesh and Elluru (2017)); metallic implants can introduce artifacts that hinder recognition and interpretation (Liu et al. (2023a)); and the presence of deciduous teeth in children may cause feature confusion (Zhou et al. (2024)). Naturally, as the importance of oral health continues to grow, an increasing number of AI researchers have turned their attention to DIA, leading to the emergence of numerous DL-based methods. The left panel of Figure 1 illustrates the annual growth trend of DIA-related publications employing DL from 2016 to September 2025. It can be observed that the number of publications began to increase rapidly from 2022 and reached its most vigorous momentum in 2024. The right panel of Figure 1 shows the geographical distribution of all included studies, with Asia—led by China—accounting for the largest share, followed by the Americas and Europe, while Africa has relatively few studies. This highlights the imbalance in the regional distribution of existing DIA research, which we will elaborate on in Section 6.

In summary, the application of DL in DIA has achieved substantial progress; however, a systematic review and synthesis are essential. From the perspectives of both dental practitioners and computer science researchers, a comprehensive examination of existing public dental imaging datasets and AI-based DL methodologies is crucial for advancing the field of intelligent DIA. Such a review not only aids in understanding how current approaches address inherent challenges and their limitations but also helps identify unresolved issues and anticipate emerging research trends in DIA.

**Relevant surveys.** Several existing surveys cover areas that are partially similar to this paper. For instance, Nambiar et al. (Nambiar and Nanjundegowda (2024)) conducted a concise review of 30 studies, categorizing them by dental specialty and illustrating the impact of DL-based models and their variants on dental practice. Al-fadley et al. (Alfadley et al. (2024)) summarized 13 studies to explore the application of UNet-based DL models in pulp segmentation. Kot et al. (Kot et al. (2025))

systematically reviewed 30 studies, categorizing them according to backbone network types, aiming to assess the evolution and performance of DL in tooth segmentation. Wu et al. (Wu et al. (2025a)) conducted a systematic review of 145 publications on dental CT image segmentation algorithms, classifying them into traditional algorithms and DL-based approaches (approximately 50 studies), and outlined the transition from classical methods to DL techniques. Compared to the aforementioned studies, our work differs in several key aspects: **(1)** Previous studies have largely overlooked a systematic investigation and meta-analysis of publicly available dental datasets. In contrast, this work offers a thorough overview of commonly used dental imaging modalities and accessible public datasets, conducting clustering analysis from multiple perspectives, including data modality, dataset scale, and acquisition methods. **(2)** Most existing research primarily focuses on panoramic (PAN) and cone-beam computed tomography (CBCT) as the main imaging modalities, emphasizing tooth segmentation tasks. Consequently, they fail to comprehensively cover the diverse tasks and modalities involved in DIA. Our review broadens this scope and provides a more holistic analysis. From a task categorization perspective, we refine dense tooth prediction into tooth-level and anatomical-level prediction, extending the review beyond tooth analysis to include lesion and maxillofacial structure prediction, along with dental classification tasks. Moreover, we incorporate previously underrepresented imaging modalities such as Intraoral Scanning (IOS), Intraoral Photography (IP), and periapical radiographs (PR). **(3)** Some prior reviews were conducted mainly from a medical perspective, with limited coverage and little attention to regional disparities. They also lacked in-depth analyses of model architectures and performance from a computer science viewpoint. In comparison, our survey systematically reviews 211 papers on DL-based DIA algorithms and 49 publicly accessible datasets, significantly surpassing previous efforts in both scope and scale. We perform clustering across all literature and datasets, providing multi-perspective analyses from a computer science standpoint. We hope that this work can offer valuable insights and guidance for researchers, fostering continuous progress and innovation within the DIA community.

**Search strategy.** We conducted a systematic search across several widely used databases in computer science and medicine, restricting the time span to 2019–September 2025. The platforms included Google Scholar, DBLP, IEEE Xplore, SpringerLink, Elsevier, and PubMed. For DBLP, the query was formulated as follows: (trans\* | deep | neural | learn\*) (oral | dent\* | tooth\* | teeth | mandib\* | maxill\* | CBCT | "cone beam" | "intraoral scan\*" | "panoramic X-ray") (segment\* | classif\* | detect\*). Similar keyword combinations were applied to the other platforms to ensure comprehensive coverage of studies relevant to this review. For dataset collection, in addition to the above academic databases, we also searched open repositories oriented to-

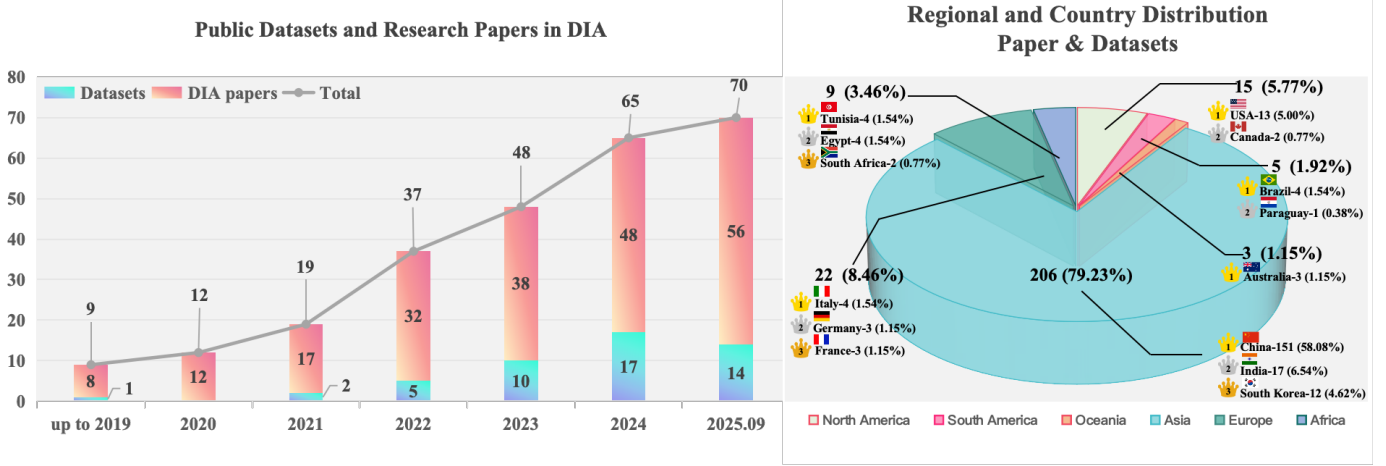


Figure 1: Statistical results from the papers reviewed in this study are visualized in two charts. The bar chart on the left depicts the annual distribution of DL-based, DIA-related papers and public datasets, with the gray line indicating their cumulative total. The pie chart on the right illustrates the regional distribution, highlighting the top three countries with the highest number of studies in each continent.

wards computer science, including HuggingFace, Kaggle, and Mendeley. The initial search yielded approximately 450 papers and 80 datasets. Two computer science researchers then independently performed the first round of screening to remove irrelevant items. The exclusion criteria for research papers were as follows:

- Studies not utilizing DL-based methods for DIA;
- Works focusing on oral cancer, oral epithelium, or cell-level microscopic imaging that fall outside the scope of dental imaging;
- Non-medical studies irrelevant to dental images (e.g., mechanical gear studies);
- Preprints from arXiv and similar platforms were meticulously examined, and low-quality non-peer-reviewed papers were excluded;
- Low-quality studies were excluded during the selection process. Specifically, datasets without public access and 2D datasets containing fewer than 50 cases or 3D datasets with fewer than 10 cases were not included in this review. For research papers, exclusion criteria encompassed studies with ambiguous methodologies, poorly designed experiments, or low reproducibility.

After applying these criteria, we identified a total of 211 research papers and 49 publicly available dental datasets. We then categorized them by publication year and geographic distribution (Fig. 1). The results indicate a significant increase in DIA-related DL research beginning in 2022, with both the number of publications and dataset releases steadily rising and reaching a peak in September 2025.

**Taxonomy.** For clarity of discussion, we categorized the publicly available dental datasets and DL-based DIA algorithms covered in this paper, as illustrated in Fig. 2. On the left side, from the perspective of datasets, all 49 datasets are clustered according to four commonly used dental imaging modalities, which include 2D X-ray imaging, CBCT, IP captured by standard cameras, and IOS. On the right side, from the perspective of models and applications, the 211 reviewed papers are classified into different DIA tasks, covering dense prediction, classification, and other related tasks. Examples of different data modalities and finer-grained literature categorizations are also provided in Fig. 2.

**Contribution.** In summary, the main contributions of this paper can be summarized as follows:

- We conducted a comprehensive and in-depth survey, reviewing 260 deep learning-based DIA studies—approximately an order of magnitude larger in scale compared to previous reviews. We also proposed an efficient taxonomy for all reviewed studies, designed to help readers from different backgrounds establish a clear and systematic understanding of this field.
- We introduced the commonly used devices and imaging modalities in dental imaging, providing a systematic and comprehensive review of publicly available dental datasets, thereby addressing a gap left by previous surveys. This effort not only offers researchers a clear understanding of the current status of available DIA datasets but also provides valuable insights into the goals and directions for future dataset development.
- We categorized all included studies based on their DIA task objectives and model architectures. Ad-

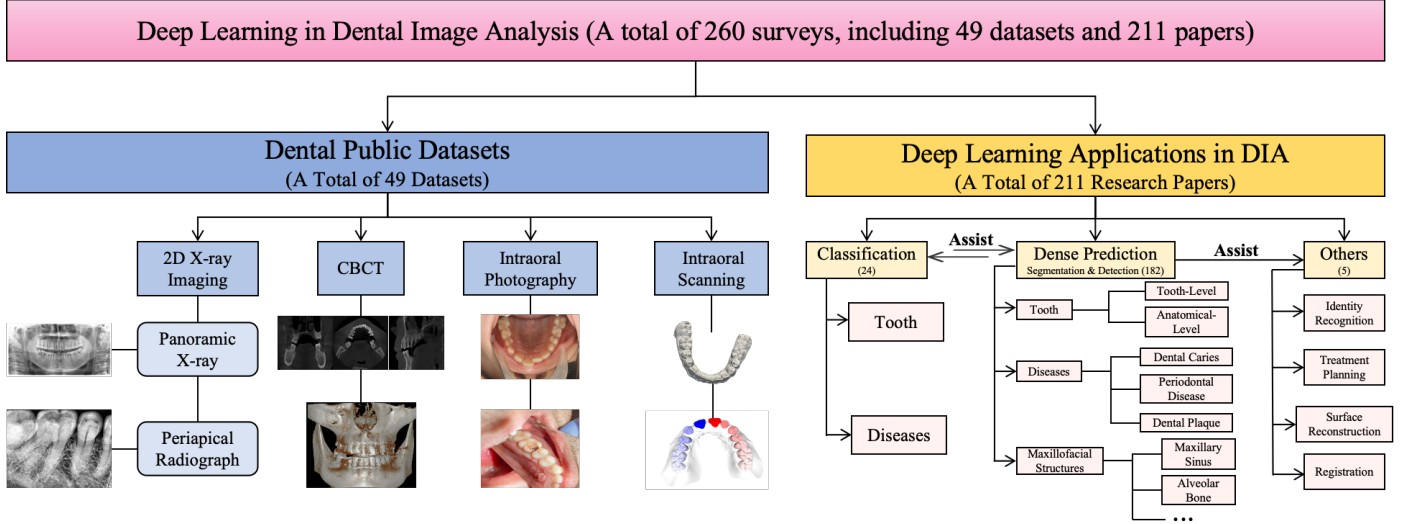


Figure 2: Taxonomy of Deep Learning Applications in Dental Image Analysis. The diagram is divided into two main sections: Dental Public Datasets and Deep Learning Applications in Dental Image Analysis.

ditionally, we provided a comprehensive statistical summary for each study, detailing the datasets used, network architectures, training parameters, hardware configurations, evaluation metrics, and performance outcomes, along with key innovations and highlights of each work.

- Furthermore, we offer a detailed discussion and analysis of the current challenges and limitations of applying DL to DIA from both the dataset and model & application perspectives. We also outline future trends, research prospects, and potential directions for exploration in this field.

**Paper Organization.** The remainder of this paper is organized as follows: Section 2 introduces commonly used dental imaging modalities and the publicly available datasets reviewed in this study. Section 3 discusses fundamental DL architectures and, in a task-oriented manner, examines DIA algorithms and applications, beginning with dense prediction. Section 5 reviews commonly used loss functions, optimizers, GPUs, and performance evaluation metrics utilized in model training and testing. Section 6 delves into the challenges and limitations of existing research, as well as potential directions for future development. Finally, Section 7 concludes the paper.

## 2. Dental Imaging and Public Datasets

The establishment of high-quality public datasets is crucial for advancing research in DL within the field of DIA and constitutes one of the essential prerequisites for developing CAD systems in dentistry. In this section, we first introduce the commonly used dental imaging modalities, followed by a detailed overview of the publicly available datasets surveyed in this paper.

### 2.1. Dental Imaging

As an essential component of medical clinical imaging, dental imaging provides a crucial reference for disease diagnosis and treatment planning in dentistry, with well-established operational standards and clearly defined functional distinctions [Vandenberghe et al. \(2010\)](#). Dental imaging can be broadly categorized into four types: 2D X-ray imaging (such as PR and PAN), CBCT, IOS, and IP. In this section, we offer a brief introduction to their imaging techniques, data formats, and clinical applications.

#### 2.1.1. 2D X-ray imaging

2D X-ray imaging remains an essential and widely utilized modality in dental clinical diagnostics. It functions by transmitting X-rays through dental structures and adjacent tissues to produce images on a detector, which can be either film or a digital sensor [Różyło-Kalinowska \(2021\)](#). Based on the acquisition method and imaging scope, 2D X-ray can be categorized into extraoral imaging (e.g., panoramic radiographs) and intraoral imaging (including periapical, bitewing, and occlusal radiographs). Fig 3 (a) and (b) respectively display the PAN and PR imaging devices photographed in a clinical setting. As illustrated in Fig 2, panoramic radiographs provide a comprehensive view of the entire dentition and maxillofacial structures in one image, making them particularly suitable for orthodontic treatment planning and the extraction of impacted wisdom teeth [Chauhan et al. \(2023\)](#). Conversely, intraoral imaging, exemplified by PR, delivers detailed localized information, typically focusing on one or a few teeth or a specific region of the dentition, and is commonly employed for caries diagnosis and preoperative planning of root canal treatment [Vandenberghe et al. \(2010\)](#); [Stera et al. \(2024\)](#). For analyzing 2D X-ray imaging, images are frequently stored in JPG or PNG formats.

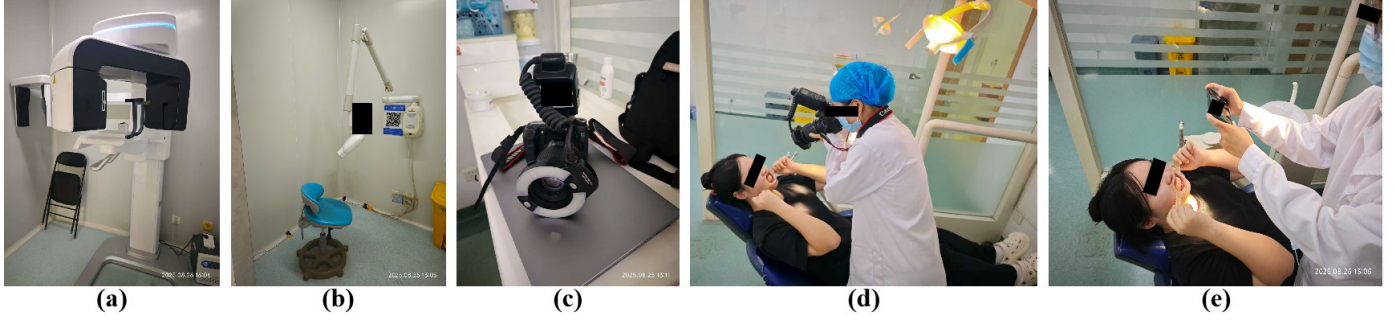


Figure 3: Dental imaging devices utilized in actual clinical practice, featuring images captured at the Department of Stomatology, First Affiliated Hospital of Nankai University. (a) 2D panoramic radiography device; (b) periapical radiography device; (c) DSLR camera for intraoral photography; (d) radiologist taking intraoral photographs using a camera; (e) radiologist taking intraoral photographs using a mobile device.

### 2.1.2. Cone-Beam Computed Tomography (CBCT)

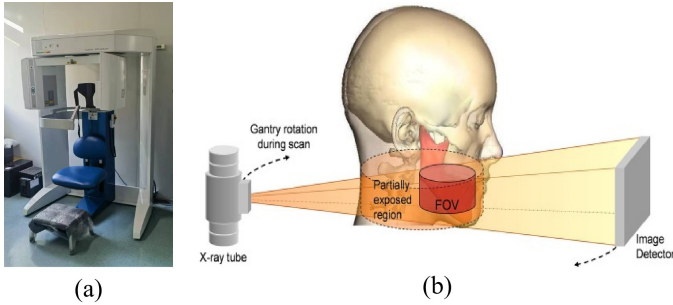


Figure 4: (a) CBCT imaging device, captured at the Department of Stomatology, First Affiliated Hospital of Nankai University; (b) The fundamental principle of CBCT scanning involves a rotating gantry, an X-ray tube, and an image detector [Kaasalainen et al. \(2021\)](#).

While 2D X-ray imaging is extensively utilized in dentistry, it presents certain limitations, such as distortions or tooth overlaps resulting from the projection of three-dimensional anatomical structures onto a two-dimensional plane [Vandenberghe et al. \(2010\)](#). In 1995, Mozzo et al. introduced the world’s first CBCT prototype, the New-Tom DVT 9000 [Mozzo et al. \(1998\)](#), marking the official advent of CBCT imaging technology. As depicted in Fig. 4, with advanced hardware based on imaging physics and continually refined reconstruction algorithms, CBCT facilitates diverse three-dimensional visualizations and high-dimensional reconstructions of volumetric data, effectively addressing many limitations of 2D X-ray imaging [Kaasalainen et al. \(2021\)](#). Consequently, CBCT has seen increased application in dentistry. CBCT images can be analyzed using either 2D slices or 3D volumetric data. The raw CBCT data are usually stored in DICOM format [Mildenberger et al. \(2002\)](#), a leading imaging standard in the medical field. Post-CBCT scan, a folder containing a series of consecutively numbered DICOM files is generated, recording not only imaging data but also basic patient information, such as sex and age. For computer science researchers, converting DICOM files into a more straightforward and efficient 3D data format, NIfTI (.nii), is preferred

[Li et al. \(2016\)](#). NIfTI-formatted data facilitate efficient preprocessing operations such as resampling, slicing, or patch extraction, enabling the creation of datasets tailored to various models and tasks. Open-source tools like ITK-Snap [Yushkevich et al. \(2006\)](#) and 3D Slicer [Fedorov et al. \(2012\)](#) aid in the 3D visualization of NIfTI data and offer rich annotation interfaces to support the construction of deep learning datasets and model training. However, despite the growing sophistication of CBCT technology, it still faces certain limitations, including higher imaging costs, low contrast, and metal artifacts. Therefore, in clinical practice, physicians typically make flexible choices between CBCT and 2D X-ray imaging based on the specific conditions of the patient [Tibúrcio-Machado et al. \(2021\)](#).

### 2.1.3. Intraoral Scanning (IOS)

Unlike the aforementioned X-ray-based radiographic imaging methods, IOS captures a 3D digital model of the teeth and oral tissues directly within the patient’s mouth. This technique is extensively applied in various clinical scenarios, such as orthodontics and dental implantology [Eggmann and Blatz \(2024\)](#). IOS acts as a digital alternative to conventional impressions, providing faster and more accurate data acquisition while significantly enhancing patient comfort during the modeling process [Mangano et al. \(2017\)](#). The acquisition and processing workflow of IOS is illustrated in Fig. 5. Initially, a dentist uses a hand-held scanning device to capture clinical point cloud data of the patient’s oral cavity (Fig. 5. a), which are then stored in point cloud format (Fig. 5. b). Subsequently, a mesh model is generated from the point cloud using triangulation or other reconstruction algorithms, consisting of vertices, edges, and faces (primarily triangular meshes, Fig. 5. c). Finally, the mesh model undergoes further processing and optimization to produce a digital dental model suitable for clinical use (Fig. 5. d).

### 2.1.4. Intraoral Photography (IP)

IP is conducted using imaging equipment, such as a DSLR camera equipped with a macro lens and flash, as illustrated in Fig. 3 (c) and (d), to capture high-resolution

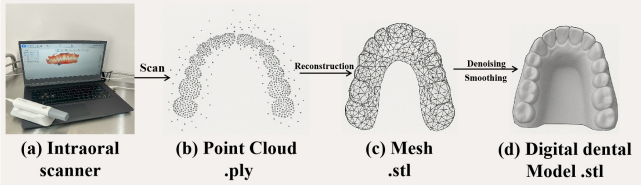


Figure 5: The construction steps of a digital dental model: (a) Intraoral scanner; (b) The intraoral scanner captures raw data, stored in point cloud format (common file extensions such as .ply or .xyz); (c) Mesh model (typically in formats like .obj or .stl); (d) Digital dental model.

2D RGB images of the teeth, gingiva, and oral mucosa within the patient’s mouth. Unlike other imaging modalities, IP offers a realistic representation of both soft and hard tissues, making it essential for oral health assessment [Guo et al. \(2021\)](#). Notably, IP is devoid of radiation exposure risks, while being cost-effective and easily performed. With minimal training, patients can utilize mobile devices to take images at home, as depicted in Fig. 3 (e), rendering it highly suitable for remote dentistry and tele-diagnosis [Mao et al. \(2025\)](#). Furthermore, because it visualizes lesion conditions directly without necessitating extensive radiological expertise, it facilitates smoother communication between patients and clinicians, aiding both parties in reaching a clearer consensus [Kulkarni and Kharad \(2024\)](#). IP does not require specialized equipment, and its imaging process is akin to that of natural images. Consequently, the output is stored in standard image formats, most commonly with extensions such as ‘.jpg’ or ‘.png’.

## 2.2. Dental Public Datasets

High-quality, large-scale public dental datasets are crucial for advancing DL research in DIA. Therefore, we have compiled 49 publicly available dental datasets in Table 1. These datasets originate from various regions and encompass a wide range of imaging modalities and research tasks. For each dataset, we also provide its acquisition method and link, facilitating efficient access and utilization of these resources by readers and researchers for study and research.

Fig 6 illustrates the distribution of dataset sizes and modalities among the datasets evaluated in this study. It can be observed that the majority of datasets consist of fewer than 10,000 samples. Regarding modalities, 2D panoramic images are the most prevalent, followed by CBCT, while other modalities occur with a roughly one order of magnitude lower frequency. 2D radiographic imaging remains the most widely employed modality in dental imaging globally and is also the predominant imaging modality in this survey, accounting for more than 61.22%. For instance, as shown in Fig. 7 (a), Panetta et al. [Panetta et al. \(2021\)](#), Hamamci et al. [Hamamci et al. \(2023\)](#), Long et al. [Loop \(2023\)](#), and Wang et al. [Wang et al. \(2025f\)](#) have proposed clinical PAN datasets of varying sizes and sources for tooth semantic or instance-level segmentation

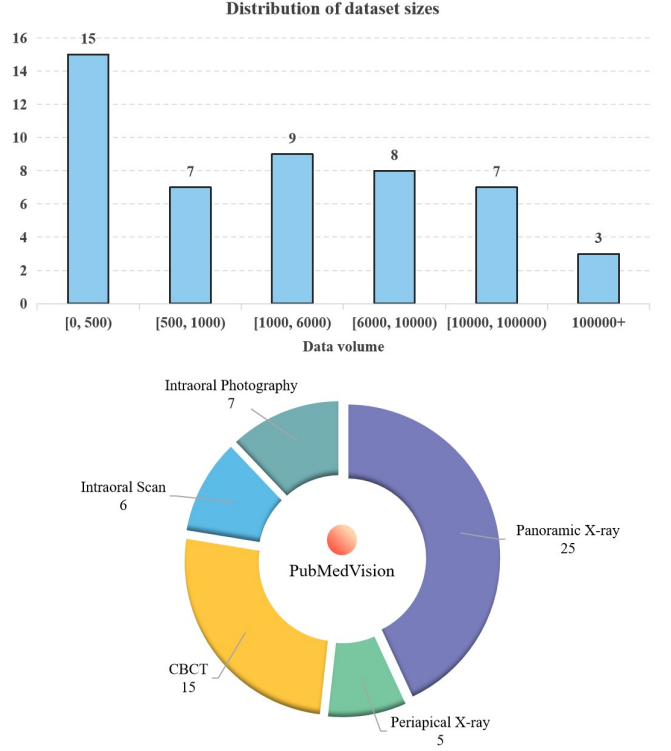


Figure 6: The distribution of dataset information. The bar chart above illustrates the data volume distribution for each dataset, while the pie chart below depicts the distribution of dataset modalities (datasets containing multiple modalities are counted more than once). PubMedVision is a special case, as its image modalities include not only various dental imaging types but also statistical charts; hence, it is separately highlighted in the center of the pie chart.

tasks. Meanwhile, datasets by Li et al. [Li et al. \(2024e\)](#), Seifossadat et al. [Seifossadat \(2023\)](#), Root Disease X-Ray UBU (2024), and OPG Kennedy Orville (2024) focus on lesion classification or anomaly detection. In addition, Dental X-Ray [Rad et al. \(2016\)](#), Zhang et al. [Zhang et al. \(2023d\)](#), DC 1000 [Wang et al. \(2023a\)](#), and Vipularya et al. [Arya \(2024\)](#) have introduced datasets dedicated to caries segmentation or lesion recognition in PAN images. Compared with PAN, publicly available PR datasets are considerably more limited, constituting only about 20% of the volume of PAN data, indicating a potential shortage of datasets. For instance, the Dental X-Ray [Rad et al. \(2016\)](#) offers a small-scale PR dataset for caries detection tasks; Root Disease X-Ray UBU (2024) includes 6,578 PR images for root lesion classification; and PRAD [Zhou et al. \(2025b\)](#) assembles a relatively large dataset containing 10K PR images with multi-structure pixel-level annotations, of which 5K are publicly accessible upon request.

With the rising popularity of CBCT, more researchers have focused on constructing 3D radiographic databases, as shown in Fig. 7 (a). CTooth [Cui et al. \(2022b\)](#) and CTooth+ [Cui et al. \(2022a\)](#) are exemplary CBCT datasets developed for semi-supervised tooth semantic segmenta-

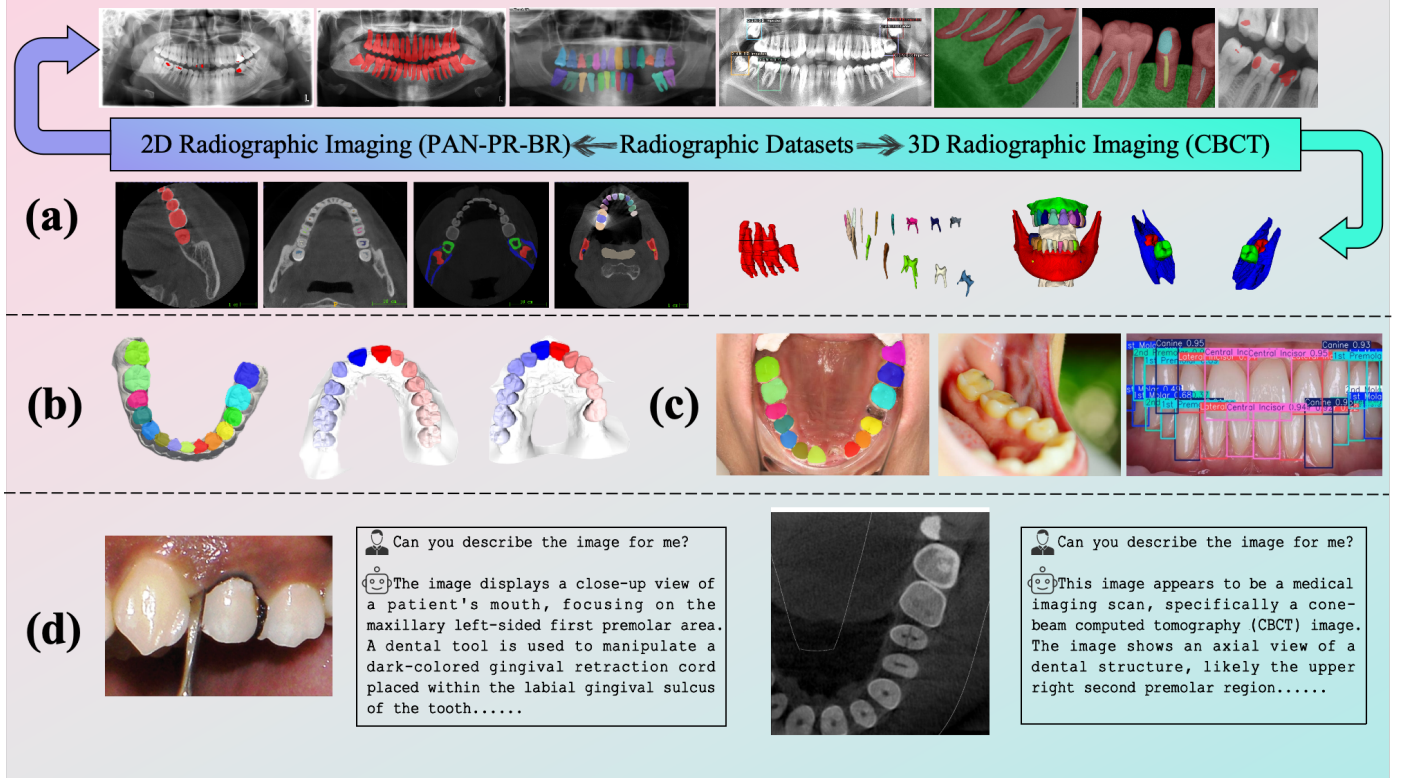


Figure 7: Examples of data modalities and annotations in publicly available dental datasets, featuring sample data drawn from selected surveyed datasets: (a) dental radiographic imaging Wang et al. (2023a); Loop (2023); Panetta et al. (2021); Hamamci et al. (2023); Zhou et al. (2025b); Cui et al. (2022b); Gamal et al. (2024); Zhou et al. (2024); Bolelli et al. (2025), (b) intraoral scanning (IOS) Zou et al. (2024); Ben-Hamadou et al. (2023), (c) intraoral photography Zou et al. (2024); Siddhartha (2024), and (d) dental image-text data represented by PubMedVision Chen et al. (2024a).

tion. Subsequently, Cui et al. Cui et al. (2022c) released a comprehensive multi-center CBCT dataset with instance-level annotations, establishing a robust benchmark for CBCT tooth segmentation. NKUT Zhou et al. (2024) specifically tackled pediatric tooth germ extraction by introducing a pediatric CBCT dataset for mandibular tooth germ and associated tissue segmentation. As the DIA field continues to evolve, segmenting individual teeth is increasingly insufficient for advanced dental CAD systems. Radiographic images inherently capture both teeth and maxillofacial anatomical structures, prompting researchers to develop datasets with finer-grained annotations and multi-task capabilities. These datasets not only enhance segmentation from the tooth level to the anatomical level but also broaden the scope of DIA applications. For example, ToothFairy Bolelli et al. (2025) goes beyond tooth instance segmentation to include maxillofacial structures and mandibular nerve canal segmentation. Pulpy3D Gamal et al. (2024) and Liang et al. Liang et al. (2024) introduced CBCT datasets for pulp segmentation. PRAD Zhou et al. (2025b) presented a large-scale PR dataset with segmentation annotations for nine anatomical structures or appliances. Dong et al. Dong et al. (2023) and DentalDS proposed PAN datasets for pediatric tooth maturity staging. These radiographic datasets have significantly advanced research in DIA.

In contrast to radiographic imaging, IP excels in directly capturing occlusal relationships and dental pathologies, enabling a more intuitive assessment of surface-level dental features that may not be clearly visible in radiographic modalities. Consequently, as shown in Fig. 7 (c), public IP datasets primarily focus on tooth detection, segmentation, and disease classification tasks. For example, the Dental Anatomy Siddhartha (2024) dataset was introduced for tooth detection; Liu et al. Liu et al. (2025b) and Huang et al. Huang et al. (2025b) developed IP datasets targeting various types of dental cracks and fractures. The Tooth Decay dataset Aminhj (2023) offers a large collection of IP images for caries classification. Furthermore, Zou et al. Zou et al. (2024) proposed an IP dataset annotated with instance-level tooth segmentation. On the other hand, IOS supports more precise applications by generating high-resolution 3D digital dental models that accurately capture surface morphology. As shown in Fig. 7 (b), these datasets are particularly valuable for orthodontic treatment planning and the fabrication of dental prostheses or aligners. Consequently, public IOS datasets are predominantly dedicated to precise tooth segmentation. For instance, IOS150K Zou et al. (2024) introduced a large-scale 2D IOS screenshot dataset for instance-level segmentation. In addition, 3D-IOSeg Li et al. (2024b), Teeth3DS+ Ben-Hamadou et al. (2023),

Table 1: Summary of Publicly Available Dental Datasets up to and including September 2025.  denotes datasets that can be directly downloaded via the provided link or upon registration, whereas  signifies datasets for which the download link must be acquired by contacting the paper’s authors.

| Dataset                                                               | Country     | Year | Modality      | Quantity | Task                                        | Overview                                                                                                                                                                                                                    | Acquisition                                                                           | License         | URL                  |
|-----------------------------------------------------------------------|-------------|------|---------------|----------|---------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------------|----------------------|
| Dental X-Ray <a href="#">Rad et al. (2016)</a>                        | Iran        | 2016 | PR            | 120      | Caries Detection                            | PAN with annotations of 152 individual teeth affected by caries.                                                                                                                                                            |    | /               | <a href="#">Link</a> |
| TDD <a href="#">Panetta et al. (2021)</a>                             | USA         | 2021 | PAN           | 1000     | Tooth Segmentation                          | PAN with corresponding tooth masks, eye-tracking maps, audio transcription texts, and ROIs for the mandibular and maxillary areas.                                                                                          |    | CC BY 4.0       | <a href="#">Link</a> |
| Dental X-Ray Images <a href="#">Kaggle (2019)</a>                     | South Korea | 2021 | PAN           | 64       | Detection<br>Tooth Segmentation             | PAN in PNG format, each accompanied by tooth detection and semantic segmentation annotations.                                                                                                                               |    | CC BY-NC-SA 4.0 | <a href="#">Link</a> |
| CTooth <a href="#">Cui et al. (2022b)</a>                             | China       | 2022 | CBCT          | 22       | Tooth Segmentation                          | CBCT in NIfTI format, pixel spacing: $0.25 \times 0.25 \text{ mm}^2$ , resolution: $256 \times 256$ .                                                                                                                       |    | CC BY 4.0       | <a href="#">Link</a> |
| CTooth+ <a href="#">Cui et al. (2022a)</a>                            | China       | 2022 | CBCT          | 168      | Tooth Segmentation                          | In addition to the CTooth dataset, 146 unlabeled CBCT scans were included.                                                                                                                                                  |    | CC BY-NC 4.0    | <a href="#">Link</a> |
| <a href="#">Cui et al. (2022c)</a>                                    | China       | 2022 | CBCT          | 150      | Tooth Segmentation                          | The CBCT data are provided with instance-level tooth segmentation masks.                                                                                                                                                    |    | CC BY 4.0       | <a href="#">Link</a> |
| Vident-lab <a href="#">Katsaros et al. (2022)</a>                     | Poland      | 2022 | Dental-Videos | 1000     | Tooth Segmentation                          | An asymmetrically annotated video dataset of natural teeth in phantom scenes is introduced for multi-task video processing, including tooth restoration, tooth segmentation, and inter-frame homography estimation.         |    | CC BY-NC 4.0    | <a href="#">Link</a> |
| ToothFairy <a href="#">Cipriano et al. (2022)</a>                     | Italy       | 2022 | CBCT          | 347      | IAN Segmentation                            | Collected from Affidea Medical Center, the dataset provides both sparse and dense annotations, including IAN segmentation masks.                                                                                            |    | /               | <a href="#">Link</a> |
| $\infty$ Panoramic Dental Xray <a href="#">Brahmi and Jdey (2023)</a> | Tunisia     | 2023 | PAN           | 107      | /                                           | 107 PAN images with a resolution of $2964 \times 1464$ , without annotations.                                                                                                                                               |    | CC BY 4.0       | <a href="#">Link</a> |
| <a href="#">Zhang et al. (2023d)</a>                                  | China       | 2023 | PAN           | 100      | Caries Segmentation<br>Disease Detection    | pediatric PAN includes annotations for overall tooth structure segmentation and detection annotations based on dental disease classification.                                                                               |    | CC BY 4.0       | <a href="#">Link</a> |
| Dentex <a href="#">Hamamci et al. (2023)</a>                          | Switzerland | 2023 | PAN           | 3,903    | Abnormal Teeth-Detection and Classification | PAN from three different institutions, annotated with quadrants, tooth numbers, and types of dental lesions.                                                                                                                |    | CC-BY 4.0       | <a href="#">Link</a> |
| <a href="#">Dong et al. (2023)</a>                                    | China       | 2023 | PAN           | 673      | Maturity Staging                            | Patients aged 3 to 14 years and include orthodontic brackets as well as slight blurring or deformation. Permanent tooth maturity was annotated using FDI tooth numbering in combination with the modified Demirjian method. |    | /               | /                    |
| Teeth3DS+ <a href="#">Ben-Hamadou et al. (2023)</a>                   | Tunisia     | 2023 | IOS           | 1800     | Tooth Segmentation                          | The dataset includes over 900 patients, featuring instance-level annotations of teeth along with key landmarks.                                                                                                             |  | CC BY-NC-ND 4.0 | <a href="#">Link</a> |
| Dental Radiography Team <a href="#">(2023)</a>                        | Iran        | 2023 | PAN           | 1272     | /                                           | Original PAN images without annotations and provide training, validation and test set grouping                                                                                                                              |  | CC BY-SA 4.0    | <a href="#">Link</a> |
| DC 1000 <a href="#">Wang et al. (2023a)</a>                           | China       | 2023 | PAN & BR      | 2986     | Caries Segmentation                         | The dataset consists of 2,389 BR and 597 PAN, each accompanied by corresponding caries segmentation labels.                                                                                                                 |  | CC BY-SA 4.0    | <a href="#">Link</a> |
| <a href="#">Loop (2023)</a>                                           | Bulgaria    | 2023 | PAN           | 598      | Tooth Segmentation                          | The dataset comprises 598 PAN with instance-level tooth segmentation labels.                                                                                                                                                |  | CC0 1.0 license | <a href="#">Link</a> |
| <a href="#">Pivoshenko (2023)</a>                                     | U.K.        | 2023 | PAN           | 116      | Mandible Segmentation                       | This dataset includes 116 panoramic radiographs (PAN) with corresponding mandibular segmentation masks, covering a wide range of dental conditions from complete dentition to partial edentulism and full edentulism.       |  | CC BY-SA 3.0    | <a href="#">Link</a> |
| MICCAI 2023 STS Challenge <a href="#">ricolee-hduu (2023)</a>         | China       | 2023 | CBCT & PAN    | 580+6500 | Tooth Segmentation                          | This dataset contains images along with corresponding semantic segmentation masks of the tooth regions.                                                                                                                     |  | CC BY-NC-ND 4.0 | <a href="#">Link</a> |

