CT-Less Attenuation Correction Using Multiview Ensemble Conditional Diffusion Model on High-Resolution Uncorrected PET Images

Alexandre St-Georges[1,2*], Gabriel Richard[1,2], Maxime Toussaint[1,2,3], Christian Thibaudeau[4], Etienne Auger[4], Étienne Croteau[5], Stephen Cunnane[5], Roger Lecomte[1,2,4], Jean-Baptiste Michaud[2,6*]

[1*]Department of Medical Imaging and Radiation Sciences, Université de Sherbrooke, Sherbrooke, QC, Canada.
[2]Sherbrooke Molecular Imaging Center, CRCHUS, Sherbrooke, QC, Canada.
[3]Nantes Université Laboratoire CRCI2NA, INSERM, CNRS, Nantes, France.
[4] Imaging Research and Technology (IR&T) Inc., Sherbrooke, Quebec, Canada.
[5]Research Center on Aging, Department of Medicine, Université de Sherbrooke, Sherbrooke, QC, Canada.
[6]Department of Electrical Engineering and Computer Engineering, Université de Sherbrooke, Sherbrooke, QC, Canada.

*Corresponding author(s). E-mail(s):
Alexandre.St-Georges@USherbrooke.ca;
Jean-Baptiste.Michaud@USherbrooke.ca;

Abstract

Accurate quantification in positron emission tomography (PET) is essential for accurate diagnostic results and effective treatment tracking. A major issue encountered in PET imaging is attenuation. Attenuation refers to the diminution of photon detected as they traverse biological tissues before reaching detectors. When such corrections are absent or inadequate, this signal degradation can introduce inaccurate quantification, making it difficult to differentiate benign from malignant conditions, and can potentially lead to misdiagnosis. Typically, this correction is done with co-computed Computed Tomography (CT) imaging to obtain structural data for calculating photon attenuation across the body. However, this methodology subjects patients to extra ionizing radiation exposure, suffers from potential spatial misregistration between PET/CT imaging sequences, and demands costly equipment infrastructure. Emerging advances in neural network architectures present an alternative approach via synthetic

CT image synthesis. Our investigation reveals that Conditional Denoising Diffusion Probabilistic Models (DDPMs) can generate high quality CT images from non attenuation corrected PET images in order to correct attenuation. By utilizing all three orthogonal views from non-attenuation-corrected PET images, the DDPM approach combined with ensemble voting generates higher quality pseudo-CT images with reduced artifacts and improved slice-to-slice consistency. Results from a study of 159 head scans acquired with the Siemens Biograph Vision PET/CT scanner demonstrate both qualitative and quantitative improvements in pseudo-CT generation. The method achieved a mean absolute error of 32 $\pm$ 10.4 HU on the CT images and an average error of (1.48 $\pm$ 0.68)% across all regions of interest when comparing PET images reconstructed using the attenuation map of the generated pseudo-CT versus the true CT.

Keywords: Deep learning, Diffusion model, attenuation correction, pseudo-CT, image generation, brain PET, CT-less attenuation correction

1 Background

In Positron Emission Tomography (PET), attenuation refers to the loss of photons resulting from photoelectric absorption and Compton scattering as photons interact with electrons in the body. This results in reduced signal intensity, especially in deeper regions and areas where tissues are denser. Such quantification inaccuracies can result in diagnostic errors or misinterpretation [1]. Attenuation correction is critical for accurate diagnostics and is systematically performed in the clinic using various methods. This process requires an attenuation map, which estimates how much the emitted photons are absorbed or scattered as they travel through the body. The attenuation map is essentially an image in which each pixel represents the attenuation coefficient of the tissue - a value that describes how much the tissue reduces the intensity of passing photons.