(continued on next page)

Table 1 (Continued from the last page. IAN: Inferior Alveolar Nerve)

| Dataset                                                        | Country    | Year | Modality    | Quantity    | Task                                      | Overview                                                                                                                                                                                                                                      | Acquisition                                                                           | License         | URL                  |
|----------------------------------------------------------------|------------|------|-------------|-------------|-------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|-----------------|----------------------|
| Pulpy3D <a href="#">Gamal et al. (2024)</a>                    | Egypt      | 2024 | CBCT        | 548         | Pulp Segmentation<br>IAN Segmentation     | An extension of the ToothFairy, with annotations including both semantic and instance-level labels for the pulp and the IAN. 150 cases were manually annotated, while the remaining annotations were generated automatically by DL model.     |    | CC BY-SA 4.0    | <a href="#">Link</a> |
| IO150K <a href="#">Zou et al. (2024)</a>                       | China      | 2024 | IOS & IP    | 150K        | Tooth Segmentation                        | The dataset comprises 2D screenshots derived from IOS along with corresponding instance-level annotations of teeth, all of which were generated using a human-in-the-loop annotation approach.                                                |    | CC BY-SA 4.0    | <a href="#">Link</a> |
| Multimodal Dental Dataset <a href="#">Huang et al. (2024)</a>  | China      | 2024 | CBCT<br>PAN | 8<br>329    | Multimodal learning                       | The dataset contains multi-modal dental images from 169 patients, where 16,203 PR were not directly captured but generated from CBCT scans.                                                                                                   |    | CC BY 4.0       | <a href="#">Link</a> |
| Orthodontic Dental Dataset <a href="#">Wang et al. (2024d)</a> | China      | 2024 | IOS         | 1060        | Tooth Segmentation<br>Landmarks Detection | The dataset includes 1,060 pre- and post-treatment dental arch models from 435 patients, along with tooth instance segmentation annotations and crown landmark annotations.                                                                   |    | CC 0            | <a href="#">Link</a> |
| NKUT <a href="#">Zhou et al. (2024)</a>                        | China      | 2024 | CBCT        | 133         | Tooth Segmentation<br>Bone Segmentation   | This dataset contains pediatric CBCT images with segmentation annotations for mandibular third molar germs, second molars, and the alveolar bone.                                                                                             |    | CC BY-NC 4.0    | <a href="#">Link</a> |
| Dental OPG X-RAY <a href="#">Rahman et al. (2024)</a>          | Bangladesh | 2024 | PAN         | 232         | Lesion detection & classification.        | This dataset includes PAN and classifies teeth into six categories: healthy teeth, dental caries, impacted teeth, infections, tooth fractures, and broken crowns or roots.                                                                    |    | CC BY-NC 4.0    | <a href="#">Link</a> |
| <a href="#">Li et al. (2024e)</a>                              | China      | 2024 | PAN         | 6536        | Lesion detection & classification         | This dataset consists of three sets of data collected from different hospitals, covering dental conditions such as impacted teeth, periodontitis, and dental caries, with annotations including segmentation masks and classification labels. |    | CC BY-NC 4.0    | <a href="#">Link</a> |
| Tooth decay <a href="#">Aminhj (2023)</a>                      | Iran       | 2024 | PR & IP     | 8579+       | Lesion classification                     | This dataset covers the detection and analysis of dental caries, including images of decayed and damaged teeth.                                                                                                                               |    | MIT             | <a href="#">Link</a> |
| <a href="#">Seifossadat (2023)</a>                             | Iran       | 2024 | PAN         | 29591       | Lesion classification                     | This dataset covers the detection and analysis of dental caries, including images of decayed and damaged teeth.                                                                                                                               |    | CC BY-SA 4.0    | <a href="#">Link</a> |
| Dental Anatomy <a href="#">Siddhartha (2024)</a>               | India      | 2024 | IP          | 724         | Tooth detection                           | used for tooth detection and is divided into training, validation, and test sets, containing 505, 112, and 107 images with corresponding annotations in each folder, respectively.                                                            |  | CC BY-SA 4.0    | <a href="#">Link</a> |
| <a href="#">Silvres (2024)</a>                                 | India      | 2024 | PAN         | 13932       | Lesion Detection                          | This dataset contains a large number of panoramic X-ray images (PAN) along with detection labels for various dental diseases, annotated in both YOLO and COCO formats.                                                                        |  | Apache 2.0      | <a href="#">Link</a> |
| <a href="#">Arya (2024)</a>                                    | India      | 2024 | PAN         | 70          | Tooth Segmentation                        | The data are high-quality samples obtained from local hospitals and trusted acquaintances, accompanied by semantic segmentation labels for the teeth.                                                                                         |  | CC BY-SA 3.0    | <a href="#">Link</a> |
| Root Disease X-Ray <a href="#">UBU (2024)</a>                  | Thailand   | 2024 | PR          | 6578        | Disease Classification                    | PR used for dental disease classification.                                                                                                                                                                                                    |  | CC BY-NC-SA 4.0 | <a href="#">Link</a> |
| PubMedVision <a href="#">Chen et al. (2024a)</a>               | China      | 2024 | Various     | 14212       | Image-Text Pretraining                    | A medical image-text paired dataset containing 646,759 image-text pairs, of which 14,212 are related to dentistry, featuring diverse formats and including both Chinese and English bilingual data.                                           |  | /               | <a href="#">Link</a> |
| 3D-IOSeg <a href="#">Li et al. (2024b)</a>                     | China      | 2024 | IOS         | 440         | Tooth Segmentation                        | The dataset includes IOS from 220 patients, covering various dental diseases with instance-level segmentation annotations, and incorporates extreme dental conditions to enhance model robustness.                                            |  | /               | <a href="#">Link</a> |
| MICCAI 2024 STS Challenge <a href="#">Wang et al. (2024f)</a>  | China      | 2024 | CBCT<br>PAN | 350<br>2400 | Tooth Segmentation                        | The 2nd STS Challenge dataset includes paired panoramic radiographs (PAN) and CBCT scans, with instance-level tooth segmentation annotations.                                                                                                 |  | /               | <a href="#">Link</a> |
| <a href="#">Liang et al. (2024)</a>                            | China      | 2024 | CBCT        | 300         | Tooth Segmentation<br>Pulp Segmentation   | CBCT scans from 300 anonymous patients, with semantic-level segmentation masks for teeth and root canals.                                                                                                                                     |  | /               | <a href="#">Link</a> |

(continued on next page)

Table 1 (Continued from the last page.)

| Dataset                                             | Country      | Year | Modality    | Quantity   | Task                                              | Overview                                                                                                                                                                                                                                          | Acquisition                                                                           | License            | URL                  |
|-----------------------------------------------------|--------------|------|-------------|------------|---------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|---------------------------------------------------------------------------------------|--------------------|----------------------|
| TRPR <a href="#">Silva et al. (2025)</a> *          | Brazil       | 2025 | PAN         | 8029       | Image-Text Pretraining                            | The TRPR dataset consists of PAN and their corresponding textual reports, without tooth segmentation annotations. The reports were written in Portuguese by two dental experts.                                                                   |    | /                  | <a href="#">Link</a> |
| MMDental <a href="#">Wang et al. (2025a)</a>        | China        | 2025 | CBCT        | 660        | Image-Text Pretraining                            | MMDental dataset includes CBCT and corresponding detailed expert medical records, along with both initial and follow-up records.                                                                                                                  |    | CC BY-NC-ND 4.0    | <a href="#">Link</a> |
| FDTooth <a href="#">Liu et al. (2025b)</a>          | HK, China    | 2025 | CBCT & IP   | 241        | Fenestration and Dehiscence Detection             | FDTooth dataset consists of paired IP and CBCT scans of both children and adults, including 2,892 annotated bounding boxes of anterior teeth with fenestration or Dehiscence.                                                                     |    | PhysioNet 1.5      | <a href="#">Link</a> |
| PRAD <a href="#">Zhou et al. (2025b)</a>            | China        | 2025 | PR          | 10K        | Anatomical-level Tooth Segmentation               | The PRAD dataset includes real clinical PR along with corresponding anatomical-level segmentation masks, covering nine types of tooth structures, lesions, and instruments.                                                                       |    | CC BY-NC 4.0       | <a href="#">Link</a> |
| DentalDS <a href="#">Wang et al. (2025c)</a>        | China        | 2025 | PAN         | 2583       | Tooth development grading                         | DentalDS dataset contains images of pediatric orthodontic patients, with annotations including bounding boxes for the mandibular teeth and developmental stage labels based on the Demirjian method, totaling 18,081 annotated teeth.             |    | CC CC BY-NC-ND 4.0 | <a href="#">Link</a> |
| STS-Tooth <a href="#">Wang et al. (2025f)</a>       | China        | 2025 | PAN<br>CBCT | 4K<br>148K | Tooth Segmentation                                | This dataset includes multimodal PAN and CBCT data from both pediatric and adult patients, providing 900 PAN images and 8,800 CBCT scans with corresponding tooth semantic segmentation masks.                                                    |    | CC BY-NC-ND 4.0    | <a href="#">Link</a> |
| <a href="#">Faizan Ahmed et al. (2025)</a>          | Pakistan     | 2025 | IP          | 6313       | Caries Detection                                  | This dataset includes high-resolution IP of children and adolescents, accompanied by dental caries detection annotations in multiple formats.                                                                                                     |    | CC BY-NC-ND 4.0    | <a href="#">Link</a> |
| ToothFairy2 <a href="#">Boelli et al. (2025)</a>    | Italy        | 2025 | CBCT        | 530        | Maxillofacial Structure Segmentation              | The dataset includes patients' CBCT scans along with voxel-level 3D instance annotations corresponding to 42 different maxillofacial structure categories.                                                                                        |    | CC BY-SA           | <a href="#">Link</a> |
| OPG <a href="#">Kennedy Orvile (2024)</a>           | Saudi Arabia | 2025 | PAN         | 622        | Lesion Classification                             | The dataset includes panoramic X-ray images (PAN) with corresponding bounding box annotations, covering categories such as Broken Roots, Periodontally Compromised Teeth, and Kennedy Classification of Partially Edentulous Arches (Class I-IV). |    | CC BY 4.0          | <a href="#">Link</a> |
| <a href="#">Ali (2024)</a>                          | Egypt        | 2025 | PAN         | 8188       | Lesion Segmentation                               | The dataset includes panoramic X-ray images (PAN) with semantic-level segmentation masks for various lesions, without distinguishing between specific disease types.                                                                              |  | /                  | <a href="#">Link</a> |
| <a href="#">Arya (2024)</a>                         | Egypt        | 2025 | PAN         | 335        | Caries Detection                                  | This dataset is adapted from other public datasets for detecting dental caries in children and includes bounding box annotations.                                                                                                                 |  | MIT                | <a href="#">Link</a> |
| MICCAI 2025 STS <a href="#">Wang et al. (2025e)</a> | China        | 2025 | CBCT<br>IOS | 780        | Tooth & Pulp segmentation, CBCT-IOS Registration. | The segmentation subset provides semantic masks of teeth and pulp in CBCT images; the registration subset includes 380 CBCT cases paired with corresponding IOS, along with affine transformation matrices.                                       |  | /                  | <a href="#">Link</a> |
| <a href="#">Huang et al. (2025b)</a>                | China        | 2025 | IP          | 2210       | Tooth Crack Detection                             | Artificial cracks were created on 432 teeth, and high-quality images were selected after photographing with a standard camera, covering multiple types of cracks.                                                                                 |  | /                  | <a href="#">Link</a> |
| <a href="#">Sehar et al. (2025)</a>                 | China        | 2025 | CBCT        | 105        | Tooth Segmentation                                | The dataset includes CBCT scans from 105 patients, totaling approximately 14,000 tooth samples, annotated according to the FDI tooth numbering system.                                                                                            |  | LLC protocol       | <a href="#">Link</a> |
| (End of Table 1)                                    |              |      |             |            |                                                   |                                                                                                                                                                                                                                                   |                                                                                       |                    |                      |

and Wang et al. Wang et al. (2024d) released datasets of varying scales in 3D point cloud format for instance-level tooth segmentation on IOS data.

With the recent rise of vision-language models (VLM) in medical image analysis, several researchers have proposed image-text datasets tailored for dental applications. As illustrated in Fig. 7 (d), PubMedVision Chen et al. (2024a) is constructed from carefully curated images and figure captions extracted from PubMed articles and medical textbooks. Among its more than 640,000 image-text pairs, approximately 14,212 pairs pertain to dentistry. The dataset includes diverse modalities, such as radiographs, IOS, IP, intraoperative color images, and statistical figures. In 2025, TRPR Silva et al. (2025) released a PAN dataset with images and corresponding Portuguese reports, intended for lesion classification and VLM pre-training. MMDental Wang et al. (2025a) introduced a multimodal dataset consisting of 403 CBCT volumes and 660 clinical cases, encompassing both initial and follow-up records, designed for multimodal learning.

Notably, since 2023, MICCAI has organized the Semi-supervised Teeth Segmentation (STS) Challenge for three consecutive years, underscoring the community’s growing focus on DIA. The STS tasks and data modalities include semantic- and instance-level segmentation of teeth on PAN and CBCT images ricolleehduu (2023); Wang et al. (2024f), as well as semantic segmentation of teeth and pulp on CBCT, and CBCT-IOS registration Wang et al. (2025e). These diverse publicly available dental datasets and community challenges, spanning multiple modalities and tasks, have established a robust foundation for applying deep learning in DIA.

### 3. Basic technologies

#### 3.1. Overview

With the continuous accumulation of data resources and growing clinical demands, the application of DL in DIA has become increasingly widespread, fostering the development of deep models based on diverse network architectures. We categorize DIA tasks into three major groups: dense prediction tasks, classification tasks, and other tasks. Among these, dense prediction tasks, primarily represented by segmentation, constitute the dominant direction of image processing in the dental field, making up the vast majority of the surveyed literature (182 papers, 86.26%). Studies indicate that precise segmentation of teeth, anatomical structures, and lesions in dental images serves as the essential foundation and prerequisite for all other tasks. In this section, we first introduce several fundamental DL network models, upon which most of the surveyed models are developed as variants or through fusion. Subsequently, we provide a comprehensive review and discussion of 211 related papers, organized by task categories.

#### 3.2. Fundamental Techniques

DL-based approaches for DIA can generally be categorized into three types: CNN-based methods, Transformer-based methods, and hybrid methods that integrate the strengths of both. The following subsections provide a detailed introduction to each.

##### 3.2.1. CNN

The basic structure of a CNN is illustrated in Fig. 8. In the early 21st century, as modern CNNs achieved breakthrough success in handwritten digit recognition tasks LeCun et al. (2002), CNNs gradually evolved into a cornerstone technology in the field of computer vision. Since then, CNNs have advanced rapidly, significantly surpassing traditional approaches in the ImageNet classification task Krizhevsky et al. (2012). The network depth has increased Simonyan and Zisserman (2014), while multi-scale feature extraction capabilities and scalability have been greatly enhanced Szegedy et al. (2015). The residual learning framework introduced by ResNet He et al. (2016) established a paradigm for deep models, effectively addressing the issues of vanishing and exploding gradients in very deep networks. Meanwhile, the advent of Fully Convolutional Networks (FCN) Long et al. (2015) enabled end-to-end semantic segmentation for the first time, equipping CNNs with pixel-level classification capability. In the field of medical image analysis, UNet Ronneberger et al. (2015) stands as the most influential deep learning model, originally proposed for cell segmentation in microscopic images. Its symmetric encoder-decoder structure and skip connections are crucial for capturing details and boundaries in medical images, which has led to its widespread adoption across diverse medical domains Chen et al. (2025a); Mirikharaji et al. (2023). Numerous variants have subsequently been proposed, such as UNet++ Zhou et al. (2019), Attention-UNet Oktay et al. (2018), 3D-UNet Çiçek et al. (2016), and VNet Milletari et al. (2016). In the dental domain, CNNs are likewise the most widely applied core technology, with a large body of research in various DIA tasks developed on the foundation of UNet and its variants.

##### 3.2.2. Transformer

Transformer Vaswani et al. (2017) was initially introduced in 2017 for machine translation and has since become a cornerstone of natural language processing. The key component of the Transformer is the self-attention mechanism, which enables the model to learn the relevance of each token to all other tokens and perform feature weighting accordingly. This mechanism endows the model with the ability to capture global features while effectively modeling long-range dependencies. In 2020, with the introduction of Vision Transformer (ViT) Dosovitskiy et al. (2020), the Transformer architecture officially entered the field of computer vision, achieving remarkable success. A significant improvement over ViT, the Swin

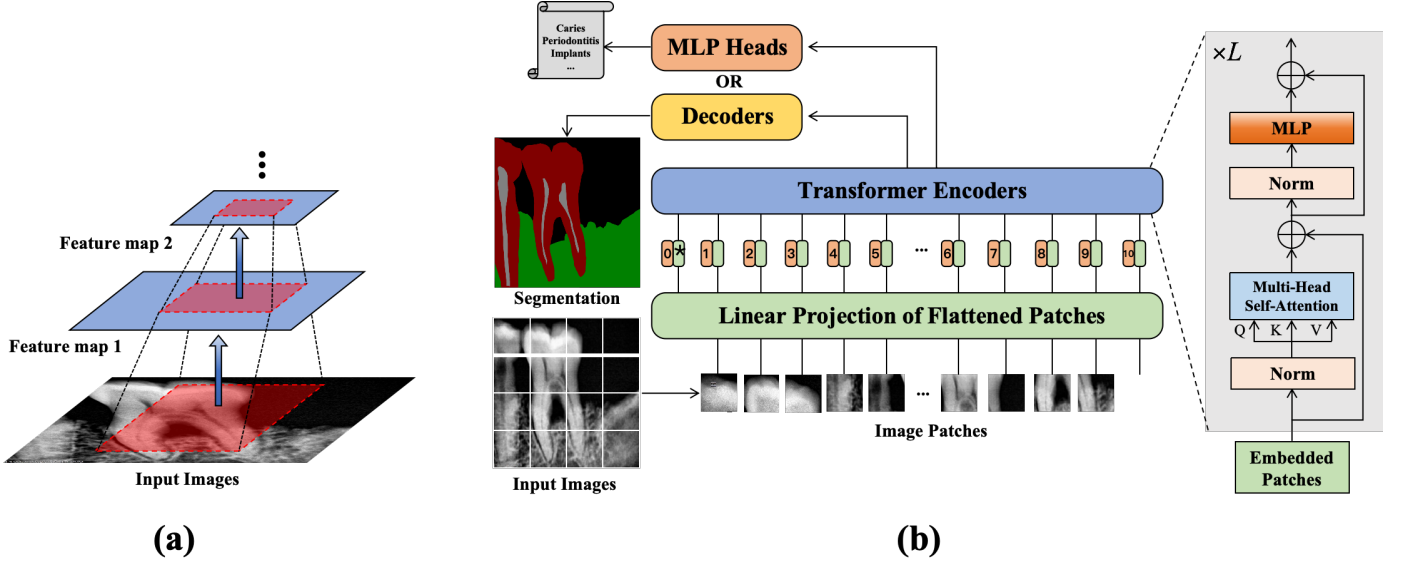


Figure 8: Illustration of fundamental model architectures in the DIA field: (a) depicts the schematic of a CNN, where convolutional kernels extract local features and progressively reduce the resolution of feature maps through hierarchical processing; (b) shows a ViT-based network structure, where the input image is divided into patches and fed into a ViT encoder to encode global features via self-attention mechanisms, followed by a decoder or classification head to generate task-specific outputs. The PR image shown in the figure is sourced from [Zhou et al. \(2025b\)](#).

Transformer [Liu et al. \(2021\)](#) introduces a hierarchical structure to mimic the pyramidal data flow of CNNs, making it more suitable for downstream dense prediction tasks, and employs the Shifted Window Attention mechanism to reduce computational overhead. In the domain of medical image analysis, ViT-based models quickly emerged and demonstrated outstanding performance across a wide range of tasks in different medical specialties [Shamshad et al. \(2023\)](#). TransUnet [Chen et al. \(2021a\)](#) is one of the earliest models to incorporate ViT into medical image analysis, enabling efficient segmentation of abdominal organs by integrating a ViT encoder into its architecture. Swin-UNet [Cao et al. \(2022\)](#) is a purely ViT-based model built upon the Swin Transformer’s pyramid-shaped hierarchical structure, demonstrating strong competitiveness in medical image segmentation tasks. MissFormer [Huang et al. \(2021b\)](#) further improves upon this by incorporating multi-scale feature extraction and attention mechanisms, exhibiting higher robustness and stronger generalization in medical image segmentation. In dental image analysis, Transformers have gradually reached parity with CNNs and even surpassed them in certain tasks.

### 3.2.3. Hybrid Model

Transformer [Vaswani et al. \(2017\)](#) was initially introduced in 2017 for machine translation and has since become a cornerstone of natural language processing. The key component of the Transformer is the self-attention mechanism, which enables the model to learn the relevance of each token to all other tokens and perform feature weighting accordingly. This mechanism endows the model with the ability to capture global features while effectively modeling long-range dependencies. In 2020, with

the introduction of Vision Transformer (ViT) [Dosovitskiy et al. \(2020\)](#), the Transformer architecture officially entered the field of computer vision, achieving remarkable success. A significant improvement over ViT, the Swin Transformer [Liu et al. \(2021\)](#) introduces a hierarchical structure to mimic the pyramidal data flow of CNNs, making it more suitable for downstream dense prediction tasks, and employs the Shifted Window Attention mechanism to reduce computational overhead. In the domain of medical image analysis, ViT-based models quickly emerged and demonstrated outstanding performance across a wide range of tasks in different medical specialties [Shamshad et al. \(2023\)](#). TransUnet [Chen et al. \(2021a\)](#) is one of the earliest models to incorporate ViT into medical image analysis, enabling efficient segmentation of abdominal organs by integrating a ViT encoder into its architecture. Swin-UNet [Cao et al. \(2022\)](#) is a purely ViT-based model built upon the Swin Transformer’s pyramid-shaped hierarchical structure, demonstrating strong competitiveness in medical image segmentation tasks. MissFormer [Huang et al. \(2021b\)](#) further improves upon this by incorporating multi-scale feature extraction and attention mechanisms, exhibiting higher robustness and stronger generalization in medical image segmentation. In dental image analysis, Transformers have gradually reached parity with CNNs and even surpassed them in certain tasks.

## 4. Dental Tasks and Models

### 4.1. Dense prediction

With the rapid development and increasing maturity of digital dentistry, dense prediction tasks, particularly

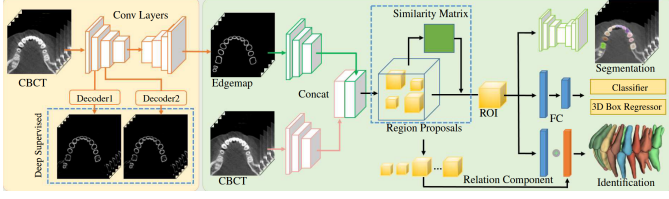


Figure 9: Overview of ToothNet. Built upon the 3D Mask R-CNN backbone, it incorporates boundary maps and similarity matrices to assist in instance-level tooth segmentation on CBCT images. ToothNet represents the first pioneering work employing DL techniques for CBCT-based tooth segmentation. Image sourced from Cui et al. (2019).

segmentation, have emerged as the most widely studied, longest-standing, and most extensively explored direction in the DIA field. Efficient segmentation and detection of dental and maxillofacial structures across various imaging modalities not only significantly reduce manual labor and provide more objective diagnostic support but also serve as a fundamental basis for numerous preoperative dental applications Dutra et al. (2016), such as orthodontic surgery planning Zou et al. (2024), endodontic treatment planning Wang et al. (2023d), and tooth extraction and implant planning Ma et al. (2023a). In this section, we provide a systematic review of the various attempts and advancements in applying ensemble DL models to dense prediction tasks in DIA. The overall categorization framework is illustrated in Fig. 2.

#### 4.1.1. Tooth

Tooth dense prediction is one of the core tasks in DIA and is often employed as a crucial preprocessing step for subsequent tasks. The application of CNN-based algorithms to DIA can be traced back to around 2016 Eun and Kim (2016); Choi et al. (2018), since which time DL has seen increasingly widespread use in DIA.

**Tooth Level.** We define the prediction of teeth as complete entities, without distinguishing their internal anatomical structures, as tooth-level tasks, as shown in Fig. ?? (a)-(f). Tooth-level tasks form the foundation of all DIA research and account for the vast majority of existing studies, as summarized in Tables 2 and 3. Based on the granularity of representation, these tasks can be further categorized into two types: semantic-level Majanga and Viriri (2019); Sivagami et al. (2020); Feng (2023); Zhao et al. (2021a) and instance-level Zhao et al. (2021b); Lin et al. (2023c); Zhao et al. (2024c); Xu et al. (2025a). The former treats all teeth as a single unified region, whereas the latter distinguishes each tooth as an independent instance. In clinical practice, the FDI tooth numbering system, as illustrated in Fig. 13 (a), is commonly used to facilitate more fine-grained analysis and diagnosis.

Prior to 2023, CNNs were the predominant architectures employed for tooth-level dense prediction tasks. As summarized in Tables 2, during this stage, many researchers

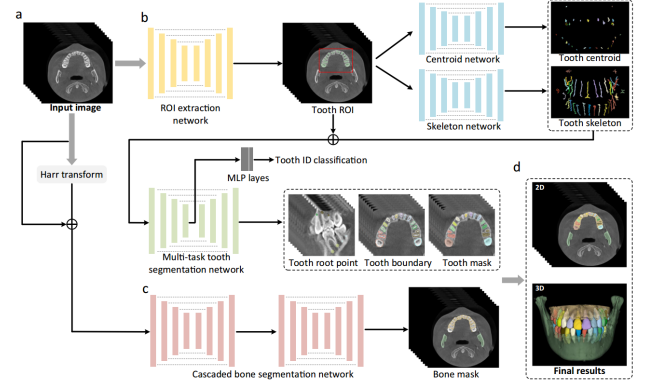


Figure 10: Overview of a multi-task CBCT segmentation framework. It introduces tooth centroid prediction and tooth skeleton prediction as auxiliary tasks to enhance the accuracy of tooth numbering and localization, ultimately achieving precise segmentation of teeth and jaw structures in CBCT images. Image sourced from Cui et al. (2022c).

also initiated various explorations of DL-based DIA, leading to a wide range of research efforts. In 2019, Cui et al. proposed ToothNet Cui et al. (2019), the first approach to apply DL to CBCT tooth instance segmentation. As shown in Fig. 9, ToothNet employs a two-stage architecture: in the first stage, a boundary extraction network is used to capture boundary information from the input; in the second stage, the extracted boundary features are fused with the original input through a similarity matrix and subsequently fed into a 3D region proposal network to perform segmentation. This multi-stage concept has also provided valuable insights for subsequent studies. For instance, Yang et al. Yang et al. (2021) combined traditional algorithms with DL by first using U-Net to localize tooth centroids, followed by level-set algorithms to achieve precise CBCT tooth segmentation. Wu et al. Wu et al. (2022) proposed a two-stage approach integrating iMesh-SegNet and PointNet to accomplish tooth segmentation and landmark detection on IOS data. Jing et al. Jing et al. (2024) adopted a similar two-stage strategy, where a coarse segmentation network was applied to CBCT images to obtain tooth ROIs, which were then refined using an improved V-Net-based segmentation network. Moreover, as shown in Fig. 10, Cui et al. further developed an automated CBCT tooth segmentation system Cui et al. (2022c) that incorporates auxiliary predictions of tooth centroids and skeletons, achieving highly accurate tooth instance segmentation on a large-scale CBCT dataset.

In addition, certain research approaches have integrated CNNs with attention mechanisms or employed cross-domain analysis to enhance the performance of tooth-level tasks across different modalities. For example, Zhong et al. Zhong et al. (2025) proposed a UNet variant incorporating a Grouped Global Attention module and cross-layer fusion, which effectively captured texture and contour information from low-level features in PAN images, thereby enabling

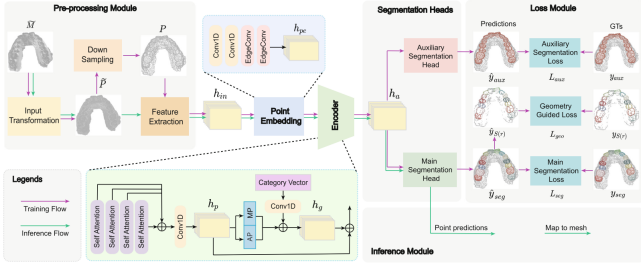


Figure 11: The overall architecture of TsegFormer, a 3D ViT-based framework designed to address the complex boundary challenges in IOS tooth segmentation. By integrating a multi-task learning paradigm with a self-attention mechanism, TsegFormer effectively captures both local and global dependencies between teeth and gingiva, enabling more precise segmentation outcomes. Image sourced from [Xiong et al. \(2023\)](#).

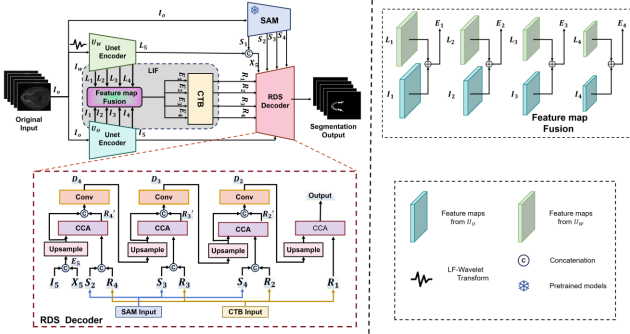


Figure 12: The overall architecture of FNet. In the Low-Frequency Image Fusion module, LF-Wavelet is introduced to enhance semantic representation and global structural integrity. Additionally, a large-scale ViT-based pretrained SAM encoder is employed to refine boundary features. FNet effectively mitigates artifact interference and semantic gaps in CBCT dental images. Image sourced from [Feng et al. \(2024\)](#).

efficient tooth-level semantic segmentation. Long et al. [Long and Yuan \(2024\)](#) introduced a Differential Feature Fusion module into the skip connections of UNet, integrating both spatial and channel attention mechanisms. This design enhanced feature extraction while alleviating information loss during the downsampling and upsampling processes, resulting in accurate tooth segmentation in CBCT images. Zhou et al. [Zhou et al. \(2025a\)](#) developed a tooth segmentation network for IOS images based on multi-scale graph convolution and cross-domain feature fusion, leveraging the integration of global and local fine-grained information to overcome variations in tooth morphology and employing attention mechanisms for cross-domain guidance. Pan et al. [Pan et al. \(2025b\)](#) proposed a multi-domain omni-perception collaborative guidance network that decomposed high- and low-frequency features and incorporated a Cross-Attention mechanism, thus achieving deep frequency information alignment and robust edge feature learning, ultimately surpassing previous methods in IP image tooth-level segmentation tasks.

As summarized in Tables 3, with the advent of the

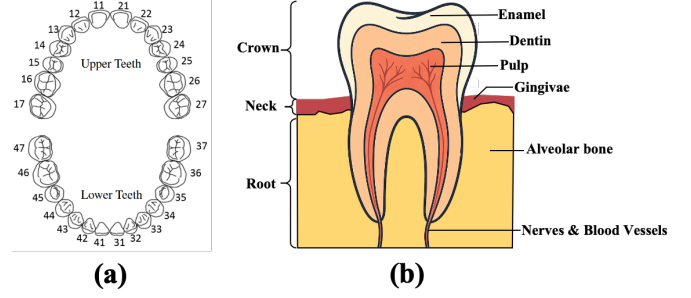


Figure 13: Fig (a) illustrates the FDI tooth numbering system, commonly utilized for tooth-level instance segmentation and labeling, while Fig (b) depicts the anatomical structure of a tooth, serving as the target for anatomy-level prediction tasks.

ViT and its growing application in medical image analysis, many researchers have focused on exploring its potential in DIA tasks. A variety of models based on both pure ViT and hybrid CNN-ViT architectures have been investigated. For pure ViT models, Xiong et al. introduced TSegFormer [Xiong et al. \(2023\)](#), as depicted in Fig. 11, a 3D Transformer-based framework designed to address complex boundaries in tooth segmentation. By leveraging a multi-task learning paradigm combined with a self-attention mechanism, the model captures both local and global dependencies between teeth and gingiva, enabling accurate instance segmentation on IOS data. Sheng et al. [Sheng et al. \(2023\)](#) collected a dataset of 100 adult PAN images and evaluated multiple networks, demonstrating that the pure ViT-based Swin-Unet [Cao et al. \(2022\)](#) outperformed other methods, thereby validating the reliability of ViT in DIA tasks. Dai et al. [Dai et al. \(2024\)](#) further proposed a pure ViT-based self-supervised framework that integrates a Masked Autoencoder with a sparse boundary prompting mechanism to achieve tooth instance segmentation on CBCT data. Compared with pure ViTs, hybrid models are more widely applied in DIA tasks. This is primarily because DIA tasks typically require models to capture global structures while simultaneously attending to local details, enabling more comprehensive and accurate feature representation. For instance, some researchers have combined CNNs and ViTs in encoder architectures to capture both rich multi-scale local details and global long-range dependencies in dental images [Manzari et al. \(2024\)](#); [Xi et al. \(2025\)](#); [Ahn and Cho \(2024\)](#); [Huang et al. \(2025a\)](#), thereby leveraging the advantages of hybrid models. Other studies have designed efficient cross-attention mechanisms tailored for dental images to achieve highly accurate tooth-level predictions [Wang et al. \(2024a\)](#); [Li et al. \(2024d,c\)](#). More recently, large-scale foundation models, such as SAM [Kirillov et al. \(2023\)](#) and Med-SAM [Ma et al. \(2024a\)](#), have been introduced into the DIA domain to harness their rich knowledge and enable efficient transfer to dental applications [Zhang et al. \(2025a\)](#); [Wang et al. \(2025b\)](#); [Feng et al. \(2024\)](#) (Fig. 12).

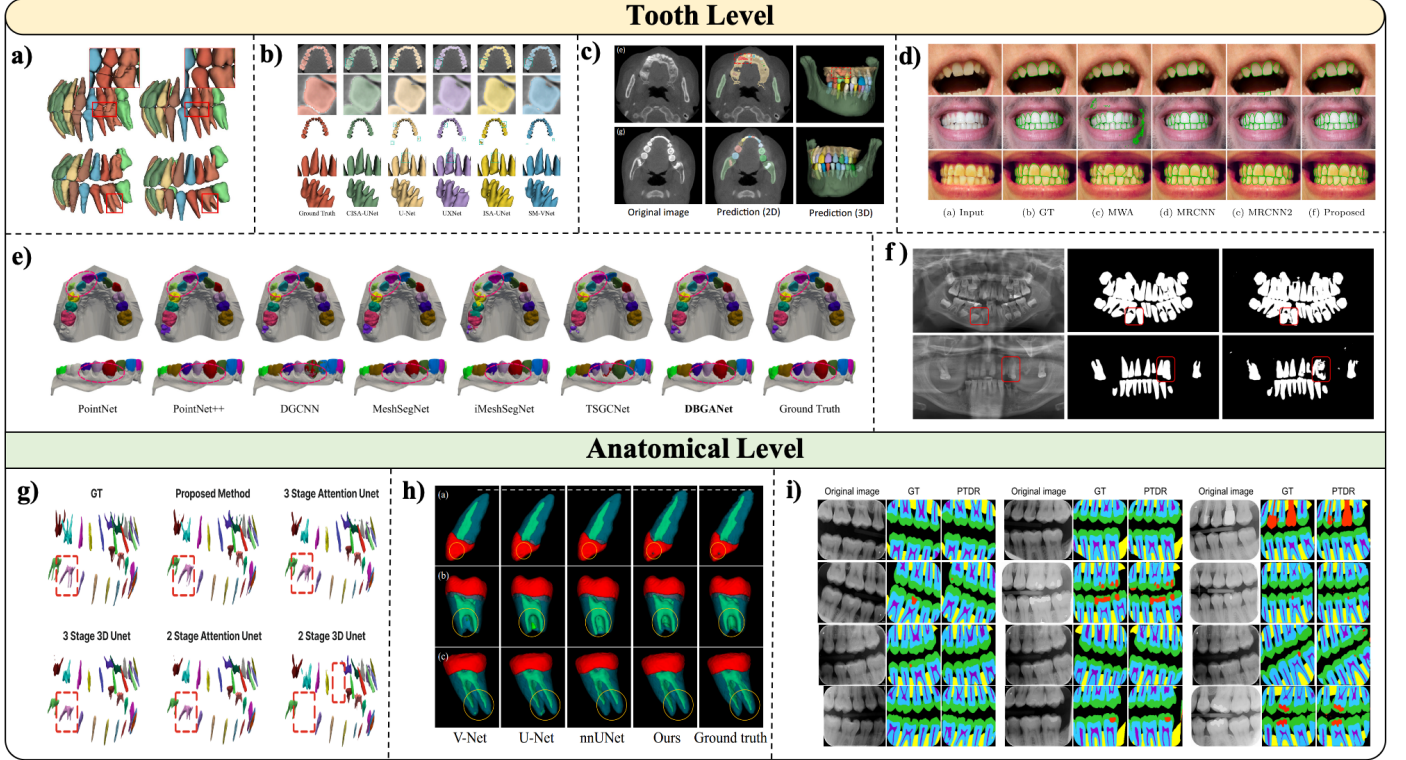


Figure 14: Visualization of dense prediction results for dental analysis. Panels (a–f) display tooth-level dense detection results: (a) CBCT tooth instance segmentation [Cui et al. \(2019\)](#); (b) CBCT tooth semantic segmentation [Lu et al. \(2025\)](#); (c) segmentation of teeth and alveolar bone in CBCT [Cui et al. \(2022c\)](#); (d) tooth semantic segmentation in intraoral photographs [Kim and Lee \(2024\)](#); (e) tooth instance segmentation in IOS [Lin et al. \(2023c\)](#); and (f) tooth semantic segmentation in PAN images [Chen et al. \(2021b\)](#). Panels (g–i) depict anatomical-level dense detection results, as provided by [Kim et al. \(2023\)](#); [Tan et al. \(2023a\)](#); [Rousseau et al. \(2023\)](#), respectively.

**Anatomical Level.** With the continuous advancement of research, the application of DL in DIA has become increasingly widespread. Tooth-level prediction has achieved remarkable results, providing insights into the arrangement of teeth and occlusal relationships. However, such predictions are limited in that they cannot capture the internal structures of teeth. Precise analysis of internal anatomical structures is critical for departments such as endodontics and serves as an essential preoperative reference for dental procedures, particularly root canal treatment. To address this limitation, researchers have extended prediction objectives from tooth-level to anatomical-level analysis, as shown in Fig. 14 (g)–(i). As shown in Table 4, leveraging the powerful feature learning capabilities of DL, a variety of algorithms and models have been proposed for anatomical-level prediction. Some studies have focused on the prediction and analysis of pulp and tooth root structures. Shi et al. [Shi et al. \(2021\)](#) investigated the performance of CNN based on LeNet-5 [LeCun et al. \(2002\)](#) in tooth root detection from PAN images, exploring different image partitioning strategies combined with edge detection operators as inputs. Li et al. developed GTUNet [Li et al. \(2021\)](#), which introduces grouped ViT into the encoder and decoder, enabling the model to preserve multi-scale information while capturing long-range dependencies for accurate root edge detection in PR. Zhang et al.

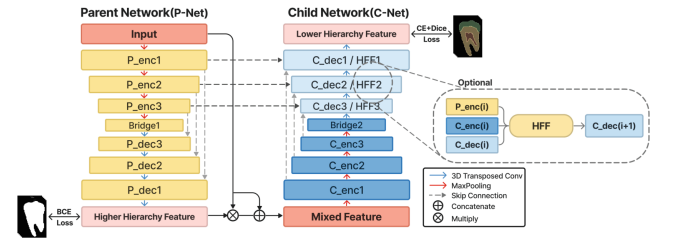


Figure 15: DHU-Net: a three-stage CBCT tooth segmentation method comprising tooth region extraction, tooth classification and patch generation, and fine segmentation of enamel, dentin, and pulp. The HFF module fuses multi-scale features via channel-spatial attention. Image sourced from [Kim et al. \(2023\)](#).

proposed RCS-Net [Zhang et al. \(2021b\)](#), which employs a global encoder and a local decoder built upon 3D U-Net to learn fine details of the pulp region, and integrates the adaptive image enhancement algorithm CLAHE to improve robustness across various CBCT imaging conditions. Wang et al. [Wang et al. \(2023e\)](#) adopted a two-stage framework that first uses a coarse-grained DentalNet to extract tooth-level ROIs from CBCT inputs, followed by 3D U-Net segmentation of the pulp within each ROI.

Beyond the pulp and tooth root, other components such as dentin and enamel have become important targets in anatomical-level dental prediction. Rousseau et al.

Table 2: A summary of tooth-level dense prediction studies based on CNN architectures, arranged in ascending order of publication year, asterisks \* indicate that the study includes a detection task, symbol # indicate that the loss function was proposed by the authors. In the table,  $\times$  indicates that a private dataset was used, while  $\checkmark$  indicates that a public dataset was used. S/I denotes the type of segmentation performed in the study: semantic segmentation (S) or instance segmentation (I). / indicates information not reported in the study, and in the S/I column, it indicates that the study did not include a segmentation task; for studies employing more than three evaluation metrics, we choose three most representative metrics. Tversky loss Salehi et al. (2017); Smo (Smoothness) Xu et al. (2018); AW (Adaptive Wing) Wang et al. (2019); GD (Generalized Dice) Hong et al. (2019); CD (Chamfer Distance); SSIM (Structural Similarity); WPW-CE (Weighted Point-wise Cross Entropy); MS-SSIM (Multi-scale Structural Similarity); RC (Region Consistency); BC (Boundary Consistency); GM (GeodesicMap); DA (Detection Accuracy); FA (Identification Accuracy); ASSD (Average Symmetric Surface Distance); MSSD (Maximum Symmetric Surface Distance); DCD (Directional Cut Discrepancy); B-F1 (Boundary F1 score); RMSE (Root Mean Squared Error); HD (Hausdorff Distance).

| Paper                        | Dataset                              | S/I | Loss                   | Optimizer   | GPU         | Evaluation                                                          | Highlights                                                                                                                                                                                                    |
|------------------------------|--------------------------------------|-----|------------------------|-------------|-------------|---------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Xu et al. (2018)             | 1200 IOS $\times$                    | I   | /                      | /           | GTX 1080    | Mean errors: 0.0848mm                                               | 3D tooth segmentation based on geometric features and boundary optimization.                                                                                                                                  |
| ToothNet Cui et al. (2019)   | 20 CBCT $\times$                     | I   | CE,BCE<br>SL1,MSE      | Adam        | GTX 1080 Ti | DSC: 91.98%<br>DA: 97.75%<br>FA: 92.79%                             | Extended 3D Mask R-CNN with edge enhancement and similarity matrix, leveraging spatial relations for tooth segmentation and recognition.                                                                      |
| Gou et al. (2019)            | 401 CBCT $\times$                    | S   | CE                     | Adam        | GTX 1070    | Accuracy: 66.7%                                                     | Tooth segmentation via level-set guided automatic annotation and improved U-Net.                                                                                                                              |
| Koch et al. (2019)           | 1500 PAN $\checkmark$                | S   | CE<br>BCE<br>Tversky   | NAdam       | /           | DSC: 93.6%                                                          | Patch-based U-Net with batch normalization and test-time augmentation for high-accuracy large-image tooth segmentation.                                                                                       |
| Majanga and Viriri (2019)    | 11,148 BR $\times$                   | S   | /                      | /           | /           | Accuracy: 89%                                                       | Canny edge-assisted features with data augmentation enhance boundary accuracy and generalization.                                                                                                             |
| Zanjani et al. (2019)        | 120 IOS $\times$                     | I   | BCE<br>#WPW-CE         | SGD<br>Adam | /           | IoU: 94%<br>Precision: 93%<br>Recall: 90%                           | End-to-end 3D point cloud segmentation with PointCNN, incorporating non-uniform sampling and adversarial discrimination.                                                                                      |
| Rao et al. (2020)            | 86 CBCT $\times$                     | S   | BCE                    | Adam        | GTX 1070    | DSC: 91.66%<br>ASSD: 0.25mm<br>MSSD: 1.18mm                         | U-Net with deep bottleneck architecture and DCRF post-processing achieves high-accuracy segmentation.                                                                                                         |
| UDS-NET Lee et al. (2020)    | 102 CBCT $\times$                    | S   | BCE                    | Adam        | Titan X     | DSC: 91.8%<br>Precision: 90.4%<br>Recall: 93.2%                     | UDS-Net leverages multi-phase training, Dense Blocks, and Spatial Dropout to enhance feature representation and generalization.                                                                               |
| TSASNet Zhao et al. (2020)   | 1500 PAN $\checkmark$                | S   | BCE<br>SSIM            | Adam        | GTX 1080 Ti | DSC: 92.72%<br>Accuracy: 96.94%<br>Specificity: 97.81%              | TSASNet integrates global and local attention mechanisms for precise boundary extraction in low-contrast panoramic dental images.                                                                             |
| Silva et al. (2020) *        | 1,321 PAN $\checkmark$               | I   | /                      | SGD         | RTX Titan   | PANet:<br>Accuracy: 96.7 $\pm$ 0.1%<br>Specificity: 98.7 $\pm$ 0.2% | Two-stage instance segmentation networks enable accurate tooth detection, segmentation, and numbering in panoramic X-rays with blurred boundaries.                                                            |
| CRIS Van Pham et al. (2020)  | 170600 IP $\times$                   | I   | /                      | /           | /           | F-measure: 81.9%                                                    | CRIS refines individual tooth contours by integrating instance masks with RCF edge features.                                                                                                                  |
| Sivagami et al. (2020)       | 1171 PAN $\checkmark$                | S   | BCE                    | Adam        | /           | DSC: 94%<br>Specificity: 95%<br>F1-Score: 93%                       | UNet-based end-to-end FCN efficiently extracts features from PAN using encoder-decoder structure.                                                                                                             |
| Zhang et al. (2020)          | 120 IOS $\times$                     | I   | Weighted-BCE           | /           | GTX 1080 Ti | DCD: 0.046mm<br>Accuracy: 98.87%                                    | Harmonic mapping with improved FCC algorithm efficiently transforms 3D tooth meshes for 2D processing and boundary refinement.                                                                                |
| Sun et al. (2020b)           | 100 IOS $\times$                     | I   | CE<br>Smo<br>#RC,#BC   | /           | /           | Accuracy: 97%                                                       | FeaStNet-based graph convolutional network performs multi-label tooth segmentation by integrating comprehensive vertex features with explicit anatomical priors to ensure precise vertex-wise correspondence. |
| c-SCN Sun et al. (2020a) *   | 100 IOS $\times$                     | I   | CE,CD<br>Smo<br>#GM    | Adam        | GTX 1080 Ti | DSC(Upper):<br>Molar: 98%<br>Premolar: 98%<br>Canine: 98%           | Use FeaStNet to produce geodesic maps, multi-label classification, and dense correspondence for vertex attribute transfer.                                                                                    |
| Yang et al. (2021)           | 5244 CBCT $\times$                   | S   | CE                     | Adam        | /           | DSC: 97.91%<br>DA: 97.33%<br>B-F1: 98.24%                           | Combining deep CNN with ellipse-shaped priors in a level set framework enables precise tooth boundary refinement.                                                                                             |
| MSLPNet Chen et al. (2021b)  | 1500 PAN $\checkmark$                | S   | Dice<br>BCE<br>MS-SSIM | Adam        | GTX 1080 Ti | DSC: 93.01%<br>Accuracy: 97.30%<br>Specificity: 98.45%              | Employ multi-scale location perception modules to enhance fuzzy root boundary and global feature representation.                                                                                              |
| ToothPix Cui et al. (2021a)  | 1500 PAN $\times$                    | S   | L1<br>BCE              | Adam        | RTX 2070    | IoU: 90.42%<br>Specificity: 98.52%<br>F1-score: 94.86%              | Use a GAN with encoder-decoder residual generator and FCN discriminator to enhance mask realism.                                                                                                              |
| Harsh et al. (2021)          | 1500 PAN $\checkmark$                | S   | Dice                   | Adam        | RTX 2060    | DSC: 94.7%<br>Accuracy: 92.8%<br>F1-score: 88.09%                   | Attention gates in U-Net skip connections retain critical features while suppressing redundant information.                                                                                                   |
| Zhao et al. (2021b)          | 80 IOS $\times$                      | I   | CE                     | Adam        | GTX 1080    | Accuracy: 94.84%<br>mIoU: 87.09%                                    | End-to-end GNN integrates local augmentation, graph attention, and global features for multi-scale 3D tooth mesh representation.                                                                              |
| DLLNet Lang et al. (2021) *  | 77 IOS $\times$                      | /   | GD<br>AW<br>MSE        | Adam        | GTX 1080 Ti | RMSE: 0.386mm                                                       | Design two branches: landmark localization and landmark area segmentation, and integrate the attention mechanism to achieve detection.                                                                        |
| TSegNet Cui et al. (2021b)   | 2000 IOS $\times$                    | I   | L1<br>CD<br>CE,BCE     | Adam        | GTX 1080 Ti | DSC(point): 98.0%<br>DSC(surface): 98.6%<br>F1-score: 94.2%         | Two-stage framework with centroid regression and confidence-aware attention enables instance-level tooth feature extraction and boundary refinement.                                                          |
| Jang et al. (2021) *         | 194 PAN $\times$<br>97 CBCT $\times$ | I   | /                      | Adam        | RTX 2070    | DSC: 94.79%<br>HD: 1.66mm<br>ASSD: 0.14 mm                          | Four-step CBCT method using multi-threshold Otsu and U-Net enables 3D tooth detection, numbering, and ROI extraction.                                                                                         |
| TSGCNet Zhang et al. (2021d) | 80 IOS $\times$                      | S   | CE                     | Adam        | GTX 1080    | mIoU: 88.99%<br>Accuracy: 95.25%                                    | TSGCNet integrates coordinate topology and normal boundaries via dual-stream attention for discriminative geometry.                                                                                           |
| TSGCN Zhao et al. (2021a)    | 80 IOS $\times$                      | I   | /                      | Adam        | GTX 1080    | mIoU: 91.69%<br>Accuracy: 96.96%                                    | An expanded version of the above work. C/N streams employ Input-Transformer and attention fusion to improve geometric alignment and robustness.                                                               |

(continued on next page)

Table 2 (Continued from last page.) Huber Ren et al. (2015); SSE (Sum of Squared Errors); ASMS (Automatically Searched Metric-Sensitive); IV (Intraclass Variance); ID (Intraclass Distance); ML-InfoNCE (Multi-level InfoNCE); PE (Potential Energy); KC (Knowledge Consistency); KL-Divergence (Kullback-Leibler Divergence); CBC (Contrastive Boundary-Constrained); SL1 (Smooth L1); ASD (Average Surface Distance); PPV (Positive Predictive Value); HD95 (95th Percentile Hausdorff Distance); VOE (Volumetric Overlap Error).

| Paper                         | Dataset     | S/I | Loss                      | Optimizer | GPU                     | Evaluation                                                               | Highlights                                                                                                                                           |
|-------------------------------|-------------|-----|---------------------------|-----------|-------------------------|--------------------------------------------------------------------------|------------------------------------------------------------------------------------------------------------------------------------------------------|
| LTPEDN Salih and Duffy (2022) | 1500 PAN ✗  | S   | Dice                      | /         | /                       | DSC: 92.42%<br>Accuracy: 94.32%                                          | LTPEDN achieves efficient DIA via multi-stage feature learning and edge enhancement while reducing computational complexity.                         |
| BDU-net Zhang et al. (2022)   | 916 PAN ✗   | I   | BCE                       | SGD       | RTX 2080 Ti             | mDSC: 92%                                                                | BDU-net enhances dental recognition and segmentation by multi-scale DFE fusion and dual boundary-region subnetworks.                                 |
| Li et al. (2022b)             | 350 CBCT ✗  | I   | CE<br>Dice                | Adam      | Tesla V100              | DSC: 91.13±0.45%<br>Precision: 92.13±0.29%<br>HD: 1.00±0.27mm            | GCN-based attention mechanism with global context improves spatial modeling and boundary recognition.                                                |
| LeFUNet Rao et al. (2022)*    | 1616 PAN ✓  | S   | /                         | Adam      | RTX 2080                | DSC: 96.94±0.74%<br>Precision: 96.81±0.23%<br>Specificity: 99.61±.83%    | LeFUNet leverages multi-stage preprocessing and SE modules to enhance feature reuse for panoramic tooth detection.                                   |
| Cui et al. (2022c)            | 4938 CBCT ✗ | I   | BCE<br>L2<br>CE           | Adam      | Tesla V100              | DSC: 94.1%<br>Sensitivity: 93.9%<br>ASD: 0.17mm                          | Three-stage V-Net framework: ROI localization, centroid and skeleton-guided instance refinement, and Haar-filtered cascaded V-Net for alveolar bone. |
| Chandrashekar et al. (2022)*  | 1500 PAN ✓  | I   | CE<br>BCE,SL1<br>IOU,SSE  | SGD       | /                       | Accuracy: 98.44%<br>F1-Score: 98.75%<br>mAP: 97.78%                      | A collaborative deep learning-based tooth detection and segmentation model, enabling precise segmentation and recognition on PAN.                    |
| Xie et al. (2022)             | 78 CBCT ✗   | I   | CE                        | /         | RTX 2080 Ti             | DSC: 88±3%<br>Precision: 98±3%<br>ASSD: 0.53±0.34mm                      | Multi-task U-Net integrates CSA modules to enhance multi-scale features for pixel-wise foreground and landmark prediction.                           |
| Aseretto and Noguera (2022)   | 1500 PAN ✓  | S   | CE                        | Adam      | GTX 1070                | F1-score: 91%<br>Precision: 92%                                          | U-Net-based FCN uses a ResNet18 encoder with transfer learning to improve performance.                                                               |
| Rundo et al. (2022)           | 52 CBCT ✗   | S   | Dice                      | Adam      | RTX 3090                | DSC: 66.2%<br>Specificity: 89.9%<br>F1-score: 65.6%                      | A 3D convolutional encoder-decoder with deep supervision and skip connections enables efficient dental arch segmentation.                            |
| Shibuya et al. (2022)         | 119 PAN ✗   | S   | CE                        | Adam      | /                       | mIoU: 67.1%                                                              | Based on DeepLab v3+, combining ASPP and decoder for segmentation, with Alpha Blending enhancing training diversity.                                 |
| IGNet Liu et al. (2022a)      | 210 IOS ✗   | S   | CE<br>*ASMS               | SGD       | Tesla V100              | mIoU: 81.25%<br>mAcc: 96.56%                                             | IGNet fuses heterogeneous coordinate and normal graphs via cross-attention and uses metric-driven automated loss search.                             |
| Caylak et al. (2022)          | 131 PAN ✗   | S   | /                         | /         | MX150                   | DSC(best): 90.26%                                                        | Exploring the performance of different UNet architectures in DIA after transfer learning with various pretrained weights.                            |
| Tian et al. (2022)*           | 210 IOS ✗   | I   | CE<br>L2<br>IV,ID         | /         | Titan X                 | mAP50%: 55.0%<br>mAP25%: 72.4%                                           | A 3D U-Net backbone with SSC jointly performs point-wise semantic segmentation and centroid prediction for improved tooth modeling.                  |
| Boudraa (2022)                | 85 IOS ✗    | I   | /                         | /         | Tesla K80               | DSC: 98.04%<br>Sensitivity: 97.99%<br>PPV: 98.19%                        | An improved MeshSegNet employs mesh-based multi-scale feature learning with 15D inputs for efficient 3D tooth segmentation.                          |
| Wu et al. (2022)*             | 136 IOS ✗   | I   | Dice<br>MSE               | /         | RTX 2080 Ti             | Sensitivity: 94.6±5.4%<br>DSC: 94.3±4.4%<br>HD: 2.420±3.140 mm           | A two-stage approach that leverages segmentation as the foundation to automatically identify anatomical landmarks in IOS.                            |
| DArch Qiu et al. (2022)       | 4773 IOS ✗  | I   | CE<br>Huber               | /         | RTX 3090                | DSC: 97.7%<br>Accuracy: 99.68%<br>Recall: 85.39%                         | A coarse-to-fine dental arch prediction using Bézier curves and GCN is proposed, with APS sampling improving detection accuracy.                     |
| Li et al. (2022d)             | 180 IOS ✗   | I   | L2<br>Dice                | /         | /                       | DSC: 95.68%<br>Sensitivity: 95.73%<br>PPV: 95.83%                        | A BEM block is proposed to resolve local context ambiguity by offset learning, integrating multi-scale contexts for accurate tooth segmentation.     |
| STSNNet Liu et al. (2022b)    | 13000 IOS ✗ | I   | CE<br>*ML-InfoNCE         | SGD       | /                       | DSC: 94.85%<br>mIoU: 93.12%<br>Accuracy: 97.22%                          | STSNNet leverages self-supervised contrastive pre-training and fine-tuning with limited labeled data to improve 3D tooth segmentation.               |
| Cao et al. (2023)             | 148 CBCT ✗  | I   | CE<br>*PE<br>Dice         | Adam      | RTX 3090                | DSC: 96.89%<br>HD95: 0.54mm<br>ASD: 0.11mm                               | Three-stage instance segmentation with spectral filtering and dilated convolutions improves tooth boundary accuracy.                                 |
| Jana et al. (2023b)           | 50 IOS ✗    | I   | /                         | /         | RTX 8000                | DSC(1st.): 92.88%<br>Accuracy(1st.): 94.12%<br>Sensitivity(1st.): 93.89% | The study evaluates general and dental-specific point cloud segmentation methods on mesh datasets.                                                   |
| KCNet Lin et al. (2023b)      | 1321 PAN ✓  | S   | CE<br>KC<br>KL-Divergence | /         | /                       | DSC: 89±3.8%<br>HD: 5.617±0.756mm<br>VOE: 0.22±0.153mm                   | Proposed lightweight KCNet uses knowledge distillation to maintain high accuracy while reducing computational cost.                                  |
| Gan et al. (2023)             | 750 CBCT ✗  | S   | CE<br>Dice                | /         | /                       | /                                                                        | Combining U-Net segmentation with Noise2Sim denoising enables region-adaptive image enhancement.                                                     |
| Tan and Xiang (2023)          | 103 IOS ✗   | I   | BCE<br>CBC<br>DICE        | /         | /                       | mIoU: 83.77%<br>Accuracy: 90.98%                                         | Graph neural network with boundary-constrained loss improves tooth and gingiva segmentation accuracy.                                                |
| BFFNet Zhang et al. (2023a)   | 106 PAN ✓   | S   | IOU<br>BCE                | Adam      | RTX 2080 Super          | DSC: 79.11%<br>HD: 0.0174mm                                              | BFFNet enhances tooth segmentation by fusing boundary features with global contextual information.                                                   |
| WMC-Net Li et al. (2023b)     | 312 CBCT ✓  | S   | CE<br>MSE<br>Dice         | SGD       | RTX 3090 Ti             | DSC: 76.98±5.59%                                                         | The weakly-consistent network integrates CBAM and EG-MT to improve tooth boundary recognition.                                                       |
| Leclercq et al. (2023)        | 78 IOS ✗    | I   | CE<br>Dice                | Adam      | Titan V                 | DSC: 96±3%<br>Sensitivity: 98±11%<br>PPV: 98±3%                          | PyTorch3D-based 2D rendering generates multi-view RGB normal maps and ground-truth labels.                                                           |
| Feng (2023)                   | 312 CBCT ✓  | S   | BCE<br>Dice               | /         | RTX 4090<br>RTX 4070 Ti | Dice: 77.5%<br>HD: 0.269mm                                               | A 3D U-Net with group normalization was employed to optimize small-batch training.                                                                   |

(continued on next page)

Table 2 (Continued from last page). FPR (False Positive Rate); NSD (Normalized Surface Dice); Lbl-T (Teeth Labeling Accuracy)

| Paper                             | Dataset                  | S/I | Loss                               | Optimizer | GPU                 | Evaluation                                                               | Highlights                                                                                                                                                                                    |
|-----------------------------------|--------------------------|-----|------------------------------------|-----------|---------------------|--------------------------------------------------------------------------|-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Zhang and Zhu (2023)              | 3000 CBCT ✓              | S   | CE<br>Dice<br>KL-Divergence        | SGD       | RTX 3090            | DSC: 72.61%<br>HD: 0.2595mm                                              | Morphology-driven preprocessing combined with dynamic convolution enables multi-head deep supervision.                                                                                        |
| Wang et al. (2023a)               | 362 CBCT ✓               | S   | /                                  | AdamW     | Tesla V100          | DSC: 80.58%<br>HD: 0.1599mm                                              | A semi-supervised domain-adaptive framework with FTA and adaptive thresholds enhances performance.                                                                                            |
| Zhao et al. (2023)                | 280 PAN ✗                | I   | CE<br>BCE<br>SL1                   | /         | Tesla K80           | mPrecision: 86.32%<br>mRecall: 90.05%<br>mF1-score: 87.30%               | Implemented tooth instance segmentation using Mask R-CNN on a self-constructed dataset.                                                                                                       |
| Teeth U-Net Hou et al. (2023)     | 1500 PAN ✗               | S   | CE                                 | RMSprop   | Titan V             | DSC: 94.28%<br>Accuracy: 98.53%<br>Precision: 95.62%                     | The model integrates multiscale and self-attention modules with dense skip connections for encoder-decoder tooth feature fusion.                                                              |
| Jana et al. (2023a)               | 589 IOS ✓                | I   | /                                  | /         | RTX 8000            | DSC(Whole Jaw): 1st.:91.05%                                              | Conducted a comparative analysis of several point cloud segmentation methods on a self-constructed dataset.                                                                                   |
| Duan and Chen (2023)              | 200 IOS ✗                | I   | CE                                 | /         | RTX 3090            | mIoU: 92.1%<br>Accuracy: 96.97%                                          | Decouple coordinate and normal features, leveraging semantic-guided graph transformation and cross-domain attention for segmentation.                                                         |
| DBGANet Lin et al. (2023c)        | 637 IOS ✓                | I   | CE<br>BCE<br>SL1                   | Adam      | RTX 3090            | Teeth3DS:<br>DSC: 94.48±0.83%<br>Sensitivity: 96.71±0.66%                | DBGANet employs voxel and surface branches with geometric attention for boundary-enhanced bi-directional feature fusion in 3D segmentation.                                                   |
| MRCM-UCTransNet Wen et al. (2024) | 30 CBCT ✗                | S   | Focal                              | /         | Tesla V100          | DSC: 90.612%<br>mIoU: 82.836%<br>Sensitivity: 93.855%                    | Enhance the U-Net backbone by integrating multi-scale residual CNN and a CTrans block for effective feature extraction and channel-wise attention fusion.                                     |
| Beser et al. (2024) *             | 3854 PAN ✗               | I   | IOU                                | SGD       | /                   | Precision: 99 %<br>Recall: 99%<br>mAP: 98%                               | Based on YOLOv5, a deep learning model was proposed for tooth detection in children during the mixed dentition period.                                                                        |
| Zannah et al. (2024)              | 389 PAN ✗                | S   | Dice                               | Adam      | RTX 3080 Ti         | DSC(L3):<br>1st.: 90.35%<br>2nd.: 90.33%                                 | Compared U-Net and its variants (Attention, Residual, U-Net++) to assess module contributions to dental segmentation performance.                                                             |
| USCT Jing et al. (2024)           | 66 CBCT ✓<br>19 CBCT ✗   | S   | MSE<br>Dice                        | SGD       | Titan XP            | DSC: 90.76%<br>HD95: 6.29mm<br>ASD: 1.76mm                               | USCT integrates symmetric consistency learning with uncertainty regularization, employing ACIM and cross-task constraints for robust semi-supervised segmentation.                            |
| CRML-Net Lyu et al. (2024)        | 129 CBCT ✓               | S   | Focal<br>BCE, Dice<br>Huber        | Adam      | RTX 4090            | DSC: 92.43%<br>HD: 1.21mm<br>ASD: 0.31mm                                 | CRML-Net skeleton fused with CBCT for features, multi-task branches jointly predict tooth masks and boundaries.                                                                               |
| Lv et al. (2024)                  | 20 CBCT ✗                | I   | L1<br>BCE<br>Dice                  | /         | /                   | DSC: 83.26%<br>Precision: 93.82%<br>HD: 2.44mm                           | Three-stage network: ROI detection, skeleton extraction, and skeleton-guided V-Net for precise single-tooth segmentation with multi-scale fusion.                                             |
| Brahmi and Jdey (2024) *          | 180 PAN ✗                | I   | /                                  | SGD       | Tesla T4            | mAP: 90%<br>Precision: 96%<br>F1-score: 63%                              | Mask R-CNN framework with ResNet101-FPN backbone for simultaneous tooth instance segmentation and identification.                                                                             |
| Kim and Lee (2024)                | 46 IP ✓                  | I   | BCE                                | Adam      | /                   | mF1-score:<br>T01:96.49%<br>T02: 97.30%                                  | ResNeSt-50 generates pseudo-edges; dual-level set active contours with tooth enhancement map refine accurate tooth segmentation.                                                              |
| DFTNet Long and Yuan (2024)       | / CBCT ✓                 | I   | CE<br>Dice                         | Adam      | /                   | DSC: 97.15%<br>Recall: 96.95%<br>Precision: 97.36%                       | V-Net with DFF performs voxel segmentation; Mask R-CNN detects instances; fused results yield precise tooth instance masks.                                                                   |
| DAE-Net Chen et al. (2024b) *     | 600 PAN ✓                | I   | IOU<br>Dice                        | Adam      | RTX 4060 Ti         | IoU: 27.0%<br>NSD: 59.5%<br>IA: 26.7%                                    | DAE-Net uses FPN and RPN with RFEM and CBAM to generate tooth instance masks and refine segmentation.                                                                                         |
| Ji et al. (2024)                  | 2400 PAN ✓               | I   | /                                  | /         | RTX 4090            | DSC(instance): 79.82%<br>NSD(instance): 84.14%<br>mIoU(instance): 75.66% | A two-stage nnU-Net framework with semi-supervised self-training performs quadrant-wise fine tooth segmentation numbering.                                                                    |
| Isensee et al. (2024)             | / CBCT ✓                 | I   | /                                  | /         | GH200<br>Tesla A100 | DSC: 92.53%<br>HD95: 18.472mm                                            | Deepened nnU-Net ResEnc L with seven residual encoder stages, optimized training and post-processing for accurate tooth segmentation.                                                         |
| UNet 3+ Sahin et al. (2024)       | 2793 PAN ✓               | S   | Dice                               | /         | RTX 3090 Ti         | DSC: 93.31%<br>Accuracy: 97.36%<br>Precision: 93.25%                     | U-Net 3+ adds multi-scale skip connections and feature fusion to improve tooth structure segmentation.                                                                                        |
| Qin et al. (2024)                 | 175 CBCT ✗<br>148 CBCT ✓ | S   | Dice                               | /         | /                   | DSC:<br>Pub: 92%<br>Cli: 92.05%                                          | 3D U-Net adds inter-layer attention, residual blocks, and attention gates to enhance boundary segmentation and multi-scale feature fusion.                                                    |
| CAMS Osama et al. (2024)          | 693 PAN ✓                | S   | /                                  | /         | /                   | mDSC: 90.41%<br>mIoU: 83.18%<br>F1-score: 90.41%                         | Three U-Net models: Specialized per quadrant, All-In-One multi-class, CAMS uses cue mask for conditional segmentation.                                                                        |
| Chen et al. (2024c)               | 350 CBCT ✓               | I   | CE<br>Dice                         | AdamW     | Tesla V100          | DSC <sub>A</sub> : 87.7%<br>NSD <sub>A</sub> : 88.7%                     | The method employs a two-stage semi-supervised framework: ROI extraction via 3D-VNet followed by mask-assisted instance segmentation.                                                         |
| RAGCNet Zhao et al. (2024c)       | 180 IOS ✓                | I   | L1<br>CE                           | Adam      | RTX 3090            | Accuracy: 95.30±0.11 %<br>mIoU: 89.20±0.25%<br>mAccuracy: 90.02±0.17%    | RAGCNet leverages a Group module and a Region-Aware Module for feature learning, achieving accurate tooth segmentation through dual-branch feature fusion and a centroid-based loss function. |
| THISNet Li et al. (2023c)         | 1200 IOS ✓               | I   | CE<br>L2<br>Dice<br>Weighted Focal | Adam      | Tesla V100          | Accuracy: 94.69±0.53%<br>mIoU: 87.05±1.68%                               | THISNet leverages region highlighting and affinity modules for tooth instance segmentation, with multi-task learning improving classification and segmentation.                               |
| CurSegNet Mao et al. (2024)       | 30 IOS ✗                 | I   | CE                                 | /         | GTX 2080            | DSC: 95.73%<br>Sensitivity: 95.20%<br>PPV: 96.39%                        | Dual-branch CurSegNet integrates GAM, CurveNet aggregation, and cross-layer fusion for 3D tooth mesh segmentation.                                                                            |
| WS-TIS Wang et al. (2024b)        | 1200 IOS ✓               | I   | CE<br>Focal                        | /         | RTX 4090            | DSC: 97.37±0.48%<br>mAP: 96.37±0.74%<br>FPR: 17.50±1.19%                 | Disentangled re-sampling and gated-attention highlight classes, Grad-CAM localizes teeth, DGCNN models fine contours.                                                                         |

(continued on next page)

Table 2 (Continued from last page). MCC (Matthew’s Correlation Coefficient); MAE (Mean Absolute Error).

| Paper                                             | Dataset                  | S/I | Loss                                | Optimizer | GPU                      | Evaluation                                                           | Highlights                                                                                                                                                                                                               |
|---------------------------------------------------|--------------------------|-----|-------------------------------------|-----------|--------------------------|----------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| HiCA<br>Li et al. (2024c)                         | 1200 IOS ✓<br>76 IOS ✗   | I   | CE                                  | Adam      | GTX 3090                 | Teeth3DS:<br>DSC: 93.44±8.60%<br>Accuracy: 94.92±5.47%               | HiCA integrates cross-stream contextual and discriminative modules for multistream feature fusion and robust missing-tooth segmentation.                                                                                 |
| TSG-GCN<br>Liu et al. (2024b)                     | 710 CBCT ✗               | I   | CE<br>MSE<br>Dice                   | Adam      | RTX 3090                 | DSC:<br>Child: 94.50±4.16%<br>Adult: 94.10±1.64%                     | TSG-GCN integrates 2D projection and 3D GCN, dynamically modeling teeth spatial relationships.                                                                                                                           |
| Zhuang et al.<br>(2025)                           | 1200 IOS ✓               | I   | CE<br>CD<br>SL1, $L_{sep}$<br>*Diff | /         | /                        | DSC: 97.63%<br>Lbl-T: 95.62%                                         | Robust tooth segmentation by fusing semantic segmentation and instance segmentation, followed by a global optimization and post-processing step for accurate final labeling.                                             |
| CMAT Chen et al. (2025b)                          | 5138 IOS ✗               | I   | CE<br>Entropy<br>KL-Divergence      | AdamW     | RTX 3090                 | mIoU:<br>CrossTooth : 82.19%<br>AbnTooth: 87.32%                     | Source prototype alignment with progressive pseudo-labels and tooth-ratio priors enables cross-domain tooth segmentation adaptation.                                                                                     |
| Arian et al. (2025)                               | 40 IOS ✗                 | S   | /                                   | /         | /                        | mIoU: 95%                                                            | Unsupervised 3D dental arch segmentation with PointNet++, using GRL and Siamese for domain adaptation.                                                                                                                   |
| Ghafoor and Lee (2025)                            | 540 PAN ✗                | I   | Dice                                | /         | RTX 3090                 | DSC: 91.72%<br>Precision: 80.44%<br>Specificity: 97.31%              | U-Net with global and teeth attention enhances boundary and salient feature representation.                                                                                                                              |
| Yaswanth et al. (2025)                            | 1000 PAN ✓               | S   | BCE                                 | Adam      | /                        | DSC: 90.7%<br>Accuracy: 97.8%<br>mAP: 96.4%                          | U-Net++ with ECA attention enhances multi-scale feature fusion and key channel representation.                                                                                                                           |
| MSCA-UNet<br>Pan et al. (2025a)                   | 2500 PAN ✓               | S   | CE<br>Dice                          | Adam      | RTX 4090                 | DSC: 92.37±0.01%                                                     | U-Net integrates adaptive convolution and multi-scale learning for global context and high-resolution detail.                                                                                                            |
| CISA-UNet<br>Lu et al. (2025)                     | 27474 CBCT ✓             | S   | DICE                                | Adam      | RTX 3090 Ti              | DSC: 94.10±1.93%<br>MCC: 94.13±1.90%<br>Accuracy: 99.94±0.020%       | CISA-UNet, built upon the 3D UNet backbone, integrates edge maps and CLAHE-enhanced images. Incorporating the CISA module for refined feature extraction, it achieves high-precision tooth segmentation from CBCT scans. |
| SMTLNet<br>Zhao et al. (2025a)                    | 1305 CBCT ✗<br>98 CBCT ✓ | I   | BCE<br>Dice<br>*Intraclass          | Adam      | RTX Titan<br>RTX 3090 Ti | DSC: 91.84±1.67%<br>Precision: 94.47±0.02%<br>HD: 1.41±0.82mm        | SMTLNet leverages self-supervised pretraining to learn 3D features, followed by coarse and fine segmentation to refine boundaries and overall results.                                                                   |
| Ma et al. (2025)                                  | 1020 PAN ✗               | I   | Dice                                | /         | RTX 3090                 | DSC: 92.0%<br>Precision: 92.3%<br>Recall: 92.5%                      | The method consists of tooth and alveolar bone segmentation, both based on nnU-Net. Optimization combines pixel aggregation and CVAE to improve completeness and anatomical plausibility.                                |
| Zhou et al. (2025a)                               | 280 IOS ✓                | I   | CE<br>GD                            | Adam      | RTX 4090                 | Accuracy: 95.68±0.12%<br>mIoU: 90.32±0.41%<br>mAccuracy: 91.41±0.59% | The network extracts tooth features via coordinate and normal vector branches and fuses them to produce final semantic labels.                                                                                           |
| GCNet Zhong et al. (2025)                         | 3187 PAN ✓               | S   | BCE<br>IOU                          | Adam      | RTX 4090                 | DSC: 93.38%<br>Recall: 94.26%<br>MAE: 0.0259mm                       | Enhanced UNet, GGA+CLF fuses multi-level features, DOD decodes edges and segmentation                                                                                                                                    |
| PXseg Wang et al. (2025d)                         | 420 PAN ✗<br>258 CBCT ✗  | I   | CE<br>BCE<br>GIOU                   | Adam      | RTX 3090                 | DSC: 88.2%<br>Accuracy: 95.7%<br>F1-score: 90.2%                     | PXseg based on BlendMask, mask branch and blender module perform tooth instance segmentation and abnormal detection                                                                                                      |
| TSegLab<br>Rekik et al. (2025)                    | 1800 IOS ✗               | I   | /                                   | SGD       | Titan V                  | DSC: 95.10%<br>Accuracy: 94.30%<br>F1-score: 98%                     | Coarse-to-fine pipeline: Mask R-CNN for detection/segmentation, GNN for tooth recognition and labeling.                                                                                                                  |
| Seg-CSAB<br>Bajpai et al. (2025)                  | 523 PAN ✓<br>802 PAN ✗   | I   | /                                   | SGD       | RTX 3090                 | DSC: 92.39%<br>Accuracy: 99.90%<br>Specificity: 99.95%               | ResNet50+FPN two-stage segmentation, CSAB enhances features, Deeper Mask Head improves tooth mask precision.                                                                                                             |
| ToothInstanceNet<br>Van Nistelrooij et al. (2025) | 716 IOS ✓                | I   | /                                   | /         | /                        | 3DTeethSeg22:<br>DSC: 98.24%<br>F1-score: 99.64%                     | ToothInstanceNet performs two-stage point cloud segmentation with Gaussian-based instance generation and MLP labeling for precise tooth segmentation.                                                                    |
| Chen et al. (2025c)                               | 39 CBCT ✗                | I   | GD<br>Focal                         | /         | RTX 3060                 | DSC: 96.33±0.89%<br>HD: 2.04±0.87<br>ASD: 0.24±0.06                  | Multi-model Attention U-Net and V-Net segmentation is fused to correct tooth positions and measure working length and canal angles.                                                                                      |

(End of Tabel. 2.)