The most common way to generate this attenuation map is by using a secondary imaging modality that provides anatomical detail, such as Computed Tomography (CT) [2] or Magnetic Resonance Imaging (MRI)[3]. CT provides a direct measurement of tissue attenuation, but since the X-ray energy range differs from that of PET annihilation photons, an energy conversion, from Hounsfield units to attenuation coefficients, is necessary to generate an accurate attenuation map [4]. MRI also offers anatomical information, but it measures proton density rather than the electron density mostly responsible for photon interactions. As a result, generating an attenuation map from MRI requires a more complex conversion process to estimate attenuation coefficients. Several methods can be used for this, including atlas-based approaches [5], tissue segmentation[6], zero echo time imaging (ZTE)[7], and deep learning techniques[8]. Another approach to this problem is to use a rotating source to directly measure the transmission through tissue [9]. However, this method requires long acquisition times, which increases as the spatial resolution of the scanner improves due to its statistical nature. This makes the approach more susceptible to patient movement and increases radiation exposure.

Some scanners lack these attenuation correction methods and must therefore resort to new emerging technologies, such as deep learning [10], [11]. Considerable work has been done recently to generate pseudo-CT from Non-Attenuation-Corrected (NAC) PET images using various architectures, such as UNET [12], Generative Adversary

Network (GAN) [13] and more recently vision transformer [14]. Other methods utilize the NAC PET to directly correct the NAC image with promising recent architectures such as the Denoising Diffusion Probabilistic Model (DDPM) [15]. DDPMs are appealing as they remove the blurry aspect provided by UNET models [16], avoid mode collapse by GAN models [17] and generate higher quality samples than vision transformers, making them a great choice for pseudo-CT generation.

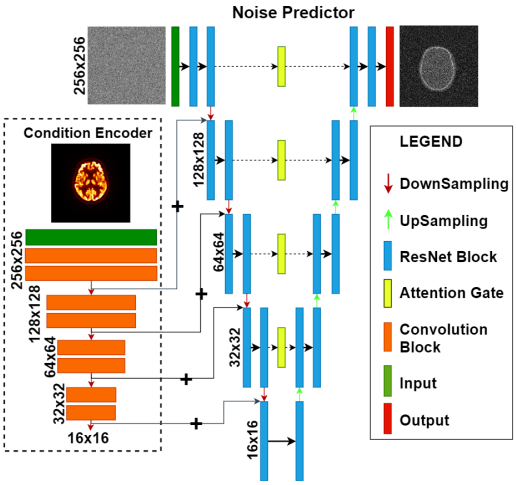
In this work, we describe an improved version of the DDPM using orthogonal planes and majority voting demonstrating that it is possible to precisely correct attenuation without the need for another imaging modality and, most importantly, without additional radiation exposure and misregistration risk with an adjunct anatomical imaging modality.

2 Methods

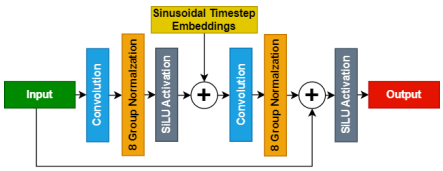
Ensuring the reliability of outputs from generative neural networks is a significant challenge because these systems can introduce artifacts and produce inconsistent results. To mitigate these concerns, our approach focuses on creating pseudo-CT (pCT) images from NAC PET scans, instead of directly producing attenuation-corrected PET images from a NAC PET image. This methodology ensures that only quantitative aspects of the PET data are modified while the anatomical structures remain untouched. This also improves the interpretability for doctors about how exactly the PET image is going to be modified for attenuation by the pCT. Furthermore, to limit artifact generation, our proposed model generates the pCT images of the same from different orthogonal planes of the NAC PET image in order to detect and eliminate artifacts before they reach the screen of medical professionals.

2.1 Model

The proposed approach uses a Denoising Diffusion Probabilistic model [16] with a UNet backbone that can be seen in Figure 1a and consists of ResNet blocks, illustrated in Figure 1b.



(a) Architecture overview



(b) Architecture of the ResNet block

Fig. 1: Model architecture.