Table 3: A summary of tooth-level dense prediction studies based on ViT and hybrid architectures, arranged in ascending order of publication year. Asterisks (\*) indicate that the study involves a detection task, while the symbol (✖) denotes that the loss function was proposed by the authors. In the table, ✖ represents the use of a private dataset, and ✓ represents the use of a public dataset. S/I denotes the segmentation type: semantic segmentation (S) or instance segmentation (I). “/” indicates information not reported in the study; in the S/I column, it denotes that the study did not include a segmentation task. For studies employing more than three evaluation metrics, the three most representative metrics are selected. AUC (Area Under the Curve); IA (Identification Accuracy); GG (Geometry-Guided); CBL (Contrastive Boundary Learning) Tang et al. (2022); AR (Annotator-Robust) Liu et al. (2017).

| Type         | Paper                          | Dataset                                               | S/I | Loss               | Optimizer | GPU             | Evaluation                                                             | Highlights                                                                                                                                                                |
|--------------|--------------------------------|-------------------------------------------------------|-----|--------------------|-----------|-----------------|------------------------------------------------------------------------|---------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Pure ViT     | TSegFormer Xiong et al. (2023) | 16000 IOS ✖                                           | I   | CE<br>✖GG          | /         | /               | DSC: 96.01%<br>mIoU: 94.34%<br>Accuracy: 97.97%                        | TSegFormer encodes point cloud geometry, employing multi-layer self-attention and auxiliary heads for fine tooth-gum boundary modeling.                                   |
|              | Sheng et al. (2023)            | 100 PAN ✖                                             | S   | BCE                | Adam      | /               | IoU:<br>PLAGH-BH: 46.89%<br>PDX: 69.56%                                | Benchmarked U-Net, LinkNet, FPN, and Swin-Unet for tooth segmentation.                                                                                                    |
|              | Dai et al. (2024)              | 158 CBCT ✖                                            | S   | Tversky            | Adam      | RTX A6000       | DSC: 83.84±4.07%<br>Precision: 86.95±3.08%<br>HD: 8.73±1.78mm          | Self-supervised MAE with graph-attention sparse boundary prompts enables CBCT tooth segmentation, improving performance with limited annotations.                         |
|              | Almalki and Latecki (2024)     | 1800 IOS ✓                                            | S   | L2<br>CD<br>MSE    | AdamW     | /               | Accuracy: 98.3%<br>DSC: 97.0%<br>PPV: 98.9%                            | DentalMAE employs a self-supervised Transformer to reconstruct masked mesh embeddings, learning transferable representations for enhanced 3D tooth segmentation.          |
|              | Tooth-ASAM Wang et al. (2025b) | 4938 CBCT ✓<br>200 CBCT ✖<br>2500 PAN ✓<br>30029 IP ✓ | S   | BCE<br>Dice        | Adam      | RTX 4090        | NC:<br>DSC: 95.5%<br>HD95: 0.234mm<br>ASSD: 0.023mm                    | Tooth-ASAM uses an image encoder to extract feature embeddings and combines them with the prompt encoder to produce masks.                                                |
| Hybrid Model | Ghafoor et al. (2023)          | 540 PAN ✖                                             | I   | Square Dice        | /         | RTX 3090        | DSC: 91.02%<br>Accuracy: 97.26%<br>Specificity: 97.30%                 | UNet with Swin-Transformer integrates TAB, using self-attention to enhance tooth boundary features.                                                                       |
|              | Zhang et al. (2023b)           | 100 CBCT ✖                                            | I   | L1<br>CE<br>Dice   | Adam      | RTX 3090        | DSC: 95.1±0.3%<br>ASD: 0.13±0.02mm<br>HD: 1.39±0.24mm                  | UNETR++ with EPA modules enables 3D tooth ROI segmentation and centroid prediction.                                                                                       |
|              | Almalki and Latecki (2023) *   | 543 PAN ✖                                             | I   | L1<br>MSE          | AdamW     | Tesla V100      | AP <sub>box</sub> : 86.1%<br>AP <sub>mask</sub> : 84.6%                | SimMIM self-supervised pretraining of Swin Transformer improves dental detection and instance segmentation.                                                               |
|              | Mohan et al. (2023) *          | 1500 PAN ✓                                            | S   | /                  | /         | /               | DSC:<br>1st.: 91.37±0.38%<br>2nd.: 89.69±1.07%<br>3rd.: 89.04±0.68     | Enhanced U-Net integrates Swin Transformer bottleneck, Teeth Attention Block, and multiscale supervision                                                                  |
|              | WCTN Ahn and Cho (2024)        | 1587 IOS ✖                                            | S   | CE                 | Adam      | RTX 3090        | mIoU: 88.71%<br>Accuracy: 94.01%<br>mAccuracy: 93.72%                  | WCTN integrates weighted sparse CNN and Transformer aggregation, employing group attention for efficient local-global feature fusion.                                     |
|              | MPRP-LAGC Wang et al. (2024c)  | 2692 CBCT ✓                                           | S   | MSE<br>BCE<br>Dice | AdamW     | /               | DSC: 91.58%<br>AUC: 92.62%<br>HD: 7.386mm                              | MPRP-LAGC integrates dual-branch pretraining and graph convolution for precise tooth segmentation.                                                                        |
|              | Trans-VNet Wang et al. (2024a) | 150 CBCT ✓<br>42 CBCT ✖                               | S   | /                  | /         | RTX 3090        | DSC:<br>MAD: 88.67%<br>Cui: 96.44%                                     | Trans-VNet integrates a Transformer into V-Net's bottleneck, fusing global context with local features under deep supervision.                                            |
|              | FlowgateUNet Cao et al. (2024) | 15748 CBCT ✖                                          | S   | CE<br>Dice         | SGD       | Tesla P40       | microCT dataset:<br>DSC: 83.86%<br>HD95: 20.33mm                       | FlowgateUNet integrates FlowFormer and gated attention within U-Net for multi-scale long-range spatial dependency modeling.                                               |
|              | Li et al. (2024d) *            | 5634 PAN ✓                                            | I   | L1<br>CE<br>Focal  | Adam      | RTX 3090 TI     | O <sup>2</sup> PR:<br>mAP: 82.4%<br>Accuracy: 99.9%<br>F1-score: 94.8% | U-Net dual-path: Transformer encodes local, CNN extracts global, fused via cross-attention for precise tooth instance segmentation.                                       |
|              | Guo et al. (2024)              | 2380 PAN ✓                                            | I   | CE<br>Dice         | Adam      | Tesla V100-SXM2 | DSC: 45.58%<br>NSD: 52.27%<br>IA: 37.73%                               | ResUnet50 with SAM-Med2D leverages semi-supervised learning and data augmentation for enhanced tooth segmentation accuracy.                                               |
|              | Wang and Bao (2024)            | 2403 PAN ✓                                            | I   | CE<br>Dice         | SGD       | RTX 4090        | DSC: 87.07%<br>NSD: 41.08%<br>IA: 4.78%                                | A semi-supervised teeth segmentation method combines U-Net and Transformer with cross-teaching, jointly training on labeled and unlabeled data.                           |
|              | FDNet Feng et al. (2024)       | 9000 CBCT ✖                                           | S   | /                  | /         | /               | DSC: 85.28±9.78%<br>Precision: 82.67±12.47%<br>Recall: 89.77±10.99%    | FDNet integrates LF-Wavelet and SAM encoder to enhance low-frequency semantics and refine boundaries, achieving accurate tooth segmentation.                              |
|              | DICL Cai et al. (2024)         | 2409 PAN ✖                                            | S   | CE<br>MSE<br>Dice  | SGD       | RTX 3060        | DSC: 23.85%<br>NSD: 66.53%<br>IA: 60.47%                               | Two-phase tooth segmentation: semi-supervised pseudo-labels, fully supervised refinement with deformable convolutions and consistency loss.                               |
|              | CrossTooth Xi et al. (2025)    | 1800 IOS ✓                                            | S   | CE<br>CBL          | Adam      | RTX 3090        | mIoU: 95.86%                                                           | CrossTooth fuses point cloud and multi-view image features with selective downsampling and boundary-aware learning.                                                       |
|              | EASAM Zhang et al. (2025a)     | 5273 PAN ✓                                            | S   | IOU<br>BCE         | Adam      | RTX 3090        | DSC: 89.03±4.96%<br>Specificity: 98.93±0.16%<br>HD: 40.84±13.27mm      | EASAM integrates ViT and edge branches, using edge-aware and cross-branch interaction for medical imaging.                                                                |
|              | LSAFormer Wang and Wang (2025) | 1800 IOS ✓                                            | I   | CE                 | Adam      | /               | mIoU: 90.5%<br>F1-score: 93.7%                                         | LSAFormer integrates local surface reconstruction and global feature perception for multi-scale point cloud segmentation.                                                 |
|              | Wathore and Gorthi (2025)      | 3909 PAN ✓                                            | S   | CE<br>Dice         | /         | /               | DSC:<br>UNet: 66.96%<br>SE-UNet: 69.89%<br>TransUNet: 76.69%           | Leverages left-right symmetry of panoramic X-rays to create flipped images, augmenting data with consistent tooth labeling.                                               |
|              | DenUnet Manzari et al. (2024)  | 543 PAN ✓                                             | I   | AR<br>BCE<br>Dice  | SGD       | RTX Titan       | mAP: 71.32%<br>AP50: 97.81%<br>AP75: 88.18%                            | DenUnet is a dual-encoder network (PDC Edge + Swin Transformer) with Local Cross-Attention Fusion for dental segmentation.                                                |
|              | UCL-Net Xu et al. (2025a) *    | 6204 PAN ✓                                            | I   | CE                 | AdamW     | RTX 3090        | Recall:<br>TDD: 58.7%<br>DentalX: 61.9%<br>CDD: 66.3%                  | It leverages an encoder-decoder framework with cross-scale attention to perform tooth instance segmentation                                                               |
|              | TSNEN Liu et al. (2025c)       | 200 IOS ✓                                             | I   | Dice               | Adam      | RTX 3090        | Accuracy: 96.83%<br>mIoU: 92.06%                                       | TSNEN dual-stream network fuses coordinate and geometric features, using attention and pooling for 3D tooth mesh segmentation.                                            |
|              | MDGFormer Pan et al. (2025b)   | 805 IP ✖                                              | S   | BCE<br>IOU<br>MSE  | Adam      | RTX 4090        | F-measure(BS): 97.7 %                                                  | MDGFormer achieves tooth segmentation via multi-domain feature extraction and interaction, while incorporating a super-resolution task to recover high-frequency details. |

Table 4: A summary of tooth anatomical-level segmentation studies, arranged in ascending order of publication year. Asterisks (\*) indicate that the study involves a detection task, while the symbol (※) denotes that the loss function was proposed by the authors. In the table, ✗ represents the use of a private dataset, and ✓ represents the use of a public dataset. For studies employing more than three evaluation metrics, the three most representative metrics are selected. Shrinkage Lu et al. (2018); Lovász-hinge Berman et al. (2018); FTM (Focal Tree-Min) Li et al. (2022a); FD (Fourier Description).

| Paper                        | Dataset     | Targets                            | Methods | Loss                                            | Optimizer | GPU         | Evaluation                                                            | Highlights                                                                                                                                                                                   |
|------------------------------|-------------|------------------------------------|---------|-------------------------------------------------|-----------|-------------|-----------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Shi et al. (2021)            | 798 PAN ✗   | Root                               | CNN     | /                                               | /         | /           | Precision:<br>Original: 88.83%<br>Sobel: 77.85%<br>Canny: 72.76%      | Patch-based edge segmentation using LeNet-5, classifying pixel centers and connecting them to form tooth contours.                                                                           |
| GT U-Net Li et al. (2021)    | 248 PR ✗    | Root                               | Hybrid  | ※FD                                             | Adam      | RTX-2080 Ti | DSC: 92.54±0.25%<br>Accuracy: 93.98±0.11%<br>Specificity: 94.79±0.14% | GT U-Net integrates Group Transformer and MHSA, progressively modeling features from local to global.                                                                                        |
| RCS-Net Zhang et al. (2021b) | 78 CBCT ✗   | Root Canal                         | CNN     | Dice                                            | Adam      | Titan-Xp    | DSC: 95.2%<br>VOE: 0.077mm<br>ASD : 0.125mm                           | RCS-Net extends 3D U-Net with dual-stage segmentation and multi-task loss for precise root canal detection.                                                                                  |
| Oztekin et al. (2022)        | 250 PAN ✗   | Dental Prosthesis                  | CNN     | Focal                                           | Adam      | /           | mIoU: 76.7%<br>Accuracy: 99.81%                                       | U-Net segments amalgam and composite resin fillings in panoramic dental radiographs automatically.                                                                                           |
| Rousseau et al. (2023)       | 2500 BR ✗   | Dentin Enamel Pulp                 | CNN     | /                                               | Adam      | Tesla-T4    | All:<br>mIoU: 76.96%                                                  | DDPM-pretrained Unet with fine-tuning enables few-shot multi-class anatomical-level semantic segmentation.                                                                                   |
| Wang et al. (2023e)          | 215 CBCT ✗  | Tooth Pulp                         | CNN     | CE,L2<br>Focal.Smo<br>Shrinkage<br>Lovász-hinge | Adam      | Tesla-V100  | Pulp:<br>DSC: 87.5±3.1%<br>HD: 0.956±0.511mm                          | DentalNet instance segmentation with 3D U-Net ROI refinement using spatial embedding clustering for tooth recognition.                                                                       |
| Kim et al. (2023)            | 70 CBCT ✗   | Enamel Dentin Pulp                 | CNN     | CE,BCE<br>DICE<br>FTM                           | /         | /           | Enamel:<br>DSC: 85.65±0.29%<br>HD95: 1.37±0.26mm                      | Three-stage CBCT segmentation with DHU-Net and HFF module for multi-level tooth structure recognition.                                                                                       |
| Tan et al. (2023a)           | 200 CBCT ✗  | Enamel Dentin Pulp                 | CNN     | BCE<br>DICE                                     | Adam      | Tesla-V100  | Enamel:<br>DSC: 97.8%<br>HD95: 0.132mm                                | HMG-Net instance segmentation with tooth skeleton and multi-scale Frangi filter improves CBCT pulp and dentin segmentation.                                                                  |
| DSIS-DPR Wang et al. (2024e) | 29112 PAN ✗ | Tooth Enamel Dentin Pulp           | CNN     | CE<br>BCE                                       | Adam      | /           | Enamel:<br>DSC: 79.14%<br>HD95: 14.39mm                               | Three-stage tooth structure segmentation: DSIS fine segmentation, ADM anomaly detection, DPR diffusion-based refinement.                                                                     |
| Tan et al. (2024)*           | 314 CBCT ✗  | Tooth Enamel Dentin Pulp           | CNN     | CE<br>SL1<br>DICE                               | /         | /           | Enamel:<br>DSC: 92.85±0.04%<br>HD95: 0.297±0.22%                      | Progressive three-stage tooth and substructure segmentation: centroid detection, instance segmentation, enamel/dentin/pulp segmentation.                                                     |
| Li et al. (2024a)            | 8892 CBCT ✓ | Enamel Dentin Pulp                 | CNN     | BCE<br>SSIM                                     | Adam      | RTX-2070    | Enamel & Dentin:<br>DSC: 95.79±0.58%<br>ASSD: 0.24±0.08mm             | A hybrid pipeline with a multi-scale attention U-Net for pulp segmentation and dual level set models for enamel-dentin segmentation, ensuring boundary accuracy and inter-slice consistency. |
| D3UNet Liang et al. (2024)   | 300 CBCT ✓  | Tooth Root Canal                   | CNN     | GD<br>CE<br>Focal                               | Adam      | RTX 4090    | Root:<br>DSC: 94.15%<br>ASSD: 0.1357mm                                | Dual-view framework localizes ROI globally and refines locally; 2.5D fusion and dual-boundary loss improve accuracy.                                                                         |
| Yu et al. (2025)             | 30 CBCT ✗   | Enamel                             | CNN     | /                                               | /         | RTX-5000    | DSC: 96.6%                                                            | 2.5D Attention U-Net segments CBCT enamel, using early stopping, external validation, and 3D thickness computation.                                                                          |
| Zhou et al. (2025b)          | 10k PR ✓    | Anatomical-structures& instruments | CNN     | CE<br>DICE                                      | Adam      | RTX-3090    | DSC: 96.6%                                                            | PRAD-10K dataset and PRNet provide high-quality annotations and benchmarks for periapical radiograph segmentation.                                                                           |

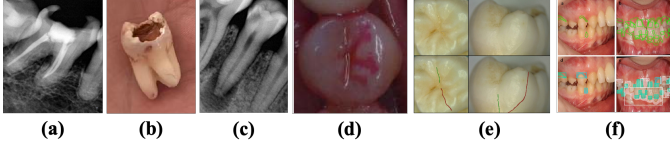


Figure 16: Illustrations of common dental diseases. Panels (a) and (c) are sourced from Zhou et al. (2025b), showing dental caries and periapical periodontitis, respectively. Panel (b) depicts an impacted third molar with caries, obtained from an extracted tooth by the author. Panel (d) shows dental plaque Shi et al. (2022); panel (e) illustrates a tooth crack Huang et al. (2025b); and panel (f) displays dental calculus, gingival hyperplasia, and inflammation Yurdakurban et al. (2025).

Rousseau et al. (2023) proposed a pretraining strategy based on the denoising diffusion probabilistic model (DDPM), where a U-Net is pretrained on a large corpus of unlabeled data and subsequently fine-tuned on a limited set of annotated samples to achieve segmentation of dentin, enamel, and pulp in BR. As shown in Fig. 15, Kim et al. Kim et al. (2023) developed DHU-Net, which incorporates a hierarchical feature fusion module and a spatial attention mechanism to extract multi-scale features from individual tooth patches and differentiate between dentin, enamel, and pulp, achieving strong performance on CBCT images. Wang et al. Wang et al. (2024e) introduced a three-stage model that first employs a GAN-based architecture to generate anatomical-level segmentation masks of teeth, followed by iterative refinement through an anomaly detection module and a diffusion prior restoration module. Their method achieved superior results on PAN images from 996 patients, outperforming previous models.

#### 4.1.2. Dental Diseases

Common dental diseases encountered in clinical practice are illustrated in Fig. 16. As shown in Fig. 19, by accurately identifying pathological features, such models can provide clinicians with objective and quantitative references, thereby improving diagnostic efficiency and treatment precision. Dental plaque represents one of the most characteristic dental diseases. Unlike most pathologies that rely on radiographic image analysis, plaque detection is typically performed using stained IP. Li et al. Li et al. (2020, 2022c) proposed a two-stage architecture that integrates CNNs with heat kernel signatures and local binary patterns to extract local features, thereby enhancing the network’s ability to learn shapes and boundaries and achieving efficient plaque segmentation. Similarly, as shown in Fig. 17, Shi et al. Shi et al. (2022) developed a contrastive learning-based dual-decoder model that employs a shared encoder to extract image features, with two independent decoders to predict plaque and tooth masks, respectively. The model leverages contrastive learning to enhance the separation between the two decoders in latent space and employs auxiliary supervision to produce sharper segmentation boundaries.

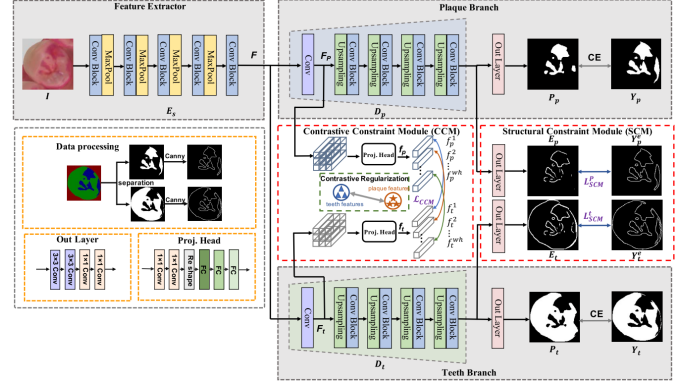


Figure 17: SDNet for tooth plaque segmentation utilizes a shared encoder and dual decoders for semantic disentanglement. The Class Conditional Module enhances class separability, while the Spatial Conditioning Module refines spatial details through boundary supervision, thereby improving accuracy and robustness. Image sourced from Shi et al. (2022).

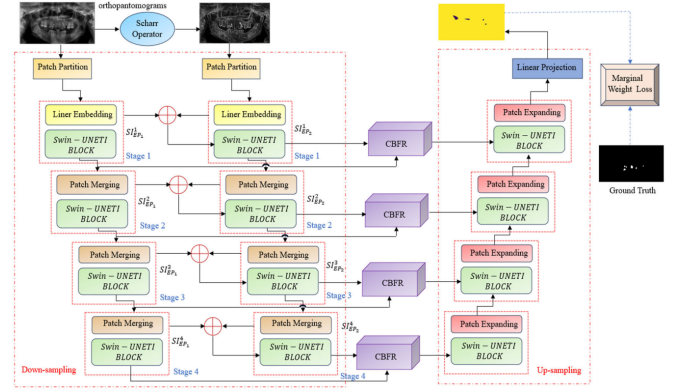


Figure 18: DentC\_BSFLNet for caries segmentation employs a multi-path encoder-decoder architecture with SwinUNET blocks and a CBFR module to fuse multi-scale features. A boundary-weighted loss function is utilized to enhance boundary accuracy and sensitivity. Image sourced from Neeraja and Anbarasi (2025).

The prediction of caries, periapical lesions, and other abnormalities has become a major research focus in the field of dental image analysis. For caries prediction, Wang et al. Wang et al. (2023b) proposed a semi-supervised framework based on uncertainty estimation to segment carious regions in PAN images. This framework employs deterministic masks derived from Monte Carlo aggregation to ensure consistency between predictions of the teacher and student models. Zhu et al. Zhu et al. (2023) introduced CariesNet, a U-shaped network with deep supervision that incorporates a full-scale axial attention module to simultaneously capture boundary cues and extract fine-grained feature maps that integrate high-level semantics and low-level details, achieving multi-stage caries segmentation in PAN images. Liu et al. Liu et al. (2025a) further developed a cascaded learning architecture that integrates YOLOX Ge et al. (2021) and FCN Long et al. (2015) for accurate caries detection and segmentation in PR images. For the detection of periapical diseases and other oral

Table 5: A summary of dental disease dense prediction studies, arranged in ascending order of publication year. Asterisks (\*) indicate that the study involves a detection task, while the symbol (✖) denotes that the loss function was proposed by the authors. In the table, ✖ represents the use of a private dataset, and ✓ represents the use of a public dataset. When multiple dental lesions are involved, “Multiple Abnormal Conditions” is used to describe the targets. For studies employing more than three evaluation metrics, the three most representative metrics are selected. WT (Wisdom Teeth); RR (Residual Roots); WS (Wear Surface); DC (Dental Calculus); Combo Taghanaki et al. (2019); FT (Focal Tversky) Abraham and Khan (2019); MW (Marginal Weight) Tran et al. (2025); BD (Boundary Dou) Sun et al. (2023); B-CE (Bootstrapped-CE) Chao et al. (2019); PA (Pixel Accuracy); FWIoU (Frequency Weighted IoU).

| Paper                                        | Dataset     | Targets                      | Methods    | Loss                    | Optimizer              | GPU         | Evaluation                                                         | Highlights                                                                                                                                                                                             |
|----------------------------------------------|-------------|------------------------------|------------|-------------------------|------------------------|-------------|--------------------------------------------------------------------|--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dense U-NET<br>Zheng et al. (2020)           | 20 CBCT ✖   | Periapical Lesions           | CNN        | CE<br>Dice<br>Focal     | Adam                   | /           | Precision: 90.0%<br>Recall: 84.0%                                  | Dense U-Net from modified FC-DenseNet with anatomical constraints for multi-class dental CBCT segmentation.                                                                                            |
| Li et al. (2020)                             | 607 IP ✖    | Dental Plaque                | CNN        | /                       | /                      | GTX-1080    | IOU:<br>Children: 85.9%<br>Adults: 64.0%                           | DeepLabV3+ with SLIC superpixels and multi-feature fusion, combined with random forest for plaque segmentation.                                                                                        |
| Li et al. (2022c)                            | 2884 IP ✓   | Dental Plaque                | CNN        | CE                      | /                      | GTX-1080    | PA: 84.59%<br>mIoU: 73.64%                                         | Self-attention CNN with local-to-global feature fusion for automatic dental plaque segmentation.                                                                                                       |
| SDNet<br>Shi et al. (2022)                   | 1869 IP ✖   | Dental Plaque                | CNN        | CE<br>BCE<br>DICE       | Adam                   | Tesla-V100  | DSC:<br>Seg-S: 95.58%<br>Seg-C: 94.94%                             | SDNet with shared encoder and dual decoders, enhanced by contrastive and structural constraints for plaque segmentation.                                                                               |
| TripleNet<br>Chen et al. (2022)              | 2884 IOS ✖  | Tooth-Defect                 | CNN        | Focal                   | Adam                   | RTX-3090    | IoU: 77.57%<br>Precision: 83.75%<br>F1-score: 86.25%               | A triple-stream multi-scale EdgeConv network integrates 18D geometric features for point-wise defect detection.                                                                                        |
| MLUA<br>Wang et al. (2023b)                  | 1000 PAN ✓  | Caries                       | CNN        | CE<br>Dice              | AdamW                  | RTX-3090    | DSC: 61.40%<br>Sensitivity: 58.77%<br>Precision: 70.04%            | Teacher-student framework with EMA and Monte Carlo sampling generates uncertainty masks for consistency learning.                                                                                      |
| CGA-UNet<br>Wang et al. (2023c)              | 591 IP ✖    | Multiple Abnormal Conditions | Hybrid     | B-CE                    | Adam                   | GTX-3090 Ti | mIoU: 63.21±1.47%<br>FWIoU: 94.35±0.84%<br>Accuracy: 96.97±0.51%   | CGA-UNet decomposes dental images into low- and high-frequency features for lesion enhancement.                                                                                                        |
| Hadzic et al. (2023)                         | 144 CBCT ✖  | Periapical Lesions           | CNN        | FT<br>Combo<br>Focal    | Adam                   | Titan X     | DSC:<br>FL: 67±3%<br>FTL: 70±4%<br>CL: 70±4%                       | SCN predicts 3D tooth coordinates and crops regions; U-Net segments apical lesions.                                                                                                                    |
| Felsch et al. (2023)                         | 18,179 IP ✖ | Caries Hypomineralization    | VIT        | /                       | /                      | RTX-A6000   | IoU: 95.9%<br>Accuracy: 97.8%<br>F1-score: 97.7%                   | SegFormer-B5 visual Transformer trained with DeepSpeed multi-server setup and augmented images simulating diverse tooth imaging conditions.                                                            |
| CariesNet<br>Zhu et al. (2023)               | 1159 PAN ✖  | Caries                       | CNN        | BCE<br>Dice             | Adam                   | RTX-2080 Ti | DSC: 93.64%<br>Accuracy: 93.61%<br>F1-score: 92.87%                | CariesNet employs U-Net with Res2Net backbone and full-scale axial attention to fuse multi-level features for caries saliency prediction.                                                              |
| LABANet<br>Chen et al. (2023) *              | 1000 PAN ✖  | Caries<br>RR<br>WT           | Hybrid     | /                       | Adam                   | Tesla-A40   | DSC:<br>WT:86.0%<br>RR: 90.3%<br>Caries: 83.6%                     | LABANet employs dual Swin Transformer backbones with enhanced SE attention head for multi-lesion oral segmentation.                                                                                    |
| Ma et al. (2023b)                            | 200 IP ✖    | WS<br>DC<br>Gingivitis       | CNN        | CE<br>Dice<br>$L_{adv}$ | Adam                   | GTX-3090    | mIoU: 72.79%<br>Accuracy: 83.62%<br>F1-score: 81.31%               | DGAN-based dental lesion segmentation combines supervised and semi-supervised learning, enhanced DeepLabV3+ for multi-scale lesions.                                                                   |
| BiseFormer<br>Ma et al. (2024b)              | 400 IP ✖    | Gingivitis<br>DC<br>WS       | Hybrid     | CE                      | /                      | RTX-3090    | mIoU: 79.76%<br>mAcc: 87.7%                                        | BiseFormer enables precise dental lesion segmentation via ConvFormer blocks, edge attention, and a hybrid CNN-ViT architecture.                                                                        |
| CenterFormer<br>Song et al. (2024b)          | 4884 IP ✓   | Dental Plaque                | Hybrid     | /                       | /                      | /           | mIoU: 75.86%<br>mPA: 85.94%<br>Accuracy: 86.51%                    | CenterFormer uses a Cluster Center Encoder, multiple granularity perceptions, and a pyramid hybrid encoder for precise dental image segmentation.                                                      |
| Jiang et al. (2024)                          | 2884 IP ✓   | Dental Plaque                | CNN        | CE<br>BCE               | /                      | /           | mIoU: 74.02%<br>mAcc: 84.04%                                       | FGN uses a high-to-low frequency dual-task pipeline with frequency-guided decoupling and a frequency-driven refinement module for plaque segmentation.                                                 |
| Koshy and Benny (2024)                       | 560 PAN ✓   | Caries                       | CNN        | BCE                     | Adam                   | /           | Accuracy: 89%                                                      | This study uses U-Net with encoder-decoder and skip connections to detect dental caries.                                                                                                               |
| Liu et al. (2025a) *                         | 5569 PR ✖   | Caries                       | CNN        | Dice                    | SGD                    | RTX-3090    | DSC: 84.6%                                                         | Cascaded YOLOX detection and FCN-8s segmentation to accurately detect and segment dental caries on periapical radiographs.                                                                             |
| Yurdakurban et al. (2025) *                  | 1000 IP ✖   | Multiple Abnormal Conditions | CNN        | Tversky                 | /                      | Tesla-T4    | F1-score:<br>Y: 63.0±0.7%<br>U+R: 64.0±4.6%                        | YOLOv8 for fast multi-object detection, and U-Net with ResNet50 for precise high-resolution lesion segmentation.                                                                                       |
| DentC-BSFLNet<br>Neeraja and Anbarasi (2025) | 153 PAN ✖   | Caries                       | Hybrid     | MW                      | AdamW                  | RTX-3050    | DSC: 95.93%<br>Accuracy: 98.53%<br>Precision: 96.41%               | DentC-BSFLNet uses a dual-path Swin-UNETI encoder, boundary refinement, and marginal weight loss for accurate caries boundary segmentation.                                                            |
| AEDD-Net<br>Huang et al. (2025a)             | 389 PR ✖    | Caries                       | Hybrid     | BD<br>Dice<br>✖Edge     | SGD                    | /           | DSC: 76.8%<br>Precision: 82.7%<br>Sensitivity: 74.9%               | AEDD-Net with a CNN-Transformer hybrid encoder and parallel CA-Decoder, Edge-Decoder for precise caries segmentation, and a multi-loss function with boundary optimization for enhanced edge accuracy. |
| Jones et al. (2025)                          | 4396 IP ✖   | Caries                       | CNN        | Focal                   | AdamW                  | RTX-4090    | Precision:<br>MB: 99%<br>VASC: 99%                                 | Attention U-Net with a frequency-guided dual-task framework enables accurate boundary segmentation of children's dental caries.                                                                        |
| Huang et al. (2025b)                         | 2210 IP ✓   | Cracked-Tooth                | CNN        | BCE<br>SL1              | SGD                    | Tesla-V100  | AP: 79.51%<br>AP50: 95.75%<br>AP75: 86.78%                         | An improved Mask R-CNN with ResNeXt backbone, crack feature enhancement, and dynamic snake convolution enables boundary-aware tooth crack segmentation.                                                |
| Tez et al. (2025)                            | 506 IP ✖    | Dental Plaque                | CNN Hybrid | /                       | Adam<br>SGD<br>RMSProp | /           | Precision:<br>1st.: 87.82%<br>2nd.: 82.03%<br>3rd.: 80.06%         | Six SOTA DL models with U-Net variants for automated dental plaque segmentation in children, and a statistical comparison with dentists for clinical validation..                                      |
| Hussein et al. (2025) *                      | 9573 PAN ✖  | Multiple Abnormal Conditions | CNN VIT    | /                       | AdamW                  | /           | mAP50:<br>YOLO: 61%<br>RT-DETR: 60%                                | RT-DETR: Transformer-based, detects overlapping dental X-rays via global attention; YOLO: CNN-based, fast single-stage detection.                                                                      |
| Ari et al. (2025)                            | 300 CBCT ✖  | Multiple Abnormal Conditions | CNN        | /                       | SGD                    | Tesla-V100  | Caries:<br>Sensitivity: 91.66%<br>Precision: 97.84%<br>AUC: 86.79% | Mask R-CNN with a ResNet101 backbone detects and segments 15 dental conditions in CBCT slices via region-based CNNs with mask prediction.                                                              |

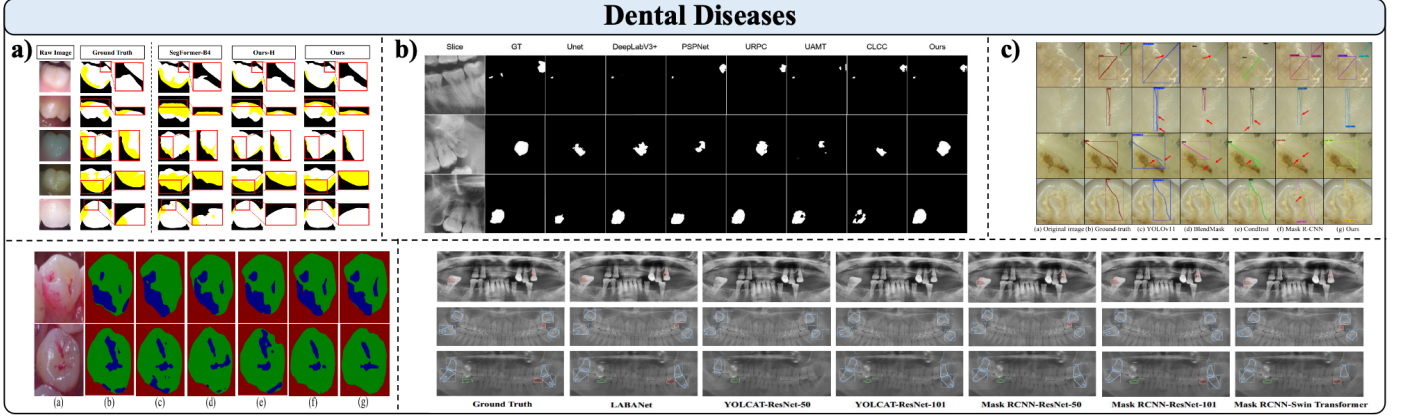


Figure 19: Visualization of dense detection results for dental lesions. Panels (a) and (b) display dental plaque, sourced from Jiang et al. (2024) and Shi et al. (2022), respectively; panel (c) depicts tooth cracks from Huang et al. (2025b); panel (d) illustrates impacted third molars, sourced from Chen et al. (2023); and panel (f) presents trends from Wang et al. (2023b).

abnormalities and infections, CNN-based architectures remain the mainstream research paradigm. Zheng et al. (2020) proposed an enhanced Dense U-Net that incorporates anatomical constraints during training, embedding oral structural knowledge into the network to achieve accurate segmentation of periapical lesions, tooth structures, and fillings in CBCT images. Chen et al. (2022) designed an edge convolution-based network that fuses outputs from multi-scale convolutional branches, enhancing the model’s adaptability to diverse tooth surface morphologies and successfully predicting tooth defect regions in IOS data.

Wang et al. (2023c) developed CGA-UNet, which decomposes the input into high- and low-frequency components and applies a class-guided attention mechanism to each frequency branch for feature reweighting. The fused representations are integrated into the encoder-decoder structure of an Attention-UNet, achieving accurate segmentation of multiple abnormal structures in IP images. Chen et al. (2023) proposed LABANet for multi-lesion segmentation in oral radiographs. It employs a Swin Transformer-based backbone and an auxiliary network. Their outputs are fused through a region proposal network and refined by an attention-based FCN, yielding precise identification of caries, residual roots, and wisdom teeth on the PAN dataset.

#### 4.1.3. Maxillofacial Structure

As shown in Fig. 20, the maxillofacial structure comprises multiple critical anatomical structures, including the maxilla, mandible, mandibular canal, inferior alveolar nerve, mental foramen, temporomandibular joint, maxillary sinus, and hard palate. These structures play vital roles in mastication, speech, and sensory function. Accurate prediction of these regions is essential for safe surgical planning, dental implant placement, and the prevention of nerve injury.

According to existing studies, predictions of maxillofacial structures are typically based on relatively fixed imag-

ing modalities such as CBCT or PAN, as radiographic imaging inherently excels in visualizing internal anatomical structures. In the early stages of research, Dasanayaka et al. (2019) proposed a semi-automatic segmentation framework for the mental foramen based on U-Net, achieving accurate prediction through customized image augmentation strategies. Kong et al. (2020) developed EED-Net, which introduced a multi-path feature extraction module into the bottleneck of the U-Net, thus improving the breadth of encoder features and maintaining high-resolution representations for efficient segmentation of maxillofacial ROIs. Zhang et al. (2021c) proposed a multi-directional resampling mechanism that generated new input images from three orthogonal directions, enabling the model to capture spatial information from multiple perspectives and successfully segment the temporomandibular joint in CBCT. Lang et al. (2022) designed a multi-region proposal network for landmark localization in CBCT images, generating multiple landmark candidates, and an FCN-based recognition network predicted bone masks from these proposals, effectively capturing the structural information surrounding the maxillofacial landmarks. With the expansion of research efforts and the increasing availability of large-scale datasets, more maxillofacial structures have been incorporated into DL studies, and hybrid architectures have gradually emerged. Jarnstedt et al. (2022) constructed a private large-scale CBCT dataset containing 1,103 patient scans and conducted a comprehensive performance analysis of various CNN models for mandibular canal segmentation. Amorim et al. (2023) employed a 3D U-Net with residual modules, achieving efficient and stable segmentation of mandibular regions across 476 CBCT scans. Zhang et al. (2023c) proposed a two-stage model named DBA-UNet, where the first stage, a boundary cross-fusion network, consists of two parallel U-Nets of identical size to refine boundary representations,

Table 6: A summary of maxillofacial structure dense prediction studies, arranged in ascending order of publication year. Asterisks (\*) indicate that the study involves a detection task, while the symbol (※) denotes that the loss function was proposed by the authors. In the table, ✕ represents the use of a private dataset, and ✓ represents the use of a public dataset. For studies employing more than three evaluation metrics, the three most representative metrics are selected. MF (Mental Foramen); LLC (Landmark Localization Craniomaxillofacial); TMJ (Temporomandibular Joint); MS (Maxillary Sinus); AB (Alveolar Bone); MC (Mandibular Canal); IAN (Inferior Alveolar Nerve); SMCD (Symmetric Mean Curve Distance); MSE (Mean Squared Error); FD (Focal Dice); SD (Soft Dice).

| Paper                         | Dataset       | Targets                        | Methods    | Loss                                                 | Optimizer     | GPU         | Evaluation                                                                                   | Highlights                                                                                                                                                                                                                                                           |
|-------------------------------|---------------|--------------------------------|------------|------------------------------------------------------|---------------|-------------|----------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Dasanayaka et al. (2019)      | 1000 PAN ✕    | MF                             | CNN        | /                                                    | /             | /           | DSC: 98.7%                                                                                   | U-Net-based semi-automatic DPT mental foramen segmentation with CLAHE, ROI cropping, and multi-augmentation.                                                                                                                                                         |
| EED-Net Kong et al. (2020)    | 2602 PAN ✕    | Maxillofacial                  | CNN        | BCE Dice                                             | Adam          | RTX-2080 Ti | IoU: 98.29%<br>Accuracy: 99.28%<br>HD: 8.32mm                                                | EED-Net combines U-Net with residual encoder, multipath Inception-ResNet, and lightweight decoder for efficient maxilla segmentation.                                                                                                                                |
| Zhang et al. (2021c)          | 89 CBCT ✕     | TMJ                            | CNN        | BCE                                                  | SGD           | RTX-2080 Ti | DSC: 98.14±0.54%<br>ASD: 0.0555±0.0198mm<br>HD: 1.5711±1.0252mm                              | Two-stage multi-view 3D U-Net fuses transverse, sagittal, coronal, and original features for final segmentation.                                                                                                                                                     |
| Huang et al. (2021a)          | 40 CBCT ✕     | Maxillofacial                  | CNN        | /                                                    | /             | /           | Accuracy<br>Lesion: 63%<br>Bone: 89%<br>Teeth: 80.1%                                         | Semi-supervised 3D U-Net integrates anatomical knowledge constraints to enhance multi-label CBCT segmentation.                                                                                                                                                       |
| Karacan and Yucebas (2022)    | 1000 PAN ✓    | Tooth<br>Maxillofacial         | CNN<br>ViT | /                                                    | Adam          | Tesla-P100  | DSC(Maxillofacial):<br>Segmenter: 98.12%<br>ConvNeXt: 95.83%<br>U-Net: 91.50%<br>ViT: 87.33% | Deep learning models segment teeth and mandible using ViT, Segmenter, ConvNeXt with mask supervision.                                                                                                                                                                |
| Lang et al. (2022) *          | 50 CBCT ✕     | LLC                            | CNN        | CE<br>L1                                             | SGD           | GTX-1080 Ti | MSE:<br>1.38 ± 0.95 mm                                                                       | Introduces a Multi-RPN module combined with Mask R-CNN to achieve accurate craniofacial landmark localization.                                                                                                                                                       |
| Du et al. (2022)              | 20 CBCT ✕     | MC                             | CNN        | Weighted-BCE                                         | Adam          | RTX-2080    | DSC: 85.91%<br>HD95: 0.5371mm                                                                | Attention-3D U-Net segments mandibular canal using scSE channel/spatial attention, fixed-point ROI cropping, and weighted BCE.                                                                                                                                       |
| Usman et al. (2022)           | 1357 CBCT ✓   | MC                             | CNN        | Dice                                                 | Adam          | RTX-Titan   | DSC: 74.8%<br>Specificity: 99%<br>F1-score: 74.9%                                            | Two-stage deep learning framework segments mandibular canal using VOI cropping, 3D attention UNet, and multi-scale residual UNet.                                                                                                                                    |
| Jarnstedt et al. (2022)       | 1103 CBCT ✕   | MC                             | CNN        | /                                                    | /             | /           | SMCD: 0.81mm                                                                                 | 3D U-Net FCN segments mandibular canal, refined via skeletonization and heuristic path selection.                                                                                                                                                                    |
| Zou et al. (2022)             | 30 CBCT ✕     | Hard Palate in the Oral Cavity | CNN        | BCE<br>Dice                                          | Adam          | RTX-2080 Ti | mDSC: 76.44%<br>mHD: 10.63mm                                                                 | Dual-branch Res2Net with channel interaction and PSA enhances palatal edges, multi-step decoder refines segmentation.                                                                                                                                                |
| Amorim et al. (2023)          | 476 CBCT ✕    | Mandible                       | CNN        | Dice                                                 | Adam          | RTX-6000    | DSC: 96±0.17%<br>HD: 0.49±0.28mm                                                             | 3D U-Net-based mandibular segmentation using internal CT/CBCT dataset.                                                                                                                                                                                               |
| MPUNet Xu et al. (2023)       | 13 CBCT ✕     | Maxilla<br>Mandible<br>Tooth   | CNN        | CE                                                   | Adam          | RTX-3090    | All:<br>DSC: 95.14±1.21%<br>HD: 7.098±1.48mm                                                 | MPUNet-based 2D-view segmentation with pre-training, weighted distance loss, and watershed for tooth separation.                                                                                                                                                     |
| DBA-UNet Zhang et al. (2023c) | 538 CBCT ✕    | MS                             | CNN        | BCE<br>Dice                                          | RMSprop       | /           | DSC: 97.48%<br>Specificity: 99.87%<br>HD95: 1.929mm                                          | DBA-UNet: dual-stage network with boundary cross-fusion and axial-attention boundary segmentation.                                                                                                                                                                   |
| Widiasri et al. (2023)        | 563 CBCT ✕    | AB<br>MC                       | CNN        | BCE                                                  | Adam          | /           | mIoU:<br>AB: 98%<br>MC: 81%                                                                  | U-Net for 2D CBCT AB and MC segmentation with encoder-decoder and skip connections.                                                                                                                                                                                  |
| X-Match Weng et al. (2024)    | 13,200 CBCT ✕ | Maxilla<br>Mandible            | CNN        | /                                                    | /             | /           | mIoU(All):<br>ResNet-50: 86.2%<br>ResNet-101 : 87.1%                                         | X-Match semi-supervised segmentation leverages wavelet transform to generate multi-scale augmentations and fuses labeled with unlabeled data via a label fusion module.                                                                                              |
| Han et al. (2024)             | 443 CBCT ✓    | IAN                            | CNN        | IOU                                                  | SGD           | /           | DSC: 73.54±10.69%<br>HD95: 7.85±8.22                                                         | Three-stage IAN segmentation: weakly supervised label generation, manual refinement, followed by AttentionUnet3D supervised training.                                                                                                                                |
| Mai et al. (2024)             | / CBCT ✓      | Maxillofacial                  | Hybrid     | CE<br>Dice                                           | /             | RTX-4090D   | /                                                                                            | Two-stage CBCT segmentation: multi-view 2D SwinUnet extracts features, improved 3D Unet classifies fused results.                                                                                                                                                    |
| DMVEL-Net Hu et al. (2025)    | 88 CBCT ✕     | TMJ                            | CNN        | BCE                                                  | SGD           | RTX-2080 Ti | DSC: 98.18±0.51%<br>ASD: 0.0554±0.0182mm<br>HD95: 0.1743±0.0550mm                            | DMVEL-Net integrates coronal and sagittal CBCT features and enforces symmetry for TMJ segmentation.                                                                                                                                                                  |
| Bolelli et al. (2024)         | 493 CBCT ✓    | MC                             | CNN        | CE,Dice<br>FD,SD<br>Focal<br>cl-Dice<br>Modifield HD | Adam<br>AdamW | Tesla-T4    | 1st.:<br>DSC: 79.6%<br>HD95: 4.49mm                                                          | The results of a medical image segmentation challenge. These methods primarily leveraged the nnU-Net, ResUNet framework and enhanced performance through model ensembling, advanced data augmentation, and specialized training strategies for precise segmentation. |
| Chen (2025)                   | 300 CBCT ✕    | MS                             | CNN        | /                                                    | RMSprop       | RTX-3060    | IoU<br>Axial plane :94.2%<br>Sagittal plane: 93.7%<br>Coronal plane: 91.6%                   | U-Net performs automatic segmentation of maxillary sinus on 2D slices, with encoder feature extraction and decoder upsampling and concatenation.                                                                                                                     |

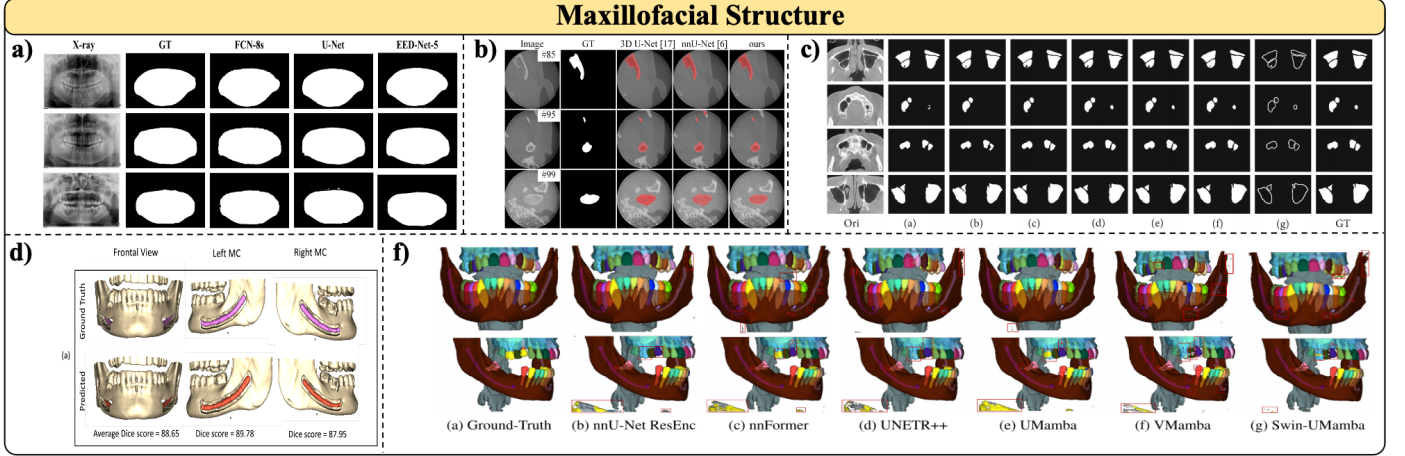


Figure 20: Visualization of segmentation results for maxillofacial structures. Panel (a) shows the maxillofacial region from Kong et al. (2020); panel (b) shows the temporomandibular joint from Zhang et al. (2021c); panel (c) presents the maxillary sinus from Zhang et al. (2023c); panel (d) illustrates the segmentation of the mandibular canal from Usman et al. (2022); and panel (f) depicts multi-structure segmentation results in the ToothFairy2 dataset, from Bolelli et al. (2025).

while the second stage, a boundary attention network with a dual-input single-output structure, enables precise segmentation of the maxillary sinus region in CBCT scans. Wen et al. Weng et al. (2024) introduced X-Match, a semi-supervised framework for maxillofacial bone segmentation in CBCT images, which applies wavelet transform to decompose images into low- and high-frequency components and leverages consistency training to enhance multi-scale feature extraction. Hu et al. Hu et al. (2025) developed DMVEL-Net, a multi-view ensemble learning network composed of an encoder-decoder backbone and multi-view feature fusion branches, which takes coronal and sagittal CBCT slices as inputs and enhances spatial representations through feature fusion, achieving accurate segmentation of the temporomandibular joint.

#### 4.2. Classification

Dental image classification tasks can generally be divided into two major directions: the identification of tooth types and conditions, and the recognition of dental diseases. DL models, through end-to-end feature extraction, enable pixel- or region-level dense predictions that provide additional structural features and well-defined ROIs, thereby enhancing the accuracy and robustness of classification performance.

##### 4.2.1. Tooth

For tooth classification tasks, dense prediction results are often employed either as foundational features or as intermediate steps to construct more intelligent and accurate automated diagnostic systems. Awari et al. Awari et al. (2024) proposed a tooth image segmentation and classification framework, where a CNN was first utilized for preliminary feature extraction to obtain tooth ROIs, followed by an enhanced CNN for refining the segmentation results.

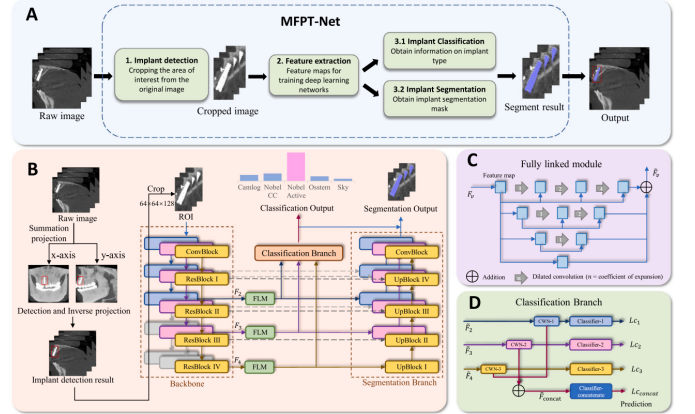


Figure 21: Overall architecture of MFPT-Net for implant classification and segmentation in CBCT images. Built on ResNet10, it employs multi-scale progressive training with Fully Linked Module and Channel-Wise Normalization modules to enhance global and channel features for robust multi-task performance. Image sourced from Zhao et al. (2025b).

Subsequently, a multilayer convolutional network was applied for tooth classification, achieving precise identification on IOS data. Alam et al. Alam et al. (2025) aimed to segment impacted molars from PAN images by integrating classification models based on CNNs, ViT, CLIP Radford et al., and hybrid architectures at the input stage to determine the presence of impacted wisdom teeth. If the classifier yielded a positive result, the image was subsequently passed to a UNet for segmentation. As shown in Fig. 21, Zhao et al. Zhao et al. (2025b) introduced MFPT-Net for implant classification and segmentation in CBCT images. After feature extraction by the backbone, a Fully Linked Module was employed to globally enhance multi-scale representations, while channel-wise normalization in the classification branch improved feature expressiveness, enabling efficient classification of implant types. Meine et

Table 7: A summary of tooth and dental disease studies, arranged in ascending order of publication year. Asterisks (\*) indicate that the study involves a detection task, while the symbol (✖) denotes that the loss function was proposed by the authors. In the table, ✖ represents the use of a private dataset, and ✓ represents the use of a public dataset. When multiple dental lesions are involved, “Multiple Abnormal Conditions” is used to describe the targets. For studies employing more than three evaluation metrics, the three most representative metrics are selected. OC (Odontogenic Cysts); FI (Furcation Involvement); MCC (Matthews Correlation Coefficient); MA (Macro-Accuracy); CR (Classification Rate); AUROC (Area Under the Receiver Operating Characteristic Curve).

| Paper                           | Dataset                  | Targets                                        | Methods                      | Loss & Optimizer                         | Evaluation                                                                               | Highlights                                                                                                                                                                                               |
|---------------------------------|--------------------------|------------------------------------------------|------------------------------|------------------------------------------|------------------------------------------------------------------------------------------|----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------|
| Tian et al. (2019)              | 600 IOS ✖                | Tooth                                          | CNN                          | CE & SGD                                 | MA:<br>Dataset1: 91.44%<br>Dataset2: 95.43%<br>Dataset3: 95.96%                          | A 3D CNN-based hierarchical classification and segmentation framework integrates sparse octree preprocessing and CRF refinement for automatic tooth analysis.                                            |
| Lakshmi and Chitra (2020)       | 1900 PAN ✓               | Caries                                         | CNN                          | /                                        | Accuracy: 96.08%<br>Precision: 78.50%<br>Recall: 84.20%                                  | Employ histogram equalization and Sobel edge detection, followed by an AlexNet-based deep CNN for early caries detection in dental X-rays.                                                               |
| Zhang et al. (2021a)            | 1000 IP ✖                | Caries                                         | CNN                          | /                                        | Accuracy: 97%                                                                            | Transfer learning via ImageNet pre-training and dental dataset fine-tuning enhances feature extraction.                                                                                                  |
| Sivasundaram and Pandian (2021) | 1171 PAN ✖               | OC                                             | CNN                          | /                                        | CR: 96.51%<br>Precision: 96.01%<br>F-score: 96%                                          | The modified LeNet-CNN performs four-class oral cyst classification and morphology-based segmentation of abnormal regions.                                                                               |
| Kumari et al. (2022)            | 120 PR ✓                 | Caries                                         | CNN RNN                      | /                                        | Accuracy: 93.67%<br>Specificity: 92.73%<br>F1-score :93.54%                              | The HSLnSSO algorithm optimizes FOC-KKC segmentation and ResNeXt-RNN classification parameters, improving model adaptability and convergence.                                                            |
| Li et al. (2023d)               | 151 CBCT ✖               | Pulp Calcification                             | ViT                          | CE + BCE<br>Dice<br>✖Triple<br>& Adam    | Accuracy: 78.47±0.25%<br>Precision: 77.31±0.37%<br>Recall: 82.08±0.23%                   | A coarse-to-fine CBCT tooth segmentation method is proposed, using a Swin Transformer and triple loss to improve segmentation, identification, and calcification detection.                              |
| PaXNet HaghaniFar et al. (2023) | 350 PAN ✓<br>120 PAN ✖   | Caries                                         | CNN                          | /                                        | Accuracy: 86.05%<br>Precision: 89.41%<br>Recall: 50.67%                                  | About 400 X-ray images were used for tooth segmentation, ROI extraction, and caries classification.                                                                                                      |
| Sarvesan et al. (2024)          | PR,BR,PAN ✖              | Periapical Lesions                             | CNN                          | /                                        | Accuracy: 97%                                                                            | A dual-branch model with feature fusion, transfer learning, and partial fine-tuning is used for periodontal classification.                                                                              |
| 3DDISC-DLTS Awari et al. (2024) | 1400 CBCT ✖              | Tooth                                          | CNN                          | /                                        | Accuracy:<br>D1: 96.47%<br>D2: 97.55%<br>D3: 97.80%                                      | The method uses U-Net for segmentation, DenseNet-169 for feature extraction, and a Tunicate Swarm-optimized Extreme Learning Machine for classification.                                                 |
| Hameed and AmalaShanthi (2024)  | 100 PAN ✓                | Tooth                                          | CNN                          | /                                        | Accuracy: 97.14%<br>Precision: 97.00%<br>Recall: 97.20%                                  | First use CM-CNN for 3D tooth segmentation, then AEG for classification, with FPN, ROI Align, and Swish/DropBlock to enhance performance.                                                                |
| MFPT-Net Zhao et al. (2025b)    | 537 CBCT ✖               | Dental-Implant                                 | CNN                          | CE<br>Dice                               | Accuracy: 92.98%<br>mPrecision: 93.15%<br>mRecall: 93.31%                                | MFPT-Net uses ResNet10 for implant classification and segmentation with multi-scale features and progressive training, while FLM enhances features and CWN normalizes channels.                          |
| Fang et al. (2025)              | 3005 PAN ✖               | Wisdom Teeth                                   | CNN                          | /                                        | Accuracy: 82.46%<br>F1-score: 77.49%<br>AUROC: 88.66%                                    | Left/right CNNs for image halves and a Full Image CNN for multi-scale features, then fuses them for multi-label wisdom tooth classification.                                                             |
| TEANet Li et al. (2025a)        | 304 IOS ✖                | Tooth                                          | CNN                          | L1<br>& Adam                             | Accuracy: 98.41%                                                                         | PointNet extracts features, a node-erasable graph models tooth adjacency, and Graph STN performs 3D alignment.                                                                                           |
| Alam et al. (2025)              | 693 PAN ✖                | Tooth                                          | CNN<br>ViT<br>CLIP<br>Hybrid | CE<br>Dice<br>&Adam                      | F1-score:<br>1st.:51%<br>2nd: 43%<br>3rd: 38%                                            | The study uses CNN, ViT, and CLIP for tooth classification, U-Net for segmentation, with feature fusion, data augmentation, and regularization to improve performance.                                   |
| Sehar et al. (2025)             | 105 CBCT ✓               | Tooth                                          | CNN                          | CE<br>& Adam                             | Accuracy: 96.40%<br>Precision: 96.75%<br>Recall: 94.92%                                  | The method segments teeth using thresholding and 3D U-Net, reconstructs point clouds, extracts features via PointNet and HamNet, and classifies with a prototype network.                                |
| Meine et al. (2025)             | 209 PAN ✖                | Tooth                                          | CNN                          | CE<br>SL1<br>BCE<br>& SGD                | Precision: 98.1%<br>Recall: 97.1%<br>Specificity: 96.2 %                                 | Mask R-CNN segments teeth with bounding boxes, masks, and FDI labels, filtering predictions by confidence and determining jaw type via Kennedy’s classification.                                         |
| Kibcak et al. (2025)            | 7696 PAN ✓               | Peri-implantitis                               | CNN                          | Adam                                     | Precision: 77.7%<br>Recall: 90.3%<br>DSC: 98.6%                                          | Implants are segmented with U-Net using skip connections, then classified with AlexNet to detect peri-implantitis.                                                                                       |
| Zhang et al. (2025b)            | 1089 PAN ✖<br>506 CBCT ✖ | FI                                             | CNN<br>ViT                   | Adam                                     | F1-score(Both):<br>MLP: 84±2.0%<br>GoogLeNet: 91±2.4%<br>VGGNet: 93±2.5%<br>ViT: 95±2.3% | The method uses six-fold stratified cross-validation; VG-Net extracts fine-grained features, GoogleNet captures multi-scale patterns, and ViT models global dependencies with multi-head self-attention. |
| Silva et al. (2025) *           | 16824 PAN ✖              | Multiple Abnormal Conditions                   | CNN<br>ViT                   | BCE<br>MCC<br>& AdamW                    | MCC: 47.5%                                                                               | HTC generates tooth pseudo-labels, ViT with MAE classifies lesions, GPT-4 extracts lesion names, and MCC evaluates performance.                                                                          |
| Arslan and Bilgin (2025) *      | 2332 PAN ✓               | Caries<br>Periapical Lesions<br>Impacted teeth | CNN<br>ViT<br>Hybrid         | /                                        | F1-score:<br>DINOv2: 79.00%<br>CNN: 66.19%<br>EfficientNet: 69.33%                       | The method integrates TransUNet, YOLO, SAM, GAN, and DINO for tooth segmentation, detection, and classification.                                                                                         |
| Li et al. (2025b)               | 358 CBCT ✖               | Multiple Abnormal Conditions                   | Hybrid                       | CE<br>Dice<br>Focal<br>& SGD             | Accuracy: 89.23%<br>Precision: 90.13%<br>Recall: 88.18%                                  | A unified network combines 3D U-Net segmentation and Transformer classification, using attention to fuse mask and image features.                                                                        |
| HC-Net Mei et al. (2025)        | 481 PAN ✖                | Periapical Lesion                              | CNN                          | Focal,L1<br>CE,Dice<br>BCE,MSE<br>& Adam | Accuracy: 92.7±1.3%<br>Precision: 90.2±1.4%<br>Recall: 95.8±2.1%                         | Hybrid framework combines tooth-level segmentation and patient-level prediction with clinical-guided fusion for periodontal diagnosis.                                                                   |
| TWGS Pang et al. (2025)         | 388 IP ✖                 | Tooth Wear Degree                              | CNN<br>ViT                   | /                                        | Precision: 91%<br>Recall: 88%<br>mPA : 95%                                               | TWGS: a Mask R-CNN-based model for single-tooth segmentation, followed by a ViT for wear severity classification on individual tooth surfaces.                                                           |
| Nasra et al. (2025)             | 11,653 IP ✖              | Multiple Abnormal Conditions                   | CNN                          | Adam                                     | Accuracy: 93%                                                                            | InceptionResNetV2 with customized linear head extracts multi-scale features for accurate oral disease classification.                                                                                    |

al. [Meine et al. \(2025\)](#) utilized Mask R-CNN for tooth instance segmentation in PAN images as a prerequisite step, and subsequently implemented an automatic five-class classification of partially edentulous jaws based on confusion matrix analysis.

#### 4.2.2. Dental Disease

Classification of dental lesions is a crucial component of computer-aided diagnosis in dentistry. Such systems typically rely on dense detection results to identify pathological or infectious regions in the teeth and maxillofacial area, providing clinicians and patients with intuitive risk alerts. Li et al. [Li et al. \(2023d\)](#) developed a ViT-based multi-task learning framework with a contrastive triple-loss mechanism to enhance the weak distinction between normal and calcified teeth, enabling both tooth segmentation and pulp calcification classification. Haghanifar et al. [Haghanifar et al. \(2023\)](#) proposed a caries classification framework that first extracts ROIs for each tooth from PAN images, followed by a CNN-based binary classification to determine caries presence. Sarvesan et al. [Sarvesan et al. \(2024\)](#) designed a dual-branch hybrid model combining pretrained VGG16 and MobileNet backbones to classify periodontal diseases from dental X-ray images. Mei et al. [Mei et al. \(2025\)](#) employed a backbone network to perform instance segmentation for detecting tooth centroids, bounding boxes, and masks, then used a pretrained encoder and classification head to achieve periodontal disease grading. As shown in Fig. 22, Pang et al. [Pang et al. \(2025\)](#) developed the TWGS system for tooth wear grading based on IP images, using Mask R-CNN for tooth instance segmentation and ViT for classifying abrasion levels across different tooth surfaces. Silva et al. [Silva et al. \(2025\)](#) proposed a multimodal dental lesion classification framework integrating PAN images and textual reports, where GPT-4 automatically extracts tooth-related lesion noun phrases from reports, and multiple ViT-based binary classifiers are trained to identify specific dental pathologies.

#### 4.3. Others

In addition to the aforementioned tasks, this study also identified several derivative tasks that build upon dense prediction and classification as fundamental components, extending their applications to fields such as biometric identification, registration, planning, and reconstruction. For example, in the field of dental biometric identification, Atacs et al. [ATAŞ et al. \(2022\)](#) analyzed PAN images from 372 patients collected between 2018 and 2020, with each patient having at least two scans taken at different times. They proposed PDR-Net, a model combining convolutional and fully connected layers for feature extraction and similarity measurement to match PAN images from the same patient. Similarly, Lai et al. [Lai et al. \(2020\)](#) utilized clinical PAN images captured before and after orthodontic treatments over the past two decades. Their model employed a shallow CNN backbone integrated

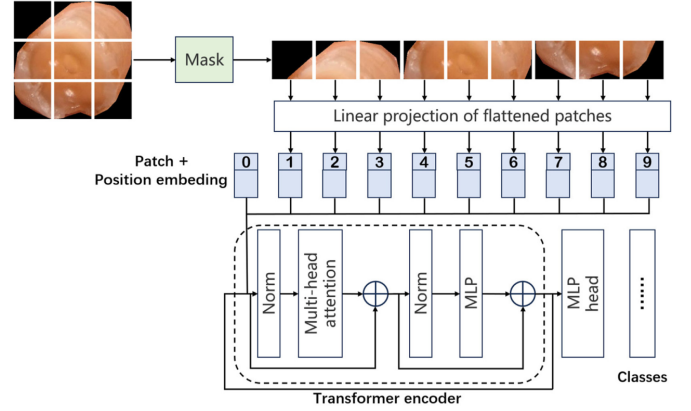


Figure 22: The architecture based on ViT is designed to classify the degree of tooth wear on different surfaces (occlusal/incisal vs. non-occlusal/non-incisal) of individual teeth in IP images. Image sourced from [Pang et al. \(2025\)](#).

with a channel attention module and a learnable connected module to enhance target tooth region representations and automatically learn cross-layer connection weights, achieving identity recognition via cosine similarity.

Beyond identification, DL has also been leveraged for registration, surgical planning, and 3D reconstruction in dental imaging. Jang et al. [Jang et al. \(2024\)](#) proposed TSIM, an automated registration and fusion framework for IOS and CBCT. TSIM-IOS and TSIM-CBCT respectively perform instance segmentation and labeling for each modality, followed by a stitching error correction module that incrementally aligns IOS data with CBCT references for accurate multimodal registration. Xing et al. [Xing et al. \(2024\)](#) developed an end-to-end framework for tooth segmentation and implant planning directly from PAN images. Using ResUNet for segmentation, they applied edge detection for contour extraction and utilized principal component analysis and linear regression to predict optimal implant locations. Tan et al. [Tan et al. \(2023b\)](#) introduced BSegNet, a graph-based network for bracket segmentation on 3D dental models, which further supports surface reconstruction. Their results demonstrate superior accuracy and efficiency compared with interactive reconstruction workflows, showing strong potential for clinical adoption.

## 5. Training and Evaluation

This section summarizes the key components involved in model training from three perspectives: loss functions, optimizers, and GPU hardware configurations. Furthermore, it categorizes and introduces the commonly used evaluation metrics in the DIA field from two dimensions: confusion-matrix-based and set-distance-based measures.

### 5.1. Loss Functions

In the training phase, the loss function serves as a crucial supervision strategy to measure the discrepancy between

the model's output and the ground truth (GT). By minimizing the loss value, the model is encouraged to align its predictions with the true labels, thereby improving accuracy. In the field of DIA, as shown in Fig. 23, Cross-Entropy (CE) loss and Dice loss [Milletari et al. \(2016\)](#) are the two most commonly used loss functions.

#### 5.1.1. L2 Loss

L2 loss, also known as Mean Squared Error (MSE), serves to quantify the discrepancy between the model's predictions and the GT. It calculates the mean of the squared differences, thus penalizing larger prediction errors more heavily, as defined in Equation 1.

$$\mathcal{L}_{\text{MSE}} = \frac{1}{N} \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (1)$$

Here,  $N$  denotes the number of training samples,  $y_i$  represents the GT of the  $i$ -th sample, and  $\hat{y}_i$  denotes the model's prediction for the  $i$ -th sample. Due to the squaring operation, MSE is a continuous and differentiable function, which facilitates gradient-based optimization. However, this squaring property can amplify errors when  $|y_i - \hat{y}_i| > 1$ . Consequently, outlier samples can exert a disproportionate influence on model parameter tuning, causing the model to overfit these anomalies while underfitting the majority of normal samples.

#### 5.1.2. L1 Loss

L1 Loss is the sum of the absolute differences between the GT and the model's predictions, mathematically defined as in Equation. 2:

$$\mathcal{L}_{\text{L1}} = \sum_{i=1}^N |y_i - \hat{y}_i| \quad (2)$$

The gradient of L1 loss is constant and non-differentiable at zero, which results in smaller model updates when errors are large and may require numerical approximations

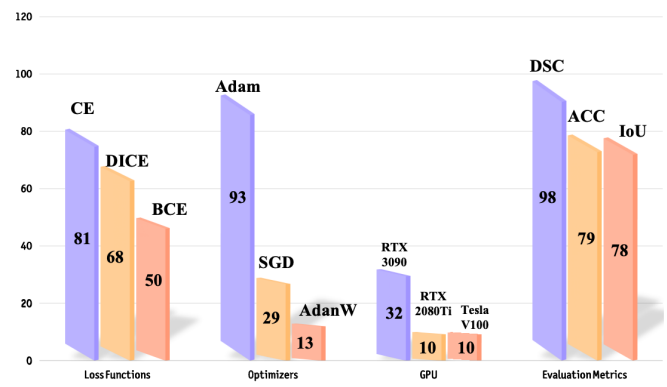


Figure 23: The figure presents statistics on the top three most frequently used loss functions, optimizers, GPUs, and evaluation metrics reported in the reviewed studies. The numbers indicate the frequency of occurrence for each item.

near zero, slowing down convergence. Unlike squared error loss, L1 loss does not amplify the discrepancies caused by outliers, making it more robust and suitable for regression problems with anomalous data.

#### 5.1.3. Smooth L1 Loss

Smooth L1 Loss combines the advantages of L1 and L2 losses, providing a balance between sensitivity to outliers and convergence stability. Its mathematical definition is shown in Equation. 3 and 4.

$$\mathcal{L}_{\text{SL1}} = \sum_{i=1}^N \ell_{\text{SL1}}(y_i - \hat{y}_i) \quad (3)$$

$$\ell_{\text{SL1}}(y_i - \hat{y}_i) = \begin{cases} 0.5(y_i - \hat{y}_i)^2, & \text{if } |y_i - \hat{y}_i| < 1 \\ |y_i - \hat{y}_i| - 0.5, & \text{otherwise} \end{cases} \quad (4)$$

It can be observed that when the error is small, the Smooth L1 Loss offers a stable gradient in a quadratic form. Conversely, when the error is large, it transitions to a linear form, enhancing the model's robustness and mitigating the influence of outliers on fitting. This property ensures precise boundary fitting for small targets while preventing gradient explosion caused by large prediction errors.

#### 5.1.4. Cross Entropy Loss

The Cross-Entropy (CE) Loss quantitatively measures the discrepancy between the model's predicted probability distribution and the true label distribution. It is the most widely used loss function in DIA. Its mathematical definition is given in Equation. 5:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N} \sum_{j=1}^C \sum_{i=1}^N y_{j,i} \log \hat{y}_{j,i} \quad (5)$$

Here,  $N$  denotes the total number of samples,  $C$  denotes the total number of classes,  $y_{j,i}$  represents the true distribution for the  $i$ -th class of the  $j$ -th sample, and  $\hat{y}_{j,i}$  denotes the predicted probability for the  $i$ -th class of the  $j$ -th sample. The softmax function is typically used to map mutually exclusive classes into a probability distribution summing to 1. A smaller cross-entropy value indicates that the predicted distribution  $\hat{y}_{j,i}$  is closer to the true distribution  $y_{j,i}$ .

For multi-target dental image detection, each sample may contain various types of teeth, such as incisors, molars, and wisdom teeth. The CE loss measures the discrepancy between the predicted probability  $\hat{y}_{j,i}$  and the GT  $y_{j,i}$  for each tooth class. Through gradient optimization, CE loss enables the model to distinguish between different classes, thereby achieving precise tooth detection.

The Binary Cross Entropy (BCE) loss is a specialized form of the CE loss function, tailored for binary classification tasks. Its mathematical definition is as the follow

Equation. 6:

$$\mathcal{L}_{\text{BCE}} = -\frac{1}{N} \sum_{i=1}^N [y_i \log \hat{y}_i + (1 - y_i) \log(1 - \hat{y}_i)] \quad (6)$$

Here,  $N$  denotes the total number of samples,  $y_i$  represents the GT for the  $i$ -th sample, and  $\hat{y}_i$  represents the model's predicted probability of the positive class. The BCE loss considers both positive and negative predictions, effectively measuring the model's overall performance in binary classification tasks.

##### 5.1.5. Dice Loss

The Dice Loss [Milletari et al. \(2016\)](#) is a commonly used loss function in segmentation tasks, derived from the Dice Similarity Coefficient (DSC), which measures the overlap between the predicted and GT segmentation regions, with values ranging from  $[0, 1]$ . By incorporating the loss minimization concept into this metric, the Dice Loss is defined in Equation. 7:

$$\mathcal{L}_{\text{Dice}} = 1 - \frac{2 \sum_{i=1}^N p_i g_i + \epsilon}{\sum_{i=1}^N p_i + \sum_{i=1}^N g_i + \epsilon} \quad (7)$$

Here,  $p_i$  denotes the predicted probability of the  $i$ -th pixel,  $g_i$  represents the corresponding GT,  $N$  is the total number of pixels, and  $\epsilon$  is a small constant added for numerical stability. Dice Loss directly optimizes the overlap between prediction and GT, effectively addressing class imbalance — a particularly important issue in medical image segmentation.