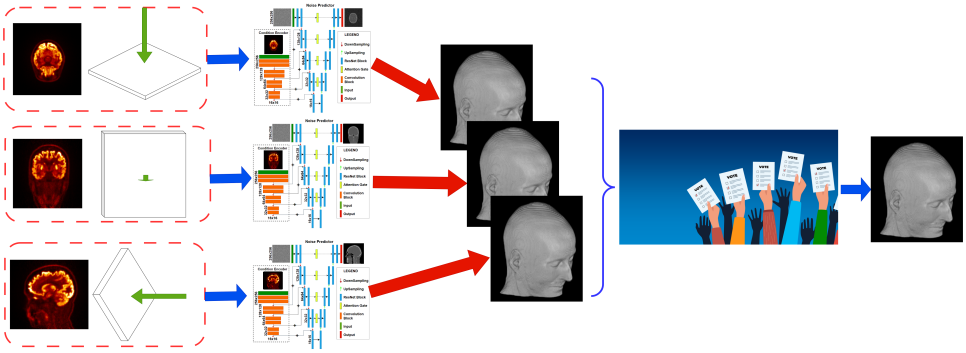


Fig. 2: Voting process of three models generating attenuation-corrected PET images of the same patient from 2D image slices along three orientations.

The proposed architecture consists of a noise predictor that removes the noise from CT images and a condition encoder that converts the PET image into a corresponding CT image. This condition encoder consists of convolution blocks that each use a 2D convolution, batch normalization and ReLU activation. The information from the condition encoder is passed to the noise predictor right after downsampling, by concatenation at stages 2 to 5. The model shown is represented for training with transverse image slices as the shape of input images is 256x256. In the case of sagittal and coronal training, this shape becomes 96x256 or 256x96 per 2D image. The attention gates come directly from Oktay [18] and are used to focus on relevant spatial regions. Each stage of the model, going from up (original image shape) to down has [64, 128, 256, 512, 1024] filters for the convolutions. All downsampling and upsampling steps are performed using convolutions and transposed convolutions with a (2, 2) stride and all convolutions use a (3x3) kernel size.

2.2 Voting

Three separate 2D models are developed, each utilizing a distinct slice orientation: transverse, sagittal, and coronal. The slices of each model are combined to create three 3D PET images of the same patient, which undergo pixel-by-pixel comparison. The voting process involves three scenarios: 1) when all three values fall within a predefined distance threshold, they are averaged together; 2) if two values meet the threshold criteria while one does not, the outlying value is excluded and the remaining two are averaged; and 3) when all three distances exceed the threshold, the two most similar values are averaged. Figure 2 illustrates this process.

2.3 Dataset

The model was trained with 159 paired PET/CT brain images from 64 men and 95 women, acquired using a *Siemens Biograph Vision* scanner. Of these, 79 used [^{18}F]-Fluorodeoxyglucose (FDG) and 80 used [^{11}C]-Acetoacetate (AcAc) [19][20]. Patients were on average 70 years old (range 55-81). For FDG, the tracer was injected 30 minutes before scanning, and for AcAc, immediately before. Scan times were 30 minutes for FDG and 18 minutes for AcAc, with doses of (338 $\pm$36) MBq and (190 $\pm$10) MBq, respectively. PET images were interpolated from (448, 448, 90) to (256, 256, 96) voxels, and CT images from (512, 512, 175) to the same resolution, with a voxel size of (1.17, 1.17, 2.75) mm, in order to make the data computationally manageable for training and to ensure consistent input dimensions across patients. All beds and

other items that do not represent the patient were segmented out of the CT image. Fifteen patients of the dataset were selected to test data (7 FDG, 8 AcAc), 4 for the validation (2 FDG, 2 AcAc) and 140 for training. All data was clipped between -1000HU and 3000HU to mitigate the effects of metal artifacts on normalization range. Normalization was done to obtain values between -1 and 1 using equations 1 and 2:

$$\widehat{NAC} = \left(\frac{NAC}{\max(NAC)} \times 2\right) - 1 \tag{1}$$

$$\widehat{CT} = \left(\frac{CT + 1000}{4000} \times 2\right) - 1 \tag{2}$$

2.4 Training

The dataset was expanded by 60% using data augmentation techniques including $\pm 15°$ rotations, $\pm 15\%$ zoom adjustments, and image flips. To augment images, we used a modified triangular distribution with uniform tails to bias sampling toward tissue-rich regions. This distribution assigns higher probabilities to central indices — where voids are less likely — and lower, uniform probabilities to outer regions. Specifically, the chosen index ranges were 0-83 (out of 96) in the transverse orientation, 30-235 in the coronal