##### 5.1.6. IoU Loss

Similar to the design principle of the Dice Loss, the IOU Loss is derived from a commonly used evaluation metric in segmentation tasks—Intersection over Union (IoU). IoU measures the overlap between the predicted segmentation and the ground truth, defined as the ratio of the intersection area to the union area of the predicted and true regions. The IOU Loss is formulated to directly optimize this overlap, thereby improving segmentation accuracy. Its mathematical definition is given in Equation. 8:

$$\mathcal{L}_{\text{IOU}} = 1 - \frac{\sum_{i=1}^N p_i g_i + \epsilon}{\sum_{i=1}^N (p_i + g_i - p_i g_i) + \epsilon} \quad (8)$$

The IOU Loss, by penalizing the non-overlapping regions between the predicted segmentation and the GT, enhances the model's segmentation accuracy, especially along the object boundaries. This focus on boundary accuracy makes it particularly effective in improving the precision of segmentation tasks.

## 5.2. Optimizer and GPU

Optimizers and GPUs are two indispensable core components in the training of deep learning models. As shown

in Fig. 23, we summarized the optimizers and GPU configurations used in all reviewed studies. The results indicate that Adam [Kingma and Ba \(2015\)](#) remains the most widely adopted optimizer in the DIA field. Its popularity stems from its combination of the advantages of momentum and adaptive learning rate, enabling faster convergence and greater stability in complex medical imaging tasks. Compared with the traditional SGD [Robbins and Monro \(1951\)](#) optimizer, Adam demonstrates superior robustness and generalization when dealing with challenges commonly found in dental imaging, such as small sample sizes, high noise levels, and class imbalance. Furthermore, since DIA tasks often involve computationally intensive structures such as 3D convolutions and attention mechanisms, the adaptive nature of Adam in parameter updates helps alleviate issues of gradient explosion or vanishing.

In terms of GPU usage, the NVIDIA RTX 3090 is currently the most commonly adopted graphics card model in the DIA field. This preference is primarily attributed to its large memory capacity (24 GB) and favorable cost-performance ratio, which effectively support batch training of high-resolution dental images and the computational demands of 3D convolutional models. In contrast, high-end GPUs such as the A100 are less frequently used in DIA research, a trend that differs somewhat from other branches of medical imaging. The fundamental reason lies in the fact that DIA still lacks sufficiently large and standardized benchmark datasets; consequently, the data volume and parameter scale of model training remain relatively modest, and the demand for ultra-high computing power is not yet pressing. Moreover, many research teams conduct experiments within clinical institutions, where hardware budgets and deployment conditions are often constrained, leading them to prefer mid-to-high-end consumer-grade GPUs as their primary computing platforms.

## 5.3. Evaluation Metric

### 5.3.1. Metrics Based on the Confusion Matrix

**Intersection over Union (IoU)**, also known as the Jaccard coefficient, is a commonly used evaluation metric in image segmentation tasks. It measures the degree of overlap between the predicted result and the GT. IoU is defined as the ratio of the intersection area to the union area of the predicted and true segmentation regions, providing a quantitative assessment of segmentation accuracy. Its definition is shown in Equation. 9.

$$\text{IoU} = \frac{TP}{TP + FP + FN} \quad (9)$$

Here,  $TP$  (True Positive) denotes the number of pixels correctly predicted as positive,  $FP$  (False Positive) denotes the number of pixels incorrectly predicted as positive, and  $FN$  (False Negative) denotes the number of pixels incorrectly predicted as negative. The IoU value ranges from  $[0, 1]$ , with higher values indicating a greater overlap between the results and the GT. The Mean Intersection over

Union (mIoU) is particularly suitable for multi-class segmentation tasks. The basic idea is to compute the IoU for each class individually and then take the average across all classes, as shown in Equation. 10:

$$\text{mIoU} = \frac{1}{N} \sum_{i=1}^N \frac{TP_i}{TP_i + FP_i + FN_i} \quad (10)$$

Here,  $N$  represents the number of classes. In DIA, IoU is widely used as an evaluation metric for multi-tooth or multi-structure segmentation tasks. Due to its intuitive geometric interpretation in object detection, IoU is also commonly applied in dental lesion detection and tooth localization tasks.

**Dice Similarity Coefficient (DSC).** is one of the most commonly used evaluation metrics in medical image segmentation and is widely applied in DIA. Its definition is given by Equation. 11:

$$\text{DSC} = \frac{2TP}{2TP + FP + FN}, \quad (11)$$

The Mean Dice Similarity Coefficient (mDSC) extends the DSC to multi-class segmentation, similar to mIoU, and is defined as follow:

$$\text{mDice} = \frac{1}{N} \sum_{i=1}^N \frac{2TP_i}{2TP_i + FP_i + FN_i} \quad (12)$$

Compared with IoU, the DSC is more sensitive to extreme class imbalance, which is common in medical imaging. In dental images, target regions such as teeth or dental pulp usually occupy only a small portion of CBCT or PAN scans; thus, using DSC provides a more reasonable reflection of segmentation performance.

**Recall (Sensitivity).** is a commonly used performance metric in classification and segmentation tasks, aimed at evaluating a model's ability to correctly identify positive samples. Its value ranges from  $[0, 1]$ , with higher values indicating better coverage of target regions. Recall is defined as follow Equation. 13:

$$\text{Recall} = \frac{TP}{TP + FN} \quad (13)$$

In the field of DIA, Recall has been widely employed in disease detection tasks. For instance, in the identification of periodontal disease or caries, a higher Recall can effectively improve lesion detection rates and reduce the risk of missed diagnoses.

**Precision.** Evaluating model performance solely based on Recall is not comprehensive because excessively high Recall may result in increased false positives. Therefore, it is essential to balance Recall with Precision. Precision measures the proportion of true positive samples among all positive predictions made by the model. The value of Precision ranges from  $[0, 1]$ , with higher values indicating

greater accuracy in positive predictions. It is defined as follow:

$$\text{Precision} = \frac{TP}{TP + FP} \quad (14)$$

In the field of DIA, a higher Precision is crucial for reducing the incidence of misdiagnoses. For instance, in caries detection on PAN images, a higher Precision indicates that most of the predicted lesions are truly present, minimizing the excessive inclusion of background or non-pathological areas. This ensures that clinicians receive more reliable information, aiding in more accurate clinical decision-making and improving patient outcomes.

**Specificity.** is an important metric that measures the model's ability to correctly identify negative samples, i.e., background or non-target regions. It is defined in Equation. 15:

$$\text{Specificity} = \frac{TN}{TN + FP} \quad (15)$$

Here,  $TN$  (True Negative) denotes the number of pixels that are truly negative and correctly predicted as negative. The value of Specificity ranges from  $[0, 1]$ , where a higher value indicates the model's superior ability to avoid false positives. In medical image analysis, Specificity is often used in conjunction with Recall to provide a comprehensive assessment of diagnostic performance. In the field of DIA, Specificity is critically important. For instance, in CBCT-based segmentation of teeth or alveolar bone, a higher Specificity indicates that the model can effectively prevent background or noise from being misclassified as target structures, thereby enhancing the reliability of segmentation outcomes.

**F1-score.** is the harmonic mean of Precision and Recall, designed to provide a balanced evaluation between these two metrics. It ranges from  $[0, 1]$ , where a higher F1-score indicates an improved ability of the model to balance Precision and Recall. In medical image analysis, the F1-score is frequently utilized to evaluate the overall accuracy and stability of models in detection and segmentation tasks. It is defined in Equation. 16.

$$\text{F1-score} = \frac{2 \cdot \text{Precision} \cdot \text{Recall}}{\text{Precision} + \text{Recall}} = \frac{2TP}{2TP + FP + FN} \quad (16)$$

In DIA, the F1-score is a valuable and intuitive evaluation metric for dental clinical applications. For example, in dental plaque detection tasks, clinicians prioritize not only accurately identifying plaque regions but also maintaining precision in distinguishing background tooth areas. The F1-score effectively balances Precision and Recall, offering a comprehensive reflection of the model's overall performance in real-world clinical scenarios.

**Accuracy.** is a fundamental and widely used evaluation metric that measures the proportion of correctly classified pixels in the overall prediction. As defined in Equation. 17, Accuracy ranges from  $[0, 1]$ , with higher values indicating a greater proportion of correctly predicted pixels, thereby reflecting the model's overall discriminative capability.

$$\text{Accuracy} = \frac{TP + TN}{TP + TN + FP + FN} \quad (17)$$

In DIA, Accuracy is frequently used to evaluate the overall performance of classification and segmentation tasks. However, because background pixels often dominate dental images, relying solely on Accuracy can be misleading in scenarios with class imbalance. Therefore, Accuracy is typically used alongside the aforementioned metrics to provide a more comprehensive and reliable performance assessment.

### 5.3.2. Metrics based on the Set and Geometry

**Hausdorff Distance (HD).** is an important metric for evaluating the discrepancy between the predicted and true boundaries in segmentation tasks. It quantifies the maximum distance between the predicted boundary and the ground truth (GT) boundary. A smaller HD value indicates closer alignment between the predicted and true boundaries, reflecting higher precision in boundary localization. The mathematical definition is presented in Equation. 17.

$$\text{HD}(A, B) = \max \left\{ \sup_{a \in A} \inf_{b \in B} \|a - b\|, \sup_{b \in B} \inf_{a \in A} \|b - a\| \right\} \quad (18)$$

Here,  $A$  and  $B$  denote the sets of predicted and GT boundary points, respectively, while  $a$  and  $b$  represent arbitrary points within each set.  $\inf$  (infimum) refers to the greatest lower bound, taking the minimum among all distances, and  $\sup$  (supremum) refers to the least upper bound, taking the maximum among all values.  $\|\cdot\|$  denotes the Euclidean distance, which measures the spatial distance between two points. In DIA, particularly in 3D CBCT analysis, HD is a commonly used metric for evaluating boundary accuracy and is widely applied in the segmentation assessment of teeth, root canals, and jawbones. Since DSC or IoU alone cannot accurately capture boundary precision, HD's sensitivity to boundary deviations makes it an important complement to overlap-based metrics.

However, the standard HD is sensitive to noise, outliers, and anomalous points. To address this issue, the 95th percentile Hausdorff Distance (HD95) is commonly used in practical studies. HD95 considers the 95th percentile of the boundary distance distribution instead of the maximum distance, thereby reducing the impact of extreme points on the evaluation. The mathematical definition is presented in Equation. 19.

$$\text{HD95}(A, B) = \text{percentile}_{95}(\{d(a, B), d(b, A)\}) \quad (19)$$

HD95 eliminates the influence of extreme outliers, providing a more stable and objective reflection of the model's actual boundary performance.

**Average Surface Distance (ASD).** quantifies the average distance between the predicted segmentation boundary points and the ground-truth boundary points. This metric provides insight into the general alignment of the predicted boundaries with the true boundaries, as defined in Equation. 20.

$$\text{ASD}(A, B) = \frac{1}{|A|} \sum_{a \in A} d(a, B), \quad (20)$$

Here,  $A$  denotes the set of predicted boundary points, and  $B$  represents the set of ground-truth boundary points.  $|A|$  is the cardinality of set  $A$ , and  $d(a, B)$  denotes the minimum distance from point  $a \in A$  to set  $B$ . A smaller ASD indicates that the predicted boundary is, on average, closer to the GT boundary. In DIA, ASD is commonly used to assess the boundary consistency of target structures such as tooth surfaces. It is frequently employed alongside HD95 to provide a more comprehensive evaluation of boundary accuracy.

**Average Symmetric Surface Distance (ASSD).** is an extension of ASD. While ASD considers only the average deviation from the predicted boundary to the GT boundary, it may underestimate errors when the GT boundary exhibits complex structures. To address this limitation, ASSD calculates the average distance from the predicted boundary to the GT and from the GT to the predicted boundary, then takes the mean of both distances. This symmetric approach provides a more comprehensive measure of boundary accuracy. Its definition is given in Equation. 21.

$$\text{ASSD}(A, B) = \frac{1}{|A| + |B|} \left( \sum_{a \in A} d(a, B) + \sum_{b \in B} d(b, A) \right), \quad (21)$$

In DIA, Average Symmetric Surface Distance (ASSD) is frequently used as a key metric for assessing boundary accuracy, offering a more intuitive and precise geometric error measurement for evaluating model performance in clinical applications.

## 6. Discussion, Challenges and Future Directions

In the preceding sections, we systematically reviewed DL-based studies in DIA from three perspectives: data, models, and tasks with applications. The findings indicate that DL and AI technologies have become the most

widely adopted core methodologies in this field. Building upon this foundation, this chapter further summarizes research progress across these aspects, analyzes the major challenges currently facing DIA studies, and discusses potential future directions and trends.

## 6.1. Discussion and Challenges

### 6.1.1. Public Datasets

Public datasets are the cornerstone and essential resource driving advancements in AI and DL research. They provide standardized benchmarks for algorithm training and performance evaluation, enabling reproducibility and fair comparison of research outcomes. However, constructing medical public datasets poses significantly greater challenges compared to natural image datasets. On one hand, medical image annotation requires the involvement of clinicians with specialized expertise, as it often involves delineating complex anatomical structures or pathological regions, making the process time-consuming and costly. On the other hand, medical imaging data are diverse in modality and complex in acquisition, necessitating professional radiological procedures and strict adherence to patient privacy and ethical regulations. Consequently, the establishment and sharing of medical public datasets face substantial technical, temporal, and regulatory obstacles compared to their natural image counterparts.

This issue is particularly pronounced in the field of DIA. Most researchers still rely heavily on private datasets for experimentation and validation. Although considerable progress has been made, the lack of publicly available and reusable datasets hinders the full verification of algorithm reproducibility and transferability. As shown in Table 1, while a number of public DIA datasets have emerged, they remain significantly smaller in scale and quantity compared to those in other medical imaging domains, such as fundus images [Wu et al. \(2024\)](#); [Li et al. \(2019\)](#) or chest X-ray analysis [Johnson et al. \(2019\)](#); [Zhao et al. \(2024b\)](#). Even when the sample size exceeds 5K [Li et al. \(2024e\)](#); [Aminhj \(2023\)](#); [Seifossadat \(2023\)](#); [UBU \(2024\)](#), many datasets only provide detection- or classification-level annotations, lacking high-quality, dense segmentation masks—thereby constraining advancements in fine-grained segmentation research.

From a modality perspective, most existing public datasets are focused on CBCT and PAN images, while modalities such as IOS, PR, and IP have received relatively limited attention. However, these data types hold substantial clinical value in orthodontics, endodontics, and oral health, respectively, and should be given greater emphasis in future research. Notably, with the increasing application and proven potential of vision-language models (VLMs) in medical image analysis [Gu et al.](#); [Wu et al. \(2025b\)](#); [Nath et al. \(2025\)](#), several DIA studies have begun constructing multimodal dental imaging datasets to explore the feasibility and benefits of multimodal learning in this field [Wang et al. \(2025a\)](#); [Silva et al. \(2025\)](#). Encouragingly, since 2023, multiple international challenges

dedicated to DIA have been launched [ricoleehduu \(2023\)](#); [Wang et al. \(2024g, 2025e\)](#); [Bolelli et al. \(2025\)](#), providing standardized platforms for fair algorithm evaluation, enhancing research engagement, and fostering the dissemination of high-quality scientific outcomes.

However, the construction of high-quality public datasets for DIA still faces several pressing challenges. First, existing datasets are limited in both scale and quality, with incomplete coverage of certain imaging modalities that require further supplementation and refinement. Second, most available datasets are derived from economically developed regions, while publicly accessible DIA data from less developed areas remain scarce. According to the World Health Organization, Africa accounts for approximately 23% of the global disease burden but possesses only 4% of the world’s healthcare workforce [Organization et al. \(2025\)](#). As of 2022, there were only about 34,405 dentists across the entire continent, averaging 0.29 dentists per 10,000 people. Under such an imbalanced ratio of medical professionals to patients, the need for intelligent diagnostic and treatment assistance systems becomes particularly urgent, and the availability of open datasets serves as the fundamental basis for their development. Furthermore, due to disparities in imaging equipment, environmental conditions, and operator expertise, models trained on data from developed regions often fail to generalize effectively to less developed areas, resulting in a significant decline in performance. Addressing this cross-domain distribution discrepancy has thus become one of the core scientific challenges in current DIA research. Finally, with the growing adoption of VLMs, the construction of multimodal datasets for DIA has become increasingly imperative. Exploring the integration of textual data into dental imaging tasks has emerged as a key research direction. Although several studies have attempted to develop such multimodal datasets, most remain single-modality and limited in scale, making them insufficient to support the training of large-scale, comprehensive VLM baseline models for dental applications.

### 6.1.2. Model and Performance

Compared to traditional algorithms, the introduction of DL has significantly enhanced the accuracy and efficiency of DIA tasks. Due to the advantages of deep models in large-scale data feature representation learning, DL methods demonstrate superior performance over traditional algorithms in handling complex dental images and diverse tasks. From a model architecture perspective, most current studies focus on CNN architectures or hybrid models combining CNN with ViT, whereas the application of pure ViT models in DIA is relatively rare. This is mainly because tooth structures are complex and variable, and the boundaries between different tissues and bones are often indistinct. The strong ability of CNNs in local feature extraction is crucial for capturing these detailed features. Combining the global modeling capability of ViT with the local feature learning ability of CNN allows for better fea-

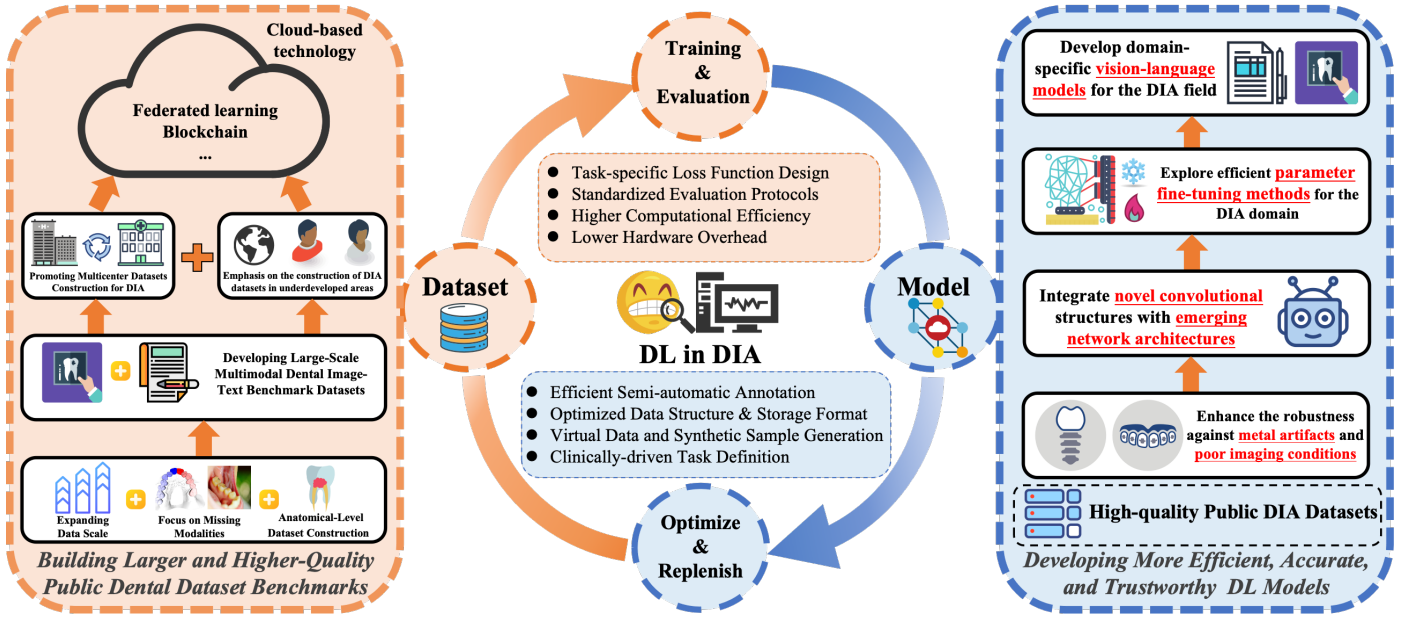


Figure 24: Overview of future research directions of DL in DIA. The framework is divided into two main parts: the left side focuses on datasets, and the right side focuses on models. These two aspects are complementary and mutually reinforcing, jointly advancing the intelligence and clinical applicability of DIA.

ture representation and complementary performance; however, pure ViT models, lacking local sensitivity, exhibit relatively limited performance in DIA tasks.

In model designs, incorporating various attention mechanisms to enhance feature representation has become a common optimization strategy. Some researchers have also attempted to integrate expert prior knowledge into the training process, explicitly improving the model’s ability to recognize and differentiate specific dental target structures. Overall, deep learning has achieved remarkable progress in core tasks such as tooth segmentation. For instance, in CBCT-based tooth identification, the average DSC of existing methods has exceeded 94%; in IOS-based tooth segmentation, the average DSC can reach over 95%; and even in PAN images, which are more susceptible to artifacts and low resolution, the instance-level tooth segmentation performance generally achieves an average DSC above 90%. These results clearly demonstrate that DL-based methods have reached a high level of performance in DIA; however, further improvements will depend on exploring more generalizable model architectures with stronger multimodal fusion capabilities.

Dentin, enamel, and dental pulp are the three primary internal structures of the tooth and represent core targets in anatomy-level prediction tasks. For the segmentation and recognition of such fine-grained structures, most studies continue to adopt CNN-based architectures, often combined with multi-scale feature extraction modules to improve prediction accuracy. Due to the lack of high-quality publicly available anatomy-level datasets, most current research relies on private data for experimentation, result-

ing in considerable variability among studies, with average DSC values ranging from 85% to 95%. In dental lesion analysis, the IP modality has seen a notable rise in application in recent years due to its ability to visually present the oral condition. Among various lesions, dental plaque and caries are the most intensively studied. To more effectively differentiate normal tissues from lesion regions, some studies, in addition to conventional attention mechanisms, have introduced contrastive learning approaches to enhance inter-class discriminability from the perspective of feature space.

In maxillofacial structure prediction tasks, CBCT is the most widely used imaging modality, and most related studies continue to rely on CNN-based architectures. Overall, high segmentation accuracy has been achieved for structures such as the temporomandibular joint, maxillary sinus, and alveolar bone, with average DSC values typically exceeding 97%. However, for finer anatomical structures such as the mandibular canal and inferior alveolar nerve, segmentation accuracy remains relatively low due to their subtle morphology and indistinct boundaries, with average DSC values ranging from 75% to 85%, indicating room for further improvement.

For classification or other non-dense network tasks, existing studies can generally be categorized into two types. The first type leverages dense prediction results as prior information to facilitate tasks such as multi-lesion or tissue classification, implant planning, and tooth surface reconstruction. The second type incorporates tooth-number classification as an auxiliary branch to enhance the precision of instance-level dense detection. Overall, studies fo-

cusing solely on classification tasks are relatively scarce, as most works tend to integrate classification and segmentation, thereby exploiting the advantages of multi-task learning in feature sharing and structural constraint.

In terms of model development and performance optimization, DIA still faces multiple challenges. Firstly, dental imaging is often affected by suboptimal acquisition angles, limited radiation doses, and the presence of dental fillings or implants, which can introduce noise or metal artifacts and interfere with image interpretation. Consequently, existing models require enhanced robustness when handling noisy inputs. Secondly, model diversity in the DIA field remains insufficient. According to recent surveys, deep learning applications in DIA are still primarily based on conventional CNN architectures, while the potential of emerging CNN paradigms [Liu et al. \(2023b\)](#); [Xu et al. \(2025b\)](#) and innovative architectures comparable to ViT (e.g., Mamba [Liu et al. \(2024a\)](#) or RWKV [Duan et al. \(2025\)](#)) has yet to be fully explored in dental scenarios. Efficiently leveraging these new models, or integrating different architectures to exploit their complementary strengths, remains a crucial direction for future research. Furthermore, some foundation models trained on ultra-large-scale medical imaging datasets (e.g., MedSAM [Ma et al. \(2024a\)](#)) perform suboptimally on dental tasks, primarily due to the scarcity of dental data during training. Therefore, developing parameter-efficient fine-tuning strategies or semi-supervised transfer approaches under limited data conditions is of significant importance for improving the predictive accuracy of foundation models in dental applications. This approach not only reduces dependence on costly annotated data but also substantially enhances the practicality and generalizability of the models.

## 6.2. Future Directions

As shown in Fig. 24, we discuss the potential future research directions of DL in DIA from two perspectives: dataset construction and model development.

### From the Perspective of Public Datasets Construction:

- Develop larger-scale, high-quality public DIA benchmark datasets, promote cross-hospital and cross-regional collaboration, and leverage key technologies such as federated learning or blockchain to expand data sources while ensuring patient privacy and ethical compliance, thereby enhancing data diversity and coverage;
- Increase focus on establishing DIA databases in underdeveloped regions by leveraging cloud-based annotation and federated sharing systems, collaborating with local universities, medical institutions, and research centers, and involving clinicians from other regions to assist with data annotation and processing, thereby enabling efficient dataset construction;

- Efforts should be intensified to construct datasets for underrepresented modalities, such as IOS and PR. Although CBCT provides more comprehensive scanning information, its high cost and radiation dose prevent it from fully replacing other imaging modalities. Therefore, developing datasets for these relatively scarce modalities remains crucial;
- The construction of multimodal dental datasets should be given high priority. Currently, publicly available multimodal DIA datasets are scarce, and existing datasets are often limited to specific modalities and tasks, lacking diversity in both image modalities and text categories. Therefore, establishing a comprehensive large-scale multimodal benchmark dataset is of great significance and serves as a crucial foundation for developing large-scale dental VLMs.

### From the Perspective of Model Developments:

- More targeted attention mechanisms or loss functions should be developed, in combination with generative training strategies such as diffusion models, to improve the model’s ability to recognize images with poor shooting angles or metal artifacts. In addition, the model’s robustness across different regions, institutions, and imaging devices should be further enhanced to ensure stable and generalizable performance in diverse and complex clinical environments;
- The potential of novel architectures in DIA tasks should be thoroughly explored, focusing on how to effectively leverage new convolutional paradigms and organically combine them with other architectures to exploit complementary strengths. Under the constraints of efficient training and low deployment overhead, such efforts should aim to further improve model performance and generalization in DIA applications;
- Efficient parameter-efficient tuning models or adapters should be designed to fully leverage foundation models pretrained on large-scale medical datasets. By integrating lightweight fine-tuning techniques such as LoRA, domain-specific dental imaging knowledge can be effectively injected into the model, enabling efficient transfer and reuse of large foundation medical image analysis models for DIA tasks;
- Expert-level VLM architectures and training frameworks tailored for the DIA domain should be developed. Unlike other medical imaging fields, dental targets are often more complex, requiring models to possess finer-grained understanding of both visual and textual information. Moreover, the model should be capable of interpreting tooth numbering and integrating positional and structural features from different tooth locations to achieve comprehensive understanding of dental images.

## 7. Conclusion

Efficient, accurate, and reproducible DIA technology serves as the foundation and key enabler for computer-aided dental diagnosis. In this paper, we provide a comprehensive review of recent advances and research trends in the application of DL to DIA. Centered on two fundamental pillars of DL research—public datasets and model architectures—we systematically summarize 49 publicly available DIA datasets and 213 related studies. We begin by introducing the research background and major task types in DIA, followed by a comparative analysis of existing review literature and a detailed description of our literature retrieval and screening methodology. Section 2 presents the main dental imaging modalities, illustrates real-world clinical imaging equipment, and systematically reviews the current landscape of public DIA datasets, including detailed information and access methods to facilitate dataset reuse by researchers. In Section 3, we categorize existing research based on task type and model architecture, covering dense prediction, classification, and other tasks, as well as CNN-based, ViT-based, and hybrid models. For dense prediction tasks, we further divide studies into four subcategories targeting tooth-level, anatomical-level, pathological, and maxillofacial structure analysis. Summarizing the key features and innovations of each work, we also analyze representative models in terms of architecture and performance, discussing the relationships and potential transferability between different task types. Furthermore, we review commonly used loss functions, optimizers, GPU configurations, and performance metrics in DIA model training to provide practical experimental guidance. Finally, we discuss the current state, limitations, and future directions of DIA research from two perspectives: data and models. We hope this review will serve as a valuable reference for the DIA research community—offering systematic background knowledge to newcomers and fresh insights to experienced researchers. Looking ahead, with the establishment of more high-quality open datasets and the emergence of innovative algorithms, DIA technologies are expected to achieve higher levels of intelligence and greater clinical applicability.

## Acknowledgments

This work is partially supported by the National Natural Science Foundation (62272248), the Natural Science Foundation of Tianjin (23JCZDJC01010).

## References

- Abraham, N., Khan, N.M., 2019. A novel focal tversky loss function with improved attention u-net for lesion segmentation, in: 2019 IEEE 16th international symposium on biomedical imaging (ISBI 2019), IEEE. pp. 683–687.
- Ahn, J.S., Cho, Y.R., 2024. Weighted sparse convolution and transformer feature aggregation networks for 3d dental segmentation. *IEEE Access* .
- Alam, S.S., Ahad, A., Ahmed, S., Dudley, J., Farook, T.H., 2025. Using deep learning to segment impacted molar teeth from panoramic radiographs. *Digital Dentistry Journal* , 100007.
- Alfadley, A., Shujaat, S., Jamleh, A., Riaz, M., Aboalela, A.A., Ma, H., Orhan, K., 2024. Progress of artificial intelligence-driven solutions for automated segmentation of dental pulp space on cone-beam computed tomography images. a systematic review. *Journal of Endodontics* 50, 1221–1232.
- Ali, M., 2024. Dental x-ray panoramic semantic segmentation task. URL: <https://www.kaggle.com/datasets/mohamedali020/dental-x-raypanoramic-semanticsegmentation-task>. accessed: 2025-08-24; This dataset includes panoramic dental X-ray images with binary and multi-class segmentation masks.
- Almalki, A., Latecki, L.J., 2023. Self-supervised learning with masked image modeling for teeth numbering, detection of dental restorations, and instance segmentation in dental panoramic radiographs, in: *Proceedings of the IEEE/CVF winter conference on applications of computer vision*, pp. 5594–5603.
- Almalki, A., Latecki, L.J., 2024. Self-supervised learning with masked autoencoders for teeth segmentation from intra-oral 3d scans, in: *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 7820–7830.
- Aminhj, 2023. Tooth decay dataset. URL: <https://www.kaggle.com/datasets/aminhj/tooth-decay>. accessed: 2025-08-23.
- Amorim, P.H., Moraes, T.F., Da Silva, J.V., Pedrini, H., 2023. Mandible segmentation from ct and cbct images based on a patch-based convolutional neural network, in: *2023 International Conference on Machine Learning and Applications (ICMLA)*, IEEE. pp. 38–44.
- Arı, T., Bayrakdar, I.S., Çelik, "O., Bilgir, E., Kuran, A., Orhan, K., 2025. Evaluation of cone-beam computed tomography images with artificial intelligence. *Journal of Imaging Informatics in Medicine* , 1–10.
- Arian, M.S.H., Sifat, F.A., Ahmed, S., Mohammed, N., Farook, T.H., 2025. Unsupervised tooth segmentation from three dimensional scans of the dental arch using domain adaptation of synthetic data. *International Journal of Medical Informatics* 195, 105769.
- Arslan, A., Bilgin, G., 2025. Tooth segmentation and abnormal tooth detection with diagnostic criterion in

- panoramic x-ray images with deep learning approach, in: 2025 33rd Signal Processing and Communications Applications Conference (SIU), IEEE. pp. 1–4.
- Arya, K.V., 2024. A collection of dental x-ray images for analysis. URL: <https://www.kaggle.com/datasets/kvipularya/a-collection-of-dental-x-ray-images-for-analysis>. accessed: 2025-08-23; Contains dental X-ray images for oral health analysis.
- Aseretto, S.G., Noguera, J.L.V., 2022. Contrast enhancement of orthopantomograms to improve tooth segmentation using u-nets, in: 2022 XVI Latin American Computer Conference (CLEI), IEEE. pp. 1–8.
- ATAŞ, M., Özdemir, C., ATAŞ, İ., Ak, B., ÖZEROĞLU, E., 2022. Biometric identification using panoramic dental radiographic images with few-shot learning. *Turkish Journal of Electrical Engineering and Computer Sciences* 30, 1115–1126.
- Awari, H., Subramani, N., Janagaraj, A., Balasubramaniapillai Thanammal, G., Thangarasu, J., Kohar, R., 2024. Three-dimensional dental image segmentation and classification using deep learning with tunicate swarm algorithm. *Expert Systems* 41, e13198.
- Bajpai, S., Kim, P., Lee, B., 2025. Advanced teeth instance segmentation using contextual spatial attention and deeper mask head, in: 2025 IEEE International Conference on Consumer Electronics (ICCE), IEEE. pp. 1–4.
- Ben-Hamadou, A., Smaoui, O., Rekik, A., Pujades, S., Boyer, E., Lim, H., Kim, M., Lee, M., Chung, M., Shin, Y.G., Leclercq, M., Cevdanes, L., Prieto, J.C., Zhuang, S., Wei, G., Cui, Z., Zhou, Y., Dascalu, T., Ibragimov, B., Yong, T.H., Ahn, H.G., Kim, W., Han, J.H., Choi, B., van Nistelrooij, N., Kempers, S., Vinayahalingam, S., Strippoli, J., Thollot, A., Setbon, H., Trosset, C., Ladroit, E., 2023. 3dteethseg'22: 3d teeth scan segmentation and labeling challenge. URL: [https://github.com/abenhamadou/3DTeethSeg\\_MICCAI\\_Challenges](https://github.com/abenhamadou/3DTeethSeg_MICCAI_Challenges). accessed: 2025-08-23.
- Benzian, H., Beltrán-Aguilar, E., 2025. The global burden of oral diseases: stronger data for stronger action. *Lancet (London, England)* 405, 868–869.
- Berman, M., Triki, A.R., Blaschko, M.B., 2018. The lovász-softmax loss: A tractable surrogate for the optimization of the intersection-over-union measure in neural networks, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 4413–4421.
- Bernabe, E., Marcenes, W., Abdulkader, R.S., Abreu, L.G., Afzal, S., Alhalaiqa, F.N., Al-Maweri, S., Alsharif, U., Anyasodor, A.E., Arora, A., et al., 2025. Trends in the global, regional, and national burden of oral conditions from 1990 to 2021: a systematic analysis for the global burden of disease study 2021. *The Lancet* 405, 897–910.
- Beser, B., Reis, T., Berber, M.N., Topaloglu, E., Gungor, E., Kılıç, M.C., Duman, S., Çelik, Ö., Kuran, A., Bayrakdar, I.S., 2024. Yolo-v5 based deep learning approach for tooth detection and segmentation on pediatric panoramic radiographs in mixed dentition. *BMC medical imaging* 24, 172.
- Bolelli, F., Lumetti, L., Vinayahalingam, S., Di Bartolomeo, M., Pellacani, A., Marchesini, K., Van Nistelrooij, N., Van Lierop, P., Xi, T., Liu, Y., et al., 2024. Segmenting the inferior alveolar canal in cbct volumes: the toothfairy challenge. *IEEE Transactions on Medical Imaging*.
- Bolelli, F., Marchesini, K., van Nistelrooij, N., Lumetti, L., Pipoli, V., Ficarra, E., Vinayahalingam, S., Grana, C., 2025. Segmenting maxillofacial structures in cbct volumes, in: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 5238–5248.
- Boudraa, O., 2022. Segmentation of 3d dental images using deep learning. *arXiv preprint arXiv:2207.09582*.
- Brahmi, W., Jdey, I., 2023. Panoramic dental x-ray dataset (version 2). <https://data.mendeley.com/datasets/73n3kz2k4k/2>. DOI: 10.17632/73n3kz2k4k.2; CC BY 4.0; Université de Kairouan; Accessed: 2025-08-23.
- Brahmi, W., Jdey, I., 2024. Automatic tooth instance segmentation and identification from panoramic x-ray images using deep cnn. *Multimedia Tools and Applications* 83, 55565–55585.
- Cai, X., Zhang, Y., Guan, Z., Sun, Q., Qu, Z., 2024. Deformable inherent consistent learning network for accurate tooth segmentation in dental panoramic radiographs, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 166–177.
- Cao, D., Cai, B., Liu, M., 2024. Flowgateunet: Dental ct image segmentation network based on flowformer and gated attention. *Signal, Image and Video Processing* 18, 1175–1182.
- Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M., 2022. Swin-unet: Unet-like pure transformer for medical image segmentation, in: European conference on computer vision, Springer. pp. 205–218.
- Cao, K., Tian, L., Li, Q., Chen, H., Li, C., Fan, Y., Ye, J., Yu, W., 2023. A stagewise deep learning framework for tooth instance segmentation in cbct images, in: Pacific Rim International Conference on Artificial Intelligence, Springer. pp. 415–425.

- Caylak, T., Yetik, I.S., Culhaoglu, A., Orhan, K., Kiliarslan, M.A., 2022. Automated dental panoramic image segmentation using transfer learning based cnns, in: 2022 29th International Conference on Systems, Signals and Image Processing (IWSSIP), IEEE. pp. 1–4.
- Chaitanya, N.C., Padmanabhan, V., Islam, M.S., Hashim, N.T., Mohammed, R., Allam, N.S., Omar, J.J., Zakaria, A., Pranathi, S., Apoorva, M., 2024. Enhancing intraoral radiographic technique: introducing “nallan’s lines”. *Frontiers in Oral Health* 5, 1498117.
- Chandrashekar, G., AlQarni, S., Bumann, E.E., Lee, Y., 2022. Collaborative deep learning model for tooth segmentation and identification using panoramic radiographs. *Computers in Biology and Medicine* 148, 105829.
- Chao, P., Kao, C.Y., Ruan, Y.S., Huang, C.H., Lin, Y.L., 2019. Hardnet: A low memory traffic network, in: Proceedings of the IEEE/CVF international conference on computer vision, pp. 3552–3561.
- Chauhan, R.B., Shah, T.V., Shah, D.H., Gohil, T.J., Oza, A.D., Jajal, B., Saxena, K.K., 2023. An overview of image processing for dental diagnosis. *Innovation and Emerging Technologies* 10, 2330001.
- Chen, H., Ge, Y., Wei, J., Xiong, H., Liu, Z., 2022. Tooth defect segmentation in 3d mesh scans using deep learning, in: CAAI International Conference on Artificial Intelligence, Springer. pp. 180–191.
- Chen, H., Huang, X., Li, Q., Wang, J., Fang, B., Chen, J., 2023. Labanet: Lead-assisting backbone attention network for oral multi-pathology segmentation, in: ICASSP 2023-2023 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), IEEE. pp. 1–5.
- Chen, J., 2025. Convolutional neural network for maxillary sinus segmentation based on the u-net architecture at different planes in the chinese population: a semantic segmentation study. *BMC Oral Health* 25, 961.
- Chen, J., Gui, C., Ouyang, R., Gao, A., Chen, S., Chen, G.H., Wang, X., Zhang, R., Cai, Z., Ji, K., et al., 2024a. Huatuogpt-vision, towards injecting medical visual knowledge into multimodal llms at scale. *arXiv preprint arXiv:2406.19280*.
- Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y., 2021a. Transunet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*.
- Chen, J., Ye, Z., Zhang, R., Li, H., Fang, B., Zhang, L.b., Wang, W., 2025a. Medical image translation with deep learning: Advances, datasets and perspectives. *Medical Image Analysis*, 103605.
- Chen, L., Zhang, D., Yan, T., Li, Z., Wei, Y., Qian, L., 2024b. Dae-net: Dual attention embedding-based tooth instance segmentation approach for panoramic x-ray images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 146–155.
- Chen, Q., Zhao, Y., Liu, Y., Sun, Y., Yang, C., Li, P., Zhang, L., Gao, C., 2021b. Mslpnet: multi-scale location perception network for dental panoramic x-ray image segmentation. *Neural Computing and Applications* 33, 10277–10291.
- Chen, R., Yang, J., Xiong, H., Xu, R., Feng, Y., Wu, J., Liu, Z., 2025b. Cross-center model adaptive tooth segmentation. *Medical Image Analysis* 101, 103443.
- Chen, Y., Wang, C., Zhao, B., 2024c. A novel two-stage approach for 3d dental tooth instance segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 188–199.
- Chen, Z., Liu, Q., Wang, J., Ji, N., Gong, Y., Gao, B., 2025c. Tooth image segmentation and root canal measurement based on deep learning. *Frontiers in Bioengineering and Biotechnology* 13, 1565403.
- Choi, J., Eun, H., Kim, C., 2018. Boosting proximal dental caries detection via combination of variational methods and convolutional neural network. *Journal of Signal Processing Systems* 90, 87–97.
- Çiçek, Ö., Abdulkadir, A., Lienkamp, S.S., Brox, T., Ronneberger, O., 2016. 3D U-Net: Learning Dense Volumetric Segmentation from Sparse Annotation. *arXiv:1606.06650*.
- Cipriano, M., Allegretti, S., Bolelli, F., Di Bartolomeo, M., Pollastri, F., Pellacani, A., Minafra, P., Anesi, A., Grana, C., 2022. Deep segmentation of the mandibular canal: a new 3d annotated dataset of cbct volumes. *Ieee Access* 10, 11500–11510.
- Cui, W., Wang, Y., Li, Y., Song, D., Zuo, X., Wang, J., Zhang, Y., Zhou, H., Chong, B.s., Zeng, L., et al., 2022a. Ctooth+: a large-scale dental cone beam computed tomography dataset and benchmark for tooth volume segmentation, in: MICCAI Workshop on Data Augmentation, Labelling, and Imperfections, Springer. pp. 64–73.
- Cui, W., Wang, Y., Zhang, Q., Zhou, H., Song, D., Zuo, X., Jia, G., Zeng, L., 2022b. Ctooth: a fully annotated 3d dataset and benchmark for tooth volume segmentation on cone beam computed tomography images, in: International Conference on Intelligent Robotics and Applications, Springer. pp. 191–200.
- Cui, W., Zeng, L., Chong, B., Zhang, Q., 2021a. Toothpix: Pixel-level tooth segmentation in panoramic x-ray

- images based on generative adversarial networks, in: 2021 IEEE 18th international symposium on biomedical imaging (ISBI), IEEE. pp. 1346–1350.
- Cui, Z., Fang, Y., Mei, L., Zhang, B., Yu, B., Liu, J., Jiang, C., Sun, Y., Ma, L., Huang, J., et al., 2022c. A fully automatic ai system for tooth and alveolar bone segmentation from cone-beam ct images. *Nature communications* 13, 2096.
- Cui, Z., Li, C., Chen, N., Wei, G., Chen, R., Zhou, Y., Shen, D., Wang, W., 2021b. Tsegnet: An efficient and accurate tooth segmentation network on 3d dental model. *Medical Image Analysis* 69, 101949.
- Cui, Z., Li, C., Wang, W., 2019. Toothnet: automatic tooth instance segmentation and identification from cone beam ct images, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 6368–6377.
- Dai, P., Ou, Y., Yang, Y., Liu, Y., Zhao, Y., 2024. Sparse anatomical prompt semi-supervised learning with masked image modeling for cbct tooth segmentation, in: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE. pp. 1–5.
- Dasanayaka, C., Dharmasena, B., Bandara, W.R., Disanayake, M.B., Jayasinghe, R., 2019. Segmentation of mental foramen in dental panoramic tomography using deep learning, in: *2019 14th conference on industrial and information systems (ICIIS)*, IEEE. pp. 81–84.
- Dong, W., You, M., He, T., Dai, J., Tang, Y., Shi, Y., Guo, J., 2023. An automatic methodology for full dentition maturity staging from opg images using deep learning. *Applied Intelligence* 53, 29514–29536.
- Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al., 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.
- Du, G., Tian, X., Song, Y., 2022. Mandibular canal segmentation from cbct image using 3d convolutional neural network with scse attention. *Ieee Access* 10, 111272–111283.
- Duan, F., Chen, L., 2023. 3d dental mesh segmentation using semantics-based feature learning with graph-transformer, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 456–465.
- Duan, Y., Wang, W., Chen, Z., Zhu, X., Lu, L., Lu, T., Qiao, Y., Li, H., Dai, J., Wang, W., 2025. Vision-RWKV: Efficient and Scalable Visual Perception with RWKV-Like Architectures. doi:10.48550/arXiv.2403.02308, *arXiv:2403.02308*.
- Dutra, K.L., Haas, L., Porporatti, A.L., Flores-Mir, C., Santos, J.N., Mezzomo, L.A., Correa, M., Canto, G.D.L., 2016. Diagnostic accuracy of cone-beam computed tomography and conventional radiography on apical periodontitis: a systematic review and meta-analysis. *Journal of endodontics* 42, 356–364.
- Eggmann, F., Blatz, M., 2024. Recent advances in intraoral scanners. *Journal of Dental Research* 103, 1349–1357.
- Eun, H., Kim, C., 2016. Oriented tooth localization for periapical dental x-ray images via convolutional neural network, in: *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, IEEE. pp. 1–7.
- Faizan Ahmed, S.M., Ghori, M.H., Khalid, A., Nooruddin, A., Adnan, N., Lal, A., Umer, F., 2025. Annotated intraoral image dataset for dental caries detection. *Scientific data* 12, 1297.
- Fang, E.P., Liew, D.J., Chang, Y.C., Fang, C.Y., 2025. Enhancing wisdom teeth detection in panoramic radiographs using multi-channel convolutional neural network with clinical knowledge. *Computers in Biology and Medicine* 192, 110368.
- Fedorov, A., Beichel, R., Kalpathy-Cramer, J., Finet, J., Fillion-Robin, J.C., Pujol, S., Bauer, C., Jennings, D., Fennessy, F., Sonka, M., et al., 2012. 3d slicer as an image computing platform for the quantitative imaging network. *Magnetic resonance imaging* 30, 1323–1341.
- Felsch, M., Meyer, O., Schlickerrieder, A., Engels, P., Schönewolf, J., Zöllner, F., Heinrich-Weltzien, R., Hesenius, M., Hickel, R., Gruhn, V., et al., 2023. Detection and localization of caries and hypomineralization on dental photographs with a vision transformer model. *NPJ digital medicine* 6, 198.
- Feng, W., 2023. Automated dental cbct segmentation using pseudo labeling method, in: *MICCAI Challenge on Semi-supervised Tooth Segmentation*. Springer, pp. 156–168.
- Feng, X., Wang, C., Wu, C., Li, Y., He, Y., Wang, S., Wang, Y., 2024. Fdnet: Feature decoupled segmentation network for tooth cbct image, in: *2024 IEEE International Symposium on Biomedical Imaging (ISBI)*, IEEE. pp. 1–5.
- Gallagher, J.E., Mattos Savage, G.C., Crummey, S.C., Sabbah, W., Makino, Y., Varenne, B., 2024. Health workforce for oral health inequity: Opportunity for action. *PLoS One* 19, e0292549.
- Gamal, M., Baraka, M., Torki, M., 2024. Automatic mandibular semantic segmentation of teeth pulp cavity and root canals, and inferior alveolar nerve on pulpy3d

- dataset, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 14–23.
- Gan, J., Yu, N., Qian, G., He, N., 2023. Concrete learning method for segmentation and denoising using cbct image, in: Proceedings of the 2023 4th International Conference on Control, Robotics and Intelligent System, pp. 41–46.
- Ge, Z., Liu, S., Wang, F., Li, Z., Sun, J., 2021. Yolox: Exceeding yolo series in 2021. arXiv preprint arXiv:2107.08430 .
- Ghafoor, A., Lee, B., 2025. Enhancing multiclass teeth segmentation with advanced attention mechanisms, in: 2025 International Conference on Artificial Intelligence in Information and Communication (ICAIIIC), IEEE. pp. 0545–0549.
- Ghafoor, A., Moon, S.Y., Lee, B., 2023. Multiclass segmentation using teeth attention modules for dental x-ray images. IEEE Access 11, 123891–123903.
- Gou, M., Rao, Y., Zhang, M., Sun, J., Cheng, K., 2019. Automatic image annotation and deep learning for tooth ct image segmentation, in: International Conference on Image and Graphics, Springer. pp. 519–528.
- Gu, D., Gao, Y., Zhou, Y., Zhou, M., Metaxas, D., . RadAlign: Advancing Radiology Report Generation with Vision-Language Concept Alignment. URL: <http://arxiv.org/abs/2501.07525>, doi:10.48550/arXiv.2501.07525, arXiv:2501.07525.
- Guo, S., Chen, Y., Mallineni, S.K., Huang, S., Liu, B., Zhang, S., Lu, C., 2021. Feasibility of oral health evaluation by intraoral digital photography: a pilot study. Journal of International Medical Research 49, 0300060520982841.
- Guo, X., Liu, W., Cui, Z., 2024. Efficient semi-supervised tooth instance segmentation in panoramic x-rays using resnet50 and sam networks, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 131–145.
- Hadzic, A., Kirnbauer, B., Stern, D., Urschler, M., 2023. Teeth localization and lesion segmentation in cbct images using spatialconfiguration-net and u-net. arXiv preprint arXiv:2312.12189 .
- Haghanifar, A., Majdabadi, M.M., Haghanifar, S., Choi, Y., Ko, S.B., 2023. Paxnet: Tooth segmentation and dental caries detection in panoramic x-ray using ensemble transfer learning and capsule classifier. Multimedia Tools and Applications 82, 27659–27679.
- Hamamci, I.E., Er, S., Simsar, E., Yuksel, A.E., Gul- tekin, S., Ozdemir, S.D., Yang, K., Li, H.B., Pati, S., Stadlinger, B., et al., 2023. Dentex: An abnormal tooth detection with dental enumeration and diagnosis benchmark for panoramic x-rays. arXiv preprint arXiv:2305.19112 .
- Hameed, T., AmalaShanthi, S., 2024. A novel teeth segmentation on three-dimensional dental model using adaptive enhanced googlenet classifier. Multimedia Tools and Applications 83, 68547–68568.
- Han, J.H., Kim, W., Ahn, H.G., 2024. Weakly-supervised convolutional neural networks for inferior alveolar nerve segmentation in cbct images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 41–48.
- Harsh, P., Chakraborty, R., Tripathi, S., Sharma, K., 2021. Attention u-net architecture for dental image segmentation, in: 2021 International Conference on Intelligent Technologies (CONIT), IEEE. pp. 1–5.
- He, A., Wang, K., Li, T., Bo, W., Kang, H., Fu, H., 2022. Progressive multiscale consistent network for multiclass fundus lesion segmentation. IEEE transactions on medical imaging 41, 3146–3157.
- He, A., Wang, K., Li, T., Du, C., Xia, S., Fu, H., 2023. H2former: An efficient hierarchical hybrid transformer for medical image segmentation. IEEE Transactions on Medical Imaging 42, 2763–2775.
- He, K., Zhang, X., Ren, S., Sun, J., 2016. Deep residual learning for image recognition, in: Proceedings of the IEEE conference on computer vision and pattern recognition, pp. 770–778.
- Hong, Y., Kim, J., Chen, G., Lin, W., Yap, P.T., Shen, D., 2019. Longitudinal prediction of infant diffusion mri data via graph convolutional adversarial networks. IEEE transactions on medical imaging 38, 2717–2725.
- Hou, S., Zhou, T., Liu, Y., Dang, P., Lu, H., Shi, H., 2023. Teeth u-net: A segmentation model of dental panoramic x-ray images for context semantics and contrast enhancement. Computers in Biology and Medicine 152, 106296.
- Hu, P., Li, J., Ma, R., Zhang, K., Guo, Y., Li, G., 2025. Temporomandibular joint cbct image segmentation via multi-view ensemble learning network. Medical & Biological Engineering & Computing 63, 693–706.
- Huang, F., Yin, J., Ma, Y., Zhang, H., Ying, S., 2025a. Precise dental caries segmentation in x-rays with an attention and edge dual-decoder network. Medical & Biological Engineering & Computing , 1–18.
- Huang, J., Yan, H., Li, J., Stewart, H.M., Setzer, F., 2021a. Combining anatomical constraints and deep learning for 3-d cbct dental image multi-label segmentation, in: 2021 IEEE 37th International Conference on Data Engineering (ICDE), IEEE. pp. 2750–2755.

- Huang, X., Deng, Z., Li, D., Yuan, X., 2021b. Missformer: An effective medical image segmentation transformer. arXiv preprint arXiv:2109.07162 .
- Huang, Y., Liu, W., Yao, C., Miao, X., Guan, X., Lu, X., Liang, X., Ma, L., Tang, S., Zhang, Z., et al., 2024. A multimodal dental dataset facilitating machine learning research and clinic services. *Scientific Data* 11, 1291.
- Huang, Y., Tang, Y., Wu, J., He, J., Wang, W., 2025b. Automatic detection and segmentation of tooth cracks based on improved mask r-cnn. *IEEE Journal of Biomedical and Health Informatics* .
- Hussein, A., Güneç, H.G., Aydin, K.C., Ates, H.F., 2025. Deep learning based tooth multi-disease detection in dental diagnostics, in: 2025 33rd Signal Processing and Communications Applications Conference (SIU), IEEE. pp. 1–4.
- Ibtehaz, N., Kihara, D., 2023. Acc-unet: A completely convolutional unet model for the 2020s, in: International conference on medical image computing and computer-assisted intervention, Springer. pp. 692–702.
- Isensee, F., Kirchhoff, Y., Kraemer, L., Rokuss, M., Ulrich, C., Maier-Hein, K.H., 2024. Scaling nnu-net for cbct segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 13–20.
- Jana, A., Maiti, A., Metaxas, D.N., 2023a. A critical analysis of the limitation of deep learning based 3d dental mesh segmentation methods in segmenting partial scans, in: 2023 45th Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC), IEEE. pp. 1–7.
- Jana, A., Subhash, H.M., Metaxas, D., 2023b. Automatic tooth segmentation from 3d dental model using deep learning: a quantitative analysis of what can be learnt from a single 3d dental model, in: 18th International Symposium on Medical Information Processing and Analysis, SPIE. pp. 42–51.
- Jang, T.J., Kim, K.C., Cho, H.C., Seo, J.K., 2021. A fully automated method for 3d individual tooth identification and segmentation in dental cbct. *IEEE transactions on pattern analysis and machine intelligence* 44, 6562–6568.
- Jang, T.J., Yun, H.S., Hyun, C.M., Kim, J.E., Lee, S.H., Seo, J.K., 2024. Fully automatic integration of dental cbct images and full-arch intraoral impressions with stitching error correction via individual tooth segmentation and identification. *Medical Image Analysis* 93, 103096.
- Järnstedt, J., Sahlsten, J., Jaskari, J., Kaski, K., Mehtonen, H., Lin, Z., Hietanen, A., Sundqvist, O., Varjonen, V., Mattila, V., et al., 2022. Comparison of deep learning segmentation and multigrader-annotated mandibular canals of multicenter cbct scans. *Scientific Reports* 12, 18598.
- Ji, C., Liu, Y., He, L., Jiang, Y., Huang, C., Wang, L., 2024. A two-stage semi-supervised nnu-net model for automated tooth segmentation in panoramic x-ray images, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 91–99.
- Jiang, Y., Song, W., Li, S., Yang, Y., Xia, B., Hao, A., Qin, H., 2024. Frequency-guided network for low-contrast staining-free dental plaque segmentation, in: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 4925–4932.
- Jing, Y., Liu, J., Liu, W., Yang, Z., Zhou, Z., Yu, Z., 2024. Usct: Uncertainty-regularized symmetric consistency learning for semi-supervised teeth segmentation in cbct. *Biomedical Signal Processing and Control* 91, 106032.
- Johnson, A.E.W., Pollard, T.J., Greenbaum, N.R., Lungren, M.P., Deng, C.y., Peng, Y., Lu, Z., Mark, R.G., Berkowitz, S.J., Horng, S., 2019. MIMIC-CXR-JPG, a large publicly available database of labeled chest radiographs. doi:10.48550/arXiv.1901.07042, arXiv:1901.07042.
- Jones, B., Lambach, M., Chen, T., Michou, S., Kilpatrick, N., Curtis, N., Burgner, D.P., Vannahme, C., Silva, M., 2025. Dental caries detection in children using intraoral scans and deep learning. *Journal of Dentistry* , 105906.
- Kaasalainen, T., Ekholm, M., Siiskonen, T., Kortensniemi, M., 2021. Dental cone beam ct: An updated review. *Physica Medica* 88, 193–217.
- Kaggle, 2019. Dental x-ray images dataset. <https://www.kaggle.com/datasets/kill92/dental-x-ray-images-dataset>. Accessed: 2025-08-23.
- Karacan, M.H., Yücebas, S.C., 2022. A deep learning model with attention mechanism for dental image segmentation, in: 2022 International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA), IEEE. pp. 1–4.
- Katsaros, E., Ostrowski, P.K., Włodarczak, K., Lewandowska, E., Ruminski, J., Siupka-Mróż, D., Lassmann, Ł., Jezierska, A., Węsierski, D., 2022. Multi-task video enhancement for dental interventions, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 177–187.
- Kibcak, E., Buhara, O., Temelci, A., Akkaya, N., Ünsal, G., Minervini, G., 2025. Deep learning-driven segmentation of dental implants and peri-implantitis detection in

- orthopantomographs: a novel diagnostic tool. *Journal of Evidence-Based Dental Practice* 25, 102058.
- Kim, S., Lee, C.O., 2024. Individual tooth segmentation in human teeth images using pseudo edge-region obtained by deep neural networks. *Signal Processing: Image Communication* 120, 117076.
- Kim, S., Song, I.S., Baek, S.J., 2023. Automatic segmentation of internal tooth structure from cbct images using hierarchical deep learning, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 703–713.
- Kingma, D.P., Ba, J., 2015. Adam: A method for stochastic optimization. *International Conference on Learning Representations (ICLR)* .
- Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A.C., Lo, W.Y., et al., 2023. Segment anything, in: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4015–4026.
- Koch, T.L., Perslev, M., Igel, C., Brandt, S.S., 2019. Accurate segmentation of dental panoramic radiographs with u-nets, in: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, IEEE. pp. 15–19.
- Kong, Z., Xiong, F., Zhang, C., Fu, Z., Zhang, M., Weng, J., Fan, M., 2020. Automated maxillofacial segmentation in panoramic dental x-ray images using an efficient encoder-decoder network. *Ieee Access* 8, 207822–207833.
- Koshy, L., Benny, A., 2024. Dental cavity detection using u-netarchitecture, in: *2024 4th International Conference on Sustainable Expert Systems (ICSSES)*, IEEE. pp. 977–982.
- Kot, W.Y., Au Yeung, S.Y., Leung, Y.Y., Leung, P.H., Yang, W.f., 2025. Evolution of deep learning tooth segmentation from ct/cbct images: a systematic review and meta-analysis. *BMC Oral Health* 25, 800.
- Krizhevsky, A., Sutskever, I., Hinton, G.E., 2012. Imagenet classification with deep convolutional neural networks. *Advances in neural information processing systems* 25.
- Kulkarni, N., Kharad, Y., 2024. A comparative evaluation of patient and orthodontist preferences and attitude in intraoral photography using a novel tongue retractor as an adjunct for occlusal photographic mirror vs. conventional techniques for capturing mandibular occlusal photographs in orthodontic practice. *Library of Progress-Library Science, Information Technology & Computer* 44.
- Kumari, A.R., Rao, S.N., Reddy, P.R., 2022. Design of hybrid dental caries segmentation and caries detection with meta-heuristic-based resnext-rnn. *Biomedical Signal Processing and Control* 78, 103961.
- Lai, Y., Fan, F., Wu, Q., Ke, W., Liao, P., Deng, Z., Chen, H., Zhang, Y., 2020. Lcanet: learnable connected attention network for human identification using dental images. *IEEE Transactions on Medical Imaging* 40, 905–915.
- Lakshmi, M.M., Chitra, P., 2020. Classification of dental cavities from x-ray images using deep cnn algorithm, in: *2020 4th International Conference on Trends in Electronics and Informatics (ICOEI)*(48184), IEEE. pp. 774–779.
- Lang, Y., Deng, H.H., Xiao, D., Lian, C., Kuang, T., Gateno, J., Yap, P.T., Xia, J.J., 2021. Dllnet: an attention-based deep learning method for dental landmark localization on high-resolution 3d digital dental models, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 478–487.
- Lang, Y., Lian, C., Xiao, D., Deng, H., Thung, K.H., Yuan, P., Gateno, J., Kuang, T., Alfi, D.M., Wang, L., et al., 2022. Localization of craniomaxillofacial landmarks on cbct images using 3d mask r-cnn and local dependency learning. *IEEE transactions on medical imaging* 41, 2856–2866.
- Leclercq, M., Ruellas, A., Gurgel, M., Yatabe, M., Bianchi, J., Cevdanes, L., Styner, M., Paniagua, B., Prieto, J.C., 2023. Dentalmodelseg: fully automated segmentation of upper and lower 3d intra-oral surfaces, in: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, IEEE. pp. 1–5.
- LeCun, Y., Bengio, Y., Hinton, G., 2015. Deep learning. *nature* 521, 436–444.
- LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., 2002. Gradient-based learning applied to document recognition. *Proceedings of the IEEE* 86, 2278–2324.
- Lee, S., Woo, S., Yu, J., Seo, J., Lee, J., Lee, C., 2020. Automated cnn-based tooth segmentation in cone-beam ct for dental implant planning. *IEEE access* 8, 50507–50518.
- Li, D., Zhu, M., Wang, S., Hu, Y., Yuan, F., Yu, J., 2024a. Accurate and automatic dental crown components segmentation with multi-scale attention based u-net and hybrid level set models. *IEEE Transactions on Automation Science and Engineering* .
- Li, H., Liu, H., Fu, H., Xu, Y., Shu, H., Niu, K., Hu, Y., Liu, J., 2023a. A generic fundus image enhancement network boosted by frequency self-supervised representation learning. *Medical Image Analysis* 90, 102945.
- Li, J., Cheng, B., Niu, N., Gao, G., Ying, S., Shi, J., Zeng, T., 2024b. A fine-grained orthodontics segmentation model for 3d intraoral scan data. *Computers in biology and medicine* 168, 107821.

- Li, K., Zhu, J., Cui, Z., Chen, X., Liu, Y., Wang, F., Zhao, Y., 2024c. A novel hierarchical cross-stream aggregation neural network for semantic segmentation of 3-d dental surface models. *IEEE Transactions on Neural Networks and Learning Systems* 36, 7382–7394.
- Li, L., Zhou, T., Wang, W., Li, J., Yang, Y., 2022a. Deep hierarchical semantic segmentation, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1246–1257.
- Li, M., Yan, Z., Lu, Q., Ma, Q., Guo, L., Zhang, Q., 2023b. A semi-supervised tooth segmentation method based on entropy-guided mean teacher and weakly mutual consistency network, in: *MICCAI Challenge on Semi-supervised Tooth Segmentation*. Springer, pp. 58–71.
- Li, P., Gao, C., Lian, C., Meng, D., 2024d. Spatial prior-guided bi-directional cross-attention transformers for tooth instance segmentation. *IEEE Transactions on Medical Imaging* 43, 3936–3948.
- Li, P., Gao, C., Liu, F., Meng, D., Yan, Y., 2023c. Thisnet: Tooth instance segmentation on 3d dental models via highlighting tooth regions. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 5229–5241.
- Li, P., Liu, Y., Cui, Z., Yang, F., Zhao, Y., Lian, C., Gao, C., 2022b. Semantic graph attention with explicit anatomical association modeling for tooth segmentation from cbct images. *IEEE Transactions on Medical Imaging* 41, 3116–3127.
- Li, S., Guo, Y., Pang, Z., Song, W., Hao, A., Xia, B., Qin, H., 2022c. Automatic dental plaque segmentation based on local-to-global features fused self-attention network. *IEEE Journal of Biomedical and Health Informatics* 26, 2240–2251.
- Li, S., Li, C., Du, Y., Ye, L., Fang, Y., Wang, C., Zhou, W., 2023d. Transformer-based tooth segmentation, identification and pulp calcification recognition in cbct, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 706–714.
- Li, S., Pang, Z., Song, W., Guo, Y., You, W., Hao, A., Qin, H., 2020. Low-shot learning of automatic dental plaque segmentation based on local-to-global feature fusion, in: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE. pp. 664–668.
- Li, T., Gao, Y., Wang, K., Guo, S., Liu, H., Kang, H., 2019. Diagnostic assessment of deep learning algorithms for diabetic retinopathy screening. *Information Sciences* 501, 511–522.
- Li, X., Chung, M., Bi, L., Huang, J., Pan, Y., Feng, D., Chen, D., Sheng, B., Kim, J., Jiang, L., 2025a. Teanet: Automated tooth extraction and arrangement with tooth-level graph spatial transformation network. *Computer Methods and Programs in Biomedicine*, 108985.
- Li, X., Liu, W., Tang, W., Guo, J., 2025b. Dense image-mask attention-guided transformer network for jaw lesions classification and segmentation in dental cone-beam computed tomography images. *Applied Intelligence* 55, 499.
- Li, X., Ma, X., Zhao, Y., Hu, J., Liu, J., Yang, Z., Han, F., Zhang, J., Liu, W., Zhou, Z., 2024e. A multi-center dental panoramic radiography image dataset for impacted teeth, periodontitis, and dental caries: Benchmarking segmentation and classification tasks. *Journal of Imaging Informatics in Medicine* 37, 831–841.
- Li, X., Morgan, P.S., Ashburner, J., Smith, J., Rorden, C., 2016. The first step for neuroimaging data analysis: Dicom to nifti conversion. *Journal of neuroscience methods* 264, 47–56.
- Li, Y., Wang, S., Wang, J., Zeng, G., Liu, W., Zhang, Q., Jin, Q., Wang, Y., 2021. Gt u-net: A u-net like group transformer network for tooth root segmentation, in: *International workshop on machine learning in medical imaging*, Springer. pp. 386–395.
- Li, Z., Liu, T., Wang, J., Zhang, C., Jia, X., 2022d. Multi-scale bidirectional enhancement network for 3d dental model segmentation, in: *2022 IEEE 19th International Symposium on Biomedical Imaging (ISBI)*, IEEE. pp. 1–5.
- Liang, J., Wang, R., Rao, S., Xu, F., Xiang, J., Wang, B., Yan, T., 2024. Dual-view dual-boundary dual u-nets for multiscale segmentation of oral cbct images, in: *Chinese Conference on Pattern Recognition and Computer Vision (PRCV)*, Springer. pp. 48–62.
- Lin, J., Dai, Q., Zhu, L., Fu, H., Wang, Q., Li, W., Rao, W., Huang, X., Wang, L., 2023a. Shifting more attention to breast lesion segmentation in ultrasound videos, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 497–507.
- Lin, S., Hao, X., Liu, Y., Yan, D., Liu, J., Zhong, M., 2023b. Lightweight deep learning methods for panoramic dental x-ray image segmentation. *Neural Computing and Applications* 35, 8295–8306.
- Lin, Z., He, Z., Wang, X., Zhang, B., Liu, C., Su, W., Tan, J., Xie, S., 2023c. Dbganet: Dual-branch geometric attention network for accurate 3d tooth segmentation. *IEEE Transactions on Circuits and Systems for Video Technology* 34, 4285–4298.

- Listl, S., Bärnighausen, T., 2025. Oral health as a marker of social progress. *Journal of Dental Research*, 00220345251324120.
- Liu, D., Tian, Y., Zhang, Y., Gelernter, J., Wang, X., 2022a. Heterogeneous data fusion and loss function design for tooth point cloud segmentation. *Neural Computing and Applications* 34, 17371–17380.
- Liu, J., Hu, T., Feng, Y., Ding, W., Liu, Z., 2023a. Tooth-segnet: image degradation meets tooth segmentation in cbct images, in: 2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI), IEEE. pp. 1–5.
- Liu, J., Zhang, H., Chen, J., Meng, R., Gao, C., Han, L., Song, Y., Tian, Y., Wang, Y., 2025a. Automated detection and segmentation of dental caries using a novel cascaded learning approach. *Biomedical Signal Processing and Control* 102, 107344.
- Liu, K., Elbatel, M., Chu, G., Shan, Z., Sum, F.H.K.M.H., Hung, K.F., Zhang, C., Li, X., Yang, Y., 2025b. Fd-tooth: Intraoral photographs and cbct images for fenestration and dehiscence detection. *Scientific Data* 12, 1007.
- Liu, S., Chen, T., Chen, X., Chen, X., Xiao, Q., Wu, B., Kärkkäinen, T., Pechenizkiy, M., Mocanu, D., Wang, Z., 2023b. More ConvNets in the 2020s: Scaling up Kernels Beyond 51x51 using Sparsity. *arXiv:2207.03620*.
- Liu, Y., Cheng, M.M., Hu, X., Wang, K., Bai, X., 2017. Richer convolutional features for edge detection, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3000–3009.
- Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Liu, Y., 2024a. VMamba: Visual State Space Model. *arXiv:2401.10166*.
- Liu, Y., Zhang, S., Wu, X., Yang, T., Pei, Y., Guo, H., Jiang, Y., Feng, Z., Xiao, W., Wang, Y.P., et al., 2024b. Individual graph representation learning for pediatric tooth segmentation from dental cbct. *IEEE Transactions on Medical Imaging*.
- Liu, Z., He, X., Wang, H., Xiong, H., Zhang, Y., Wang, G., Hao, J., Feng, Y., Zhu, F., Hu, H., 2022b. Hierarchical self-supervised learning for 3d tooth segmentation in intra-oral mesh scans. *IEEE Transactions on Medical Imaging* 42, 467–480.
- Liu, Z., Lin, Y., Cao, Y., Hu, H., Wei, Y., Zhang, Z., Lin, S., Guo, B., 2021. Swin Transformer: Hierarchical Vision Transformer using Shifted Windows, in: 2021 IEEE/CVF International Conference on Computer Vision (ICCV), IEEE, Montreal, QC, Canada. pp. 9992–10002. doi:10.1109/ICCV48922.2021.00986.
- Liu, Z., Tang, H., Xue, J., Liao, Y., 2025c. Two-stream structure-oriented neighbor enhancement network for dental model segmentation. *Engineering Applications of Artificial Intelligence* 150, 110618.
- Long, J., Shelhamer, E., Darrell, T., 2015. Fully convolutional networks for semantic segmentation, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3431–3440.
- Long, M., Yuan, T., 2024. Dftnet: Differential feature fusion model for tooth segmentation in cbct images, in: 2024 International Symposium on Digital Home (ISDH), IEEE. pp. 55–60.
- Loop, H.I.T., 2023. Teeth Segmentation on dental X-ray images. URL: <https://www.kaggle.com/datasets/humansintheloop/teeth-segmentation-on-dental-xray-images>, doi:10.34740/KAGGLE/DSV/5884500.
- Lu, J., Huang, X., Song, C., Li, C., Hu, Y., Xin, R., Emam, M., 2025. Cisa-unet: Dual auxiliary information for tooth segmentation from cbct images. *Alexandria Engineering Journal* 114, 543–555.
- Lu, X., Ma, C., Ni, B., Yang, X., Reid, I., Yang, M.H., 2018. Deep regression tracking with shrinkage loss, in: *Proceedings of the European conference on computer vision (ECCV)*, pp. 353–369.
- Lv, W., Niu, Q., Wang, L., Song, S., Huo, J., Wang, L., Xiao, R., 2024. Tooth segmentation from cone beam computed tomography using skeleton-guided v-net, in: *Proceedings of the 2024 4th International Conference on Bioinformatics and Intelligent Computing*, pp. 297–302.
- Lyu, Y., Liu, Z., Zhang, Y., Chen, H., Xu, Z., 2024. Crml-net: Cross-modal reasoning and multi-task learning network for tooth image segmentation. *Computer Vision and Image Understanding* 248, 104138.
- Ma, J., He, Y., Li, F., Han, L., You, C., Wang, B., 2024a. Segment anything in medical images. *Nature Communications* 15, 654.
- Ma, L., Xue, P., Gu, Y., Zhao, Y., Zhu, M., Ding, Z., Shen, D., 2023a. Accurate 3d prediction of missing teeth in diverse patterns for precise dental implant planning. *arXiv preprint arXiv:2307.07953*.
- Ma, T., Qian, J., Qin, X., Yu, D., 2024b. Lightweight dental lesion segmentation algorithm based on transformer and cnn, in: 2024 9th International Conference on Image, Vision and Computing (ICIVC), IEEE. pp. 183–187.
- Ma, T., Shi, Y., Zhou, X., Qin, X., 2023b. Semi-supervised tooth lesion segmentation algorithm based on generative adversarial network, in: 2023 IEEE International Conference on Signal Processing, Communications and Computing (ICSPCC), IEEE. pp. 1–5.

- Ma, Y., Qu, Z., Liu, X., Lin, J., Jin, Z., 2025. A high-precision framework for teeth instance segmentation in panoramic radiographs. *Displays* , 103064.
- Mai, H., Dang, X., Chen, J., Guo, J., Chen, X., 2024. A multi-axial network for oral structure segmentation, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 49–62.
- Majanga, V., Viriri, S., 2019. A deep learning approach for automatic segmentation of dental images, in: *International Conference on Mining Intelligence and Knowledge Exploration*, Springer. pp. 143–152.
- Mangano, F., Gandolfi, A., Luongo, G., Logozzo, S., 2017. Intraoral scanners in dentistry: a review of the current literature. *BMC oral health* 17, 149.
- Manzari, O.N., Bayrami, F., Khaloo, H., Khodakaramimagsoud, Z., Shokouhi, S.B., 2024. Denunet: enhancing dental image segmentation through edge and body fusion. *Multimedia Tools and Applications* , 1–21.
- Mao, J., Liu, Z., Gu, J., Wang, C., Chan, S., 2024. Cursegnet: 3d dental model segmentation network based on curve feature aggregation, in: *International Conference on Artificial Neural Networks*, Springer. pp. 245–259.
- Mao, K., Thu, K.M., Hung, K.F., Yu, O.Y., Hsung, R.T.C., Lam, W.Y.H., 2025. Artificial intelligence in detecting periodontal disease from intraoral photographs: A systematic review. *International Dental Journal* 75, 100883.
- Mei, L., Deng, K., Cui, Z., Fang, Y., Li, Y., Lai, H., Tonetti, M.S., Shen, D., 2025. Clinical knowledge-guided hybrid classification network for automatic periodontal disease diagnosis in x-ray image. *Medical Image Analysis* 99, 103376.
- Meine, H., Metzger, M.C., Weingart, P., Wüster, J., Schmelzeisen, R., Rörich, A., Georgii, J., Brandenburg, L.S., 2025. Determination of kennedy’s classification in panoramic x-rays by automated tooth labeling. *International Journal of Computer Assisted Radiology and Surgery* , 1–9.
- Mildenberger, P., Eichelberg, M., Martin, E., 2002. Introduction to the dicom standard. *European radiology* 12, 920–927.
- Milletari, F., Navab, N., Ahmadi, S.A., 2016. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation, in: *2016 Fourth International Conference on 3D Vision (3DV)*, IEEE, Stanford, CA, USA. pp. 565–571. doi:10.1109/3DV.2016.79.
- Mirikharaji, Z., Abhishek, K., Bissoto, A., Barata, C., Avila, S., Valle, E., Celebi, M.E., Hamarneh, G., 2023. A survey on deep learning for skin lesion segmentation. *Medical Image Analysis* 88, 102863.
- Mohan, R., Arunmozhi, R., Rajinikanth, V., 2023. Deep-learning segmentation and recognition of tooth in thresholded panoramic x-ray, in: *2023 Winter Summit on Smart Computing and Networks (WiSSCoN)*, IEEE. pp. 1–5.
- Mozzo, P., Procacci, C., Tacconi, A., Tinazzi Martini, P., Bergamo Andreis, I., 1998. A new volumetric ct machine for dental imaging based on the cone-beam technique: preliminary results. *European radiology* 8, 1558–1564.
- Nambiar, R., Nanjundegowda, R., 2024. A comprehensive review of ai and deep learning applications in dentistry: From image segmentation to treatment planning. *Journal of Robotics and Control (JRC)* 5, 1744–1752.
- Nasra, P., Gupta, S., Jagadish, S., Singh, A.R., 2025. Deep learning for oral disease detection: Leveraging inceptionresnetv2 for automated diagnosis, in: *2025 4th OPJU International Technology Conference (OTCON) on Smart Computing for Innovation and Advancement in Industry 5.0*, IEEE. pp. 1–6.
- Nath, V., Li, W., Yang, D., Myronenko, A., Zheng, M., Lu, Y., Liu, Z., Yin, H., Tang, Y., Guo, P., Zhao, C., Xu, Z., He, Y., Heinrich, G., Law, Y.M., Simon, B., Harmon, S., Aylward, S., Edgar, M., Zephyr, M., Han, S., Molchanov, P., Turkbey, B., Roth, H., Xu, D., 2025. VILA-M3: Enhancing Vision-Language Models with Medical Expert Knowledge. doi:10.48550/arXiv.2411.12915, arXiv:2411.12915.
- Neeraja, R., Anbarasi, L.J., 2025. Dentic\_bsfl net: Boundary refined and spatial adaptive feature learning network for enhanced dental caries segmentation. *IEEE Access* .
- Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al., 2018. Attention u-net: Learning where to look for the pancreas. *arXiv preprint arXiv:1804.03999* .
- Organization, W.H., 2022. Global oral health status report: towards universal health coverage for oral health by 2030. World Health Organization.
- Organization, W.H., 2024. Global strategy and action plan on oral health 2023–2030. World Health Organization.
- Organization, W.H., et al., 2025. Workforce factsheet: achieving universal coverage of oral health services by 2030, in: *Workforce factsheet: achieving universal coverage of oral health services by 2030*.
- Orville, 2024. Dental opg kennedy dataset. URL: <https://www.kaggle.com/datasets/orville/dental-opg-kennedy-dataset>. accessed: 2025-08-24; This dataset includes panoramic dental X-ray images annotated with Kennedy classification.

- Osama, R., Selim, S., Elattar, M., 2024. Cue-aided multi-class segmentation (cams): A novel approach for enhanced modelling in dental panoramic imaging, in: 2024 International Conference on Innovation and Intelligence for Informatics, Computing, and Technologies (3ICT), IEEE. pp. 37–42.
- Oztekin, F., Katar, O., Sadak, F., Aydogan, M., Yildirim, T.T., Plawiak, P., Yildirim, O., Talo, M., Karabatak, M., 2022. Automatic semantic segmentation for dental restorations in panoramic radiography images using u-net model. *International Journal of Imaging Systems and Technology* 32, 1990–2001.
- Pan, P., Zhang, C., Sun, J., Guo, L., 2025a. Multi-scale conv-attention u-net for medical image segmentation. *Scientific Reports* 15, 12041.
- Pan, X., Chen, Y., Chen, X., Chen, L., Dai, C., Chen, Q., Jia, Y., 2025b. Mdgformer: Multi-domain omni sensing and co-guided for dental segmentation. *Expert Systems with Applications* 289, 128182.
- Panetta, K., Rajendran, R., Ramesh, A., Rao, S.P., Agaian, S., 2021. Tufts dental database: a multimodal panoramic x-ray dataset for benchmarking diagnostic systems. *IEEE journal of biomedical and health informatics* 26, 1650–1659.
- Pang, Y.N., Yang, Z., Zhang, L.X., Liu, X.q., Dong, X.S., Sheng, X., Tan, J.g., Mao, X.Y., Liu, M.y., 2025. Establishment and evaluation of a deep learning-based tooth wear severity grading system using intraoral photographs. *Journal of Dental Sciences* 20, 477–486.
- Pivoshenko, V., 2023. Panoramic dental x-rays with segmented mandibles. URL: <https://www.kaggle.com/datasets/volodymyrpivoshenko/panoramic-dental-x-rays-with-segmented-mandibles>. accessed: 2025-08-23; Published: 2023-10-10 (according to SelectDataset).
- Qin, F., Li, X., Liu, M., Huang, Z., Han, J., Zhuo, L., 2024. An inter-layer volumetric attention network for teeth segmentation in cross-fov cbct dental images, in: 2024 IEEE International Conference on Progress in Informatics and Computing (PIC), IEEE. pp. 254–258.
- Qiu, L., Ye, C., Chen, P., Liu, Y., Han, X., Cui, S., 2022. Darch: Dental arch prior-assisted 3d tooth instance segmentation with weak annotations, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 20752–20761.
- Rad, A.E., Rahim, M.S.M., Rehman, A., Saba, T., 2016. Digital dental x-ray database for caries screening. *3D Research* 7, 18.
- Radford, A., Kim, J.W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., Sastry, G., Askell, A., Mishkin, P., Clark, J., Krueger, G., Sutskever, I., . Transferable visual models from natural language supervision .
- Rahman, M.M., Marculescu, R., 2025. Effiddec3d: An optimized decoder for high-performance and efficient 3d medical image segmentation, in: Proceedings of the Computer Vision and Pattern Recognition Conference, pp. 10435–10444.
- Rahman, R.B., Tanim, S.A., Alfaz, N., Shrestha, T.E., Miah, M.S.U., Mridha, M., 2024. A comprehensive dental dataset of six classes for deep learning based object detection study. *Data in Brief* 57, 110970.
- Rao, Y., Nartey, O.T., Zeng, S., Acheampong, K.N., Haruna, C.R., Sun, J., 2022. Lefunet: unet with learnable feature connections for teeth identification and segmentation in dental panoramic x-ray images, in: 2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 2110–2118.
- Rao, Y., Wang, Y., Meng, F., Pu, J., Sun, J., Wang, Q., 2020. A symmetric fully convolutional residual network with dcrf for accurate tooth segmentation. *IEEE Access* 8, 92028–92038.
- Rekik, A., Ben-Hamadou, A., Smaoui, O., Bouzguenda, F., Pujades, S., Boyer, E., 2025. Tseglab: Multi-stage 3d dental scan segmentation and labeling. *Computers in Biology and Medicine* 185, 109535.
- Ren, S., He, K., Girshick, R., Sun, J., 2015. Faster r-cnn: Towards real-time object detection with region proposal networks. *Advances in neural information processing systems* 28.
- ricoleehduu, 2023. Semi-supervised teeh segmentation challenges (sts-challenge). <https://github.com/ricoleehduu/STS-Challenge>. GitHub repository; accessed 23 August 2025; MIT License.
- Robbins, H., Monro, S., 1951. A stochastic approximation method. *The Annals of Mathematical Statistics* 22, 400–407.
- Ronneberger, O., Fischer, P., Brox, T., 2015. U-net: Convolutional networks for biomedical image segmentation, in: International Conference on Medical image computing and computer-assisted intervention, Springer. pp. 234–241.
- Rousseau, J., Alaka, C., Covili, E., Mayard, H., Misrachi, L., Au, W., 2023. Pre-training with diffusion models for dental radiography segmentation, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 174–182.
- Różyło-Kalinowska, I., 2021. Panoramic radiography in dentistry. *Clinical dentistry reviewed* 5, 26.

- Rundo, F., Pino, C., Sarpietro, R.E., Spampinato, C., Salanitri, F.P., 2022. A hierarchical 3d segmentation model for cone-beam computed tomography dental-arch scans, in: International Conference on Pattern Recognition, Springer. pp. 589–601.
- Şahin, M.E., Ulutaş, H., Süzgen, E.E., 2024. Automated segmentation of dental structures in panoramic radiographs using u-net 3+, in: 2024 8th International Artificial Intelligence and Data Processing Symposium (IDAP), IEEE. pp. 1–6.
- Salehi, S.S.M., Erdogmus, D., Gholipour, A., 2017. Tversky loss function for image segmentation using 3d fully convolutional deep networks, in: International workshop on machine learning in medical imaging, Springer. pp. 379–387.
- Salih, O., Duffy, K.J., 2022. The local ternary pattern encoder–decoder neural network for dental image segmentation. IET Image Processing 16, 1520–1530.
- Sarvesan, B., Sridhar, J., Soumya, S., Tejaswini, V., Yamini, G., Avula, S., 2024. Enhancing dental x-ray segmentation using deep learning, in: 2024 International Conference on Smart Technologies for Sustainable Development Goals (ICSTSDG), IEEE. pp. 1–7.
- Sehar, U., Xiong, J., Xia, Z., 2025. Automatic tooth labeling after segmentation using prototype-based meta-learning. Machine Intelligence Research , 1–14.
- Seifossadat, A., 2023. Dental radiography segmentation dataset. URL: <https://www.kaggle.com/datasets/abbasseifossadat/dental-radiography-segmentation>. accessed: 2025-08-23; This dataset includes dental radiography images with segmentation masks for various conditions such as cavities and implants.
- Shamshad, F., Khan, S., Zamir, S.W., Khan, M.H., Hayat, M., Khan, F.S., Fu, H., 2023. Transformers in medical imaging: A survey. Medical image analysis 88, 102802.
- Shao, Y., Wang, M., Mai, J., Fu, X., Li, M., Zheng, J., Diao, Z., Yin, A., Chen, Y., Xiao, J., et al., 2022. Lidp: A lung image dataset with pathological information for lung cancer screening, in: International Conference on Medical Image Computing and Computer-Assisted Intervention, Springer. pp. 770–779.
- Shen, C., Lian, C., Zhang, W., Wang, F., Zhang, J., Fan, S., Wei, X., Wang, G., Li, K., Mu, H., et al., 2025. Large-vocabulary forensic pathological analyses via prototypical cross-modal contrastive learning. Nature Communications 16, 6773.
- Sheng, C., Wang, L., Huang, Z., Wang, T., Guo, Y., Hou, W., Xu, L., Wang, J., Yan, X., 2023. Transformer-based deep learning network for tooth segmentation on panoramic radiographs. Journal of systems science and complexity 36, 257–272.
- Shi, J., Sun, B., Ye, X., Wang, Z., Luo, X., Liu, J., Gao, H., Li, H., 2022. Semantic decomposition network with contrastive and structural constraints for dental plaque segmentation. IEEE Transactions on Medical Imaging 42, 935–946.
- Shi, R., Liu, C., Tao, J., Li, G., Xiao, K., Xie, Z., Wang, J., Gharabaghi, B., 2021. Panoramic radiographic x-ray image tooth root segmentation based on lenet-5 networks, in: International Conference on Advanced Machine Learning Technologies and Applications, Springer. pp. 134–144.
- Shibuya, K., Murata, K., Nakayama, R., Hizukuri, A., 2022. Semantic segmentation method for tooth type with data augmentation using alpha-blending, in: 2022 Joint 12th International Conference on Soft Computing and Intelligent Systems and 23rd International Symposium on Advanced Intelligent Systems (SCIS&ISIS), IEEE. pp. 1–2.
- Siddhartha, K.S., 2024. Dental anatomy dataset - yolov8. URL: <https://www.kaggle.com/datasets/saisiddhartha69/dental-anatomy-dataset-yolov8>. accessed: 2025-08-23; Annotated using Roboflow.
- Silva, B., Fontinele, J., Vieira, C.L.Z., Tavares, J.M.R., Cury, P.R., Oliveira, L., 2025. A holistic approach for classifying dental conditions from textual reports and panoramic radiographs. Medical Image Analysis , 103709.
- Silva, B., Pinheiro, L., Oliveira, L., Pithon, M., 2020. A study on tooth segmentation and numbering using end-to-end deep neural networks, in: 2020 33rd SIBGRAPI conference on graphics, patterns and images (SIBGRAPI), IEEE. pp. 164–171.
- Silvres, L., 2024. Dental disease panoramic detection dataset. URL: <https://www.kaggle.com/datasets/lokisilvres/dental-disease-panoramic-detection-dataset>. accessed: 2025-08-23; This dataset includes panoramic dental X-ray images with annotations in YOLO and COCO formats.
- Simonyan, K., Zisserman, A., 2014. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 .
- Sivagami, S., Chitra, P., Kailash, G.S.R., Muralidharan, S., 2020. Unet architecture based dental panoramic image segmentation, in: 2020 International Conference on Wireless Communications Signal Processing and Networking (WiSPNET), IEEE. pp. 187–191.

- Sivasundaram, S., Pandian, C., 2021. Performance analysis of classification and segmentation of cysts in panoramic dental images using convolutional neural network architecture. *International Journal of Imaging Systems and Technology* 31, 2214–2225.
- Song, A.H., Williams, M., Williamson, D.F., Chow, S.S., Jaume, G., Gao, G., Zhang, A., Chen, B., Baras, A.S., Serafin, R., et al., 2024a. Analysis of 3d pathology samples using weakly supervised ai. *Cell* 187, 2502–2520.
- Song, W., Wang, X., Guo, Y., Li, S., Xia, B., Hao, A., 2024b. Centerformer: a novel cluster center enhanced transformer for unconstrained dental plaque segmentation. *IEEE Transactions on Multimedia*.
- Stera, G., Giusti, M., Magnini, A., Calistri, L., Izzetti, R., Nardi, C., 2024. Diagnostic accuracy of periapical radiography and panoramic radiography in the detection of apical periodontitis: a systematic review and meta-analysis. *La radiologia medica* 129, 1682–1695.
- Sun, D., Pei, Y., Li, P., Song, G., Guo, Y., Zha, H., Xu, T., 2020a. Automatic tooth segmentation and dense correspondence of 3d dental model, in: *International conference on medical image computing and computer-assisted intervention*, Springer. pp. 703–712.
- Sun, D., Pei, Y., Song, G., Guo, Y., Ma, G., Xu, T., Zha, H., 2020b. Tooth segmentation and labeling from digital dental casts, in: *2020 IEEE 17th International Symposium on Biomedical Imaging (ISBI)*, IEEE. pp. 669–673.
- Sun, F., Luo, Z., Li, S., 2023. Boundary difference over union loss for medical image segmentation, in: *International conference on medical image computing and computer-assisted intervention*, Springer. pp. 292–301.
- Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A., 2015. Going deeper with convolutions, in: *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1–9.
- Taghanaki, S.A., Zheng, Y., Zhou, S.K., Georgescu, B., Sharma, P., Xu, D., Comaniciu, D., Hamarneh, G., 2019. Combo loss: Handling input and output imbalance in multi-organ segmentation. *Computerized Medical Imaging and Graphics* 75, 24–33.
- Tan, M., Cui, Z., Zhong, T., Fang, Y., Zhang, Y., Shen, D., 2024. A progressive framework for tooth and substructure segmentation from cone-beam ct images. *Computers in Biology and Medicine* 169, 107839.
- Tan, M., Fang, Y., Ma, L., Zhang, Y., Cui, Z., Shen, D., 2023a. Dental anatomy segmentation from cone beam ct images, in: *2023 IEEE 20th International Symposium on Biomedical Imaging (ISBI)*, IEEE. pp. 1–4.
- Tan, Y., Xiang, X., 2023. Boundary-constrained graph network for tooth segmentation on 3d dental surfaces, in: *International Workshop on Machine Learning in Medical Imaging*, Springer. pp. 94–103.
- Tan, Y., Xiang, X., Chen, Y., Jing, H., Ye, S., Xue, C., Xu, H., 2023b. Coupling bracket segmentation and tooth surface reconstruction on 3d dental models, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 411–420.
- Tang, L., Zhan, Y., Chen, Z., Yu, B., Tao, D., 2022. Contrastive boundary learning for point cloud segmentation, in: *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 8489–8499.
- Team, I.K., 2023. Dental radiography dataset. URL: <https://www.kaggle.com/datasets/imtkaggleteam/dental-radiography>. accessed: 2025-08-23.
- Tez, B.Ç., Güzel, Y., Kızıltan Eliaçık, B.B., Aydın, Z., 2025. Deep-learning-based ai-model for predicting dental plaque in the young permanent teeth of children aged 8–13 years. *Children* 12, 475.
- Tian, S., Dai, N., Zhang, B., Yuan, F., Yu, Q., Cheng, X., 2019. Automatic classification and segmentation of teeth on 3d dental model using hierarchical deep learning networks. *Ieee Access* 7, 84817–84828.
- Tian, Y., Zhang, Y., Chen, W.G., Liu, D., Wang, H., Xu, H., Han, J., Ge, Y., 2022. 3d tooth instance segmentation learning objectness and affinity in point cloud. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM)* 18, 1–16.
- Tibúrcio-Machado, C., Michelon, C., Zanatta, F., Gomes, M.S., Marin, J.A., Bier, C.A., 2021. The global prevalence of apical periodontitis: a systematic review and meta-analysis. *International endodontic journal* 54, 712–735.
- Tran, N.D., Tran, T.T., Nguyen, Q.H., Vu, M.H., Pham, V.T., 2025. Litenext: A novel lightweight convmixer-based model with self-embedding representation parallel for medical image segmentation. *Biomedical Signal Processing and Control* 107, 107773.
- UBU, E., 2024. Root disease x-ray imaging dataset. URL: <https://www.kaggle.com/datasets/engineeringubu/root-disease-x-ray-imaging>. accessed: 2025-08-23; This dataset includes X-ray images for analyzing root diseases.
- Usman, M., Rehman, A., Saleem, A.M., Jawaid, R., Byon, S.S., Kim, S.H., Lee, B.D., Heo, M.S., Shin, Y.G., 2022. Dual-stage deeply supervised attention-based convolutional neural networks for mandibular canal segmentation in cbct scans. *Sensors* 22, 9877.

- Van Nistelrooij, N., de Medeiros Maia, H.C., Cao, L., Vinayahalingam, S., Loomans, B., Cenci, M.S., Mendes, F.M., 2025. Automated detection and numbering of primary and permanent teeth in digital impressions of children using artificial intelligence. *Journal of Dentistry*, 105976.
- Van Pham, T., Lucas, Y., Treuillet, S., Debraux, L., 2020. Object contour refinement using instance segmentation in dental images, in: *International Conference on Advanced Concepts for Intelligent Vision Systems*, Springer. pp. 99–107.
- Vandenberghe, B., Jacobs, R., Bosmans, H., 2010. Modern dental imaging: A review of the current technology and clinical applications in dental practice. *European Radiology* 20, 2637–2655. doi:10.1007/s00330-010-1836-1.
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I., 2017. Attention is all you need. *Advances in neural information processing systems* 30.
- Venkatesh, E., Elluru, S.V., 2017. Cone beam computed tomography: Basics and applications in dentistry. *Journal of Istanbul University Faculty of Dentistry* 51, S102–S121. doi:10.17096/jiufd.00289.
- Wang, C., Yang, J., Wu, B., Liu, R., Yu, P., 2024a. Transvnet: Transformer-based tooth semantic segmentation in cbct images. *Biomedical Signal Processing and Control* 97, 106666.
- Wang, C., Zhang, Y., Wu, C., Liu, J., Wu, L., Wang, Y., Huang, X., Feng, X., Wang, Y., 2025a. Mmdental-a multimodal dataset of tooth cbct images with expert medical records. *Scientific Data* 12, 1172.
- Wang, H., Li, K., Zhu, J., Wang, F., Lian, C., Ma, J., 2024b. Weakly supervised tooth instance segmentation on 3d dental models with multi-label learning, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 723–733.
- Wang, H., Li, Y., Liu, L., 2024c. Multi-path restorative pre-training with localized adaptive graph convolution for 2d tooth segmentation, in: *2024 4th International Conference on Industrial Automation, Robotics and Control Engineering (IARCE)*, IEEE. pp. 250–254.
- Wang, L., Bao, W., 2024. Semi-supervised 2d dental image segmentation via cross teaching network, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 178–187.
- Wang, P., Gu, H., Sun, Y., 2025b. Tooth segmentation on multimodal images using adapted segment anything model. *Scientific Reports* 15, 13874.
- Wang, P., He, A., Wang, A., Zhou, Z., Guan, X., Li, T., 2025c. Towards automated pediatric dental development staging: A dataset and model, in: *International Conference on Medical Image Computing and Computer-Assisted Intervention*, Springer. pp. 605–615.
- Wang, R., Cheng, F., Dai, G., Zhang, J., Fan, C., Yu, J., Li, J., Jiang, F., 2025d. Pxseg: automatic tooth segmentation, numbering and abnormal morphology detection based on cbct and panoramic radiographs. *BMC Oral Health* 25, 1230.
- Wang, S., Lei, C., Liang, Y., Sun, J., Xie, X., Wang, Y., Zuo, F., Bai, Y., Li, S., Liu, Y.J., 2024d. A 3d dental model dataset with pre/post-orthodontic treatment for automatic tooth alignment. *Scientific Data* 11, 1277.
- Wang, X., Bo, L., Fuxin, L., 2019. Adaptive wing loss for robust face alignment via heatmap regression, in: *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 6971–6981.
- Wang, X., Gao, S., Jiang, K., Zhang, H., Wang, L., Chen, F., Yu, J., Yang, F., 2023a. Multi-level uncertainty aware learning for semi-supervised dental panoramic caries segmentation. *Neurocomputing* 540, 126208.
- Wang, X., Gao, S., Jiang, K., Zhang, H., Wang, L., Chen, F., Yu, J., Yang, F., 2023b. Multi-level uncertainty aware learning for semi-supervised dental panoramic caries segmentation. *Neurocomputing* 540, 126208.
- Wang, X., He, Z., Liu, C., Zhang, B., Lin, Z., Guo, J., Xie, S., 2023c. Cga-unet: Category-guide attention unet for dental abnormality detection and segmentation from dental-maxillofacial images. *IEEE Transactions on Instrumentation and Measurement* 72, 1–11.
- Wang, X., Wang, L., Yang, Z., Zhou, J., Zheng, Y., Chen, F., Hong, R., Yu, J., Yang, F., 2024e. Dsis-dpr: Structured instance segmentation and diffusion prior refinement for dental anatomy learning. *IEEE Transactions on Multimedia* 26, 9464–9476.
- Wang, X., Wang, Z., 2025. Lsaformer: A 3d tooth segmentation method integrating local surface reconstruction and transformer, in: *2025 10th International Conference on Intelligent Computing and Signal Processing (ICSP)*, IEEE. pp. 433–436.
- Wang, Y., Chen, X., Zhang, H., Wu, C., Ye, F., Zhang, Q., Zhang, Y., Li, Y., 2024f. Sts 2024: 2nd semi-supervised teeth segmentation miccai challenge. <https://sts-challenge.github.io/miccai2024/organizers.html>. Accessed: 2025-08-24; Organized by Communication University of Zhejiang, Hangzhou Dianzi University, Shenzhen University, Queen Mary University of London, and Hangzhou Dental Hospital Group.

- Wang, Y., Chen, X., Zhang, H., Wu, C., Ye, F., Zhang, Q., Zhang, Y., Li, Y., 2024g. STS 2024: 2nd semi-supervised teeth segmentation MICCAI challenge. <https://sts-challenge.github.io/miccai2024/index.html>. Accessed: 2025-08-24.
- Wang, Y., Chen, X., Zhang, H., Wu, C., Ye, F., Zhang, Q., Zhang, Y., Li, Y., 2025e. Sts 2025: 3rd semi-supervised teeth segmentation and registration miccai challenge. <https://songhen15.github.io/STSdevelop.github.io/miccai2025/organizers.html>. Accessed: 2025-08-24; Organized by Communication University of Zhejiang, Hangzhou Dianzi University, Shenzhen University, Queen Mary University of London, and Hangzhou Dental Hospital Group.
- Wang, Y., Xia, W., Yan, Z., Zhao, L., Bian, X., Liu, C., Qi, Z., Zhang, S., Tang, Z., 2023d. Root canal treatment planning by automatic tooth and root canal segmentation in dental CBCT with deep multi-task feature learning. *Medical Image Analysis* 85, 102750. doi:10.1016/j.media.2023.102750.
- Wang, Y., Xia, W., Yan, Z., Zhao, L., Bian, X., Liu, C., Qi, Z., Zhang, S., Tang, Z., 2023e. Root canal treatment planning by automatic tooth and root canal segmentation in dental cbct with deep multi-task feature learning. *Medical image analysis* 85, 102750.
- Wang, Y., Ye, F., Chen, Y., Wang, C., Wu, C., Xu, F., Ma, Z., Liu, Y., Zhang, Y., Cao, M., et al., 2025f. A multimodal dental dataset for semi-supervised deep learning image segmentation. *Scientific Data* 12, 117.
- Wathore, S., Gorthi, S., 2025. Bilateral symmetry-based augmentation method for improved tooth segmentation in panoramic x-rays. *Pattern Recognition Letters* 188, 1–7.
- Wen, X., Liu, Z., Chu, Y., Le, M., Li, L., 2024. Mrcm-uctransnet: Automatic and accurate 3d tooth segmentation network from cone-beam ct images. *International Journal of Imaging Systems and Technology* 34, e23139.
- Weng, Z., Zheng, S., Yu, C., Zhu, D., Yu, H., Jia, L., Zhang, D., Lin, H., 2024. X-match: A semi-supervised framework for oral jawbones segmentation using wavelet transform for enhanced consistency learning, in: 2024 IEEE International Conference on Bioinformatics and Biomedicine (BIBM), IEEE. pp. 7156–7158.
- Widiasri, M., Suciati, N., Fatichah, C., Astuti, E.R., Putra, R.H., Arifin, A.Z., 2023. Alveolar bone and mandibular canal segmentation on cone beam computed tomography images using u-net, in: 2023 8th International Conference on Instrumentation, Control, and Automation (ICA), IEEE. pp. 36–41.
- Wu, D., Jiang, J., Wang, J., Bi, Z., Yu, G., 2025a. Segmentation algorithms of dental ct images: A comprehensive review from classical to deep learning trend. *Expert Systems with Applications* , 126853.
- Wu, L., Nie, Y., He, S., Zhuang, J., Chen, H., 2025b. UniBiomed: A Universal Foundation Model for Grounded Biomedical Image Interpretation. doi:10.48550/arXiv.2504.21336, arXiv:2504.21336.
- Wu, R., Zhang, C., Zhang, J., Zhou, Y., Zhou, T., Fu, H., 2024. MM-Retinal: Knowledge-Enhanced Foundational Pretraining with Fundus Image-Text Expertise. doi:10.48550/arXiv.2405.11793, arXiv:2405.11793.
- Wu, T.H., Lian, C., Lee, S., Pastewait, M., Piers, C., Liu, J., Wang, F., Wang, L., Chiu, C.Y., Wang, W., et al., 2022. Two-stage mesh deep learning for automated tooth segmentation and landmark localization on 3d intraoral scans. *IEEE transactions on medical imaging* 41, 3158–3166.
- Wu, Y., Wang, Z., Yang, X., Kang, H., He, A., Li, T., 2025c. Trans-sam: Transfer segment anything model to medical image segmentation with parameter-efficient fine-tuning. *Knowledge-Based Systems* 310, 112909.
- Xi, S., Liu, Z., Chang, J., Wu, H., Wang, X., Hao, A., 2025. 3d dental model segmentation with geometrical boundary preserving, in: *Proceedings of the Computer Vision and Pattern Recognition Conference*, pp. 10476–10485.
- Xie, L., Liu, B., Cao, Y., Yang, C., 2022. Automatic individual tooth segmentation in cone-beam computed tomography based on multi-task cnn and watershed transform, in: 2022 IEEE 24th Int Conf on High Performance Computing & Communications; 8th Int Conf on Data Science & Systems; 20th Int Conf on Smart City; 8th Int Conf on Dependability in Sensor, Cloud & Big Data Systems & Application (HPCC/DSS/SmartCity/DependSys), IEEE. pp. 749–754.
- Xing, Y., Liao, P., Alasleh, R.A., Khampatee, V., Alizadeh-Shabdiz, F., 2024. Dental x-ray segmentation and auto implant design based on convolutional neural network, in: 2024 IEEE 7th International Conference on Multimedia Information Processing and Retrieval (MIPR), IEEE. pp. 243–246.
- Xiong, H., Li, K., Tan, K., Feng, Y., Zhou, J.T., Hao, J., Ying, H., Wu, J., Liu, Z., 2023. Tsegformer: 3d tooth segmentation in intraoral scans with geometry guided transformer, in: *International conference on medical image computing and computer-assisted intervention*, Springer. pp. 421–432.
- Xu, P., Gholamalizadeh, T., Moshfeghifar, F., Darkner, S., Erleben, K., 2023. Deep-learning-based segmentation of individual tooth and bone with periodontal ligament interface details for simulation purposes. *IEEE Access* 11, 102460–102470.

- Xu, X., Chen, J., Yin, J., 2025a. Tooth instance segmentation and disease detection with uncertainty-aware contrastive learning and cross-scale attention. *IEEE Journal of Biomedical and Health Informatics* .
- Xu, X., Liu, C., Zheng, Y., 2018. 3d tooth segmentation and labeling using deep convolutional neural networks. *IEEE transactions on visualization and computer graphics* 25, 2336–2348.
- Xu, X., Liu, H., Zhang, T., Xiong, H., Yu, W., 2025b. PreCM: The Padding-based Rotation Equivariant Convolution Mode for Semantic Segmentation. *IEEE Transactions on Image Processing* , 1–doi:10.1109/TIP.2025.3558425, [arXiv:2411.01624](#).
- Yang, Y., Xie, R., Jia, W., Chen, Z., Yang, Y., Xie, L., Jiang, B., 2021. Accurate and automatic tooth image segmentation model with deep convolutional neural networks and level set method. *Neurocomputing* 419, 108–125.
- Yaswanth, J.S., Kumar, K.J., Singh, R.P., 2025. Automated tooth segmentation in x-ray images using attention integrated u-net++ model, in: 2025 3rd International Conference on Intelligent Systems, Advanced Computing and Communication (ISACC), IEEE. pp. 1317–1322.
- Yu, W., Wang, X., Yang, H., 2025. Clinically oriented automatic three-dimensional enamel segmentation via deep learning. *BMC Oral Health* 25, 133.
- Yurdakurban, E., Bayırlı, A.B., Uytun, M., Suçeken, Ö., Karasoy, O., Özkara, O., Topsakal, K.G., 2025. Automated periodontal assessment in orthodontic patients: a dual cnn framework. *Clinical Oral Investigations* 29, 328.
- Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G., 2006. User-guided 3D active contour segmentation of anatomical structures: Significantly improved efficiency and reliability. *Neuroimage* 31, 1116–1128.
- Zanjani, F.G., Moin, D.A., Verheij, B., Claessen, F., Cherici, T., Tan, T., et al., 2019. Deep learning approach to semantic segmentation in 3d point cloud intra-oral scans of teeth, in: *International Conference on Medical Imaging with Deep Learning*, PMLR. pp. 557–571.
- Zannah, R., Bashar, M., Mushfiq, R.B., Chakrabarty, A., Hossain, S., Jung, Y.J., 2024. Semantic segmentation on panoramic dental x-ray images using u-net architectures. *IEEE Access* 12, 44598–44612.
- Zhang, B., Zhu, X., 2023. High-precision semi-supervised 3d dental segmentation based on nnunet, in: *MICCAI Challenge on Semi-supervised Tooth Segmentation*. Springer, pp. 180–191.
- Zhang, D., Li, Z., Zeng, F., Wei, Y., 2023a. Boundary feature fusion network for tooth image segmentation, in: *MICCAI Challenge on Semi-supervised Tooth Segmentation*. Springer, pp. 110–123.
- Zhang, F., Zhu, J., Hao, P., Wu, F., Zheng, Y., 2022. Bdu-net: toward accurate segmentation of dental image using border guidance and feature map distortion. *International Journal of Imaging Systems and Technology* 32, 1221–1230.
- Zhang, J., Li, C., Song, Q., Gao, L., Lai, Y.K., 2020. Automatic 3d tooth segmentation using convolutional neural networks in harmonic parameter space. *Graphical Models* 109, 101071.
- Zhang, J., Li, X., Gao, Z., Chen, J., 2021a. Image detection of dental diseases based on deep transfer learning, in: *2021 2nd International Conference on Artificial Intelligence and Computer Engineering (ICAICE)*, IEEE. pp. 774–778.
- Zhang, J., Lin, M., Hou, H., Sun, B., Hu, F., Yu, Y., Li, M., 2025a. Easam: an edge-aware sam-based paradigm for tooth segmentation. *Signal, Image and Video Processing* 19, 673.
- Zhang, J., Xia, W., Dong, J., Tang, Z., Zhao, Q., 2021b. Root canal segmentation in cbct images by 3d u-net with global and local combination loss, in: *2021 43rd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*, IEEE. pp. 3097–3100.
- Zhang, K., Li, J., Ma, R., Li, G., 2021c. 3d temporomandibular joint cbct image segmentation via multi-directional resampling ensemble learning network, in: *International Workshop on Machine Learning in Medical Imaging*, Springer. pp. 634–643.
- Zhang, L., Zhao, Y., Meng, D., Cui, Z., Gao, C., Gao, X., Lian, C., Shen, D., 2021d. Tsgcnet: Discriminative geometric feature learning with two-stream graph convolutional network for 3d dental model segmentation, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 6699–6708.
- Zhang, X., Guo, E., Liu, X., Zhao, H., Yang, J., Li, W., Wu, W., Sun, W., 2025b. Enhancing furcation involvement classification on panoramic radiographs with vision transformers. *BMC Oral Health* 25, 153.
- Zhang, Y., Luo, Z., Li, S., 2023b. A multi-stage network with self-attention for tooth instance segmentation, in: *CCF Conference on Computer Supported Cooperative Work and Social Computing*, Springer. pp. 438–450.
- Zhang, Y., Qian, K., Zhu, Z., Yu, H., Zhang, B., 2023c. Dba-unet: a double u-shaped boundary attention network for maxillary sinus anatomical structure segmentation in cbct images. *Signal, Image and Video Processing* 17, 2251–2257.

- Zhang, Y., Ye, F., Chen, L., Xu, F., Chen, X., Wu, H., Cao, M., Li, Y., Wang, Y., Huang, X., 2023d. Children's dental panoramic radiographs dataset for caries segmentation and dental disease detection. *Scientific Data* 10, 380.
- Zhao, J., Zhou, Y., Chen, Z., Fu, H., Wan, L., 2024a. Topicwise separable sentence retrieval for medical report generation. *IEEE Transactions on Medical Imaging* .
- Zhao, J., Zhou, Y., Chen, Z., Fu, H., Wan, L., 2024b. Topicwise Separable Sentence Retrieval for Medical Report Generation. *IEEE Transactions on Medical Imaging* , 1–doi:10.1109/TMI.2024.3507076.
- Zhao, X., Xu, T., Peng, L., Li, S., Zhao, Y., Liu, H., He, J., Liang, S., 2023. Recognition and segmentation of teeth and mandibular nerve canals in panoramic dental x-rays by mask rcnn. *Displays* 78, 102447.
- Zhao, Y., Cheng, B., Niu, N., Wang, J., Zeng, T., Zhang, G., Shi, J., Li, J., 2024c. Few sampling meshes-based 3d tooth segmentation via region-aware graph convolutional network. *Expert Systems with Applications* 252, 124255.
- Zhao, Y., Li, P., Gao, C., Liu, Y., Chen, Q., Yang, F., Meng, D., 2020. Tsasnet: Tooth segmentation on dental panoramic x-ray images by two-stage attention segmentation network. *Knowledge-Based Systems* 206, 106338.
- Zhao, Y., Wu, R., Dai, P., Huang, H., Liu, Y., 2025a. Smtl-net: Domain prior-inspired tooth segmentation based on self-supervised manifold transfer learning. *IEEE Transactions on Neural Networks and Learning Systems* .
- Zhao, Y., Zhang, L., Liu, Y., Meng, D., Cui, Z., Gao, C., Gao, X., Lian, C., Shen, D., 2021a. Two-stream graph convolutional network for intra-oral scanner image segmentation. *IEEE Transactions on Medical Imaging* 41, 826–835.
- Zhao, Y., Zhang, L., Yang, C., Tan, Y., Liu, Y., Li, P., Huang, T., Gao, C., 2021b. 3d dental model segmentation with graph attentional convolution network. *Pattern Recognition Letters* 152, 79–85.
- Zhao, Y., Zhu, L., Wang, W., Lv, L., Li, Q., Liu, Y., Xi, J., Yi, C., 2025b. Progressive multi-task learning for fine-grained dental implant classification and segmentation in cbct image. *Computers in Biology and Medicine* 189, 109896.
- Zheng, Z., Yan, H., Setzer, F.C., Shi, K.J., Mupparapu, M., Li, J., 2020. Anatomically constrained deep learning for automating dental cbct segmentation and lesion detection. *IEEE Transactions on Automation Science and Engineering* 18, 603–614.
- Zhong, W., Ren, X., Zhang, H., 2025. Automatic x-ray teeth segmentation with grouped attention. *Scientific Reports* 15, 64.
- Zhou, M., Zheng, H., Zhang, M., Qin, X., Zhu, Y., Wu, Z., 2025a. Enhancing 3d intra-oral tooth segmentation through multi-scale graph convolution neural networks and cross-domain integration. *Journal of Medical and Biological Engineering* , 1–14.
- Zhou, Z., Chen, Y., He, A., Que, X., Wang, K., Yao, R., Li, T., 2024. Nkut: dataset and benchmark for pediatric mandibular wisdom teeth segmentation. *IEEE Journal of Biomedical and Health Informatics* 28, 3523–3533.
- Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N., Liang, J., 2019. Unet++: Redesigning skip connections to exploit multiscale features in image segmentation. *IEEE transactions on medical imaging* 39, 1856–1867.
- Zhou, Z., Zhang, Y., Xu, R., Zhao, X., Li, T., 2025b. Prad: Periapical radiograph analysis dataset and benchmark model development. *arXiv preprint arXiv:2504.07760* .
- Zhu, H., Cao, Z., Lian, L., Ye, G., Gao, H., Wu, J., 2023. Cariesnet: a deep learning approach for segmentation of multi-stage caries lesion from oral panoramic x-ray image. *Neural Computing and Applications* 35, 16051–16059.
- Zhuang, S., Wei, G., Cui, Z., Zhou, Y., 2025. Robust hybrid learning for automatic teeth segmentation and labeling on 3d dental models. *IEEE Trans. Multim.* 27, 792–803. URL: <https://doi.org/10.1109/TMM.2023.3289760>, doi:10.1109/TMM.2023.3289760.
- Zou, B., Wang, S., Liu, H., Sun, G., Wang, Y., Zuo, F., Quan, C., Zhao, Y., 2024. Teeth-seg: an efficient instance segmentation framework for orthodontic treatment based on multi-scale aggregation and anthropic prior knowledge, in: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 11601–11610.
- Zou, K., Tao, T., Yuan, X., Shen, X., Lai, W., Long, H., 2022. An interactive dual-branch network for hard palate segmentation of the oral cavity from cbct images. *Applied Soft Computing* 129, 109549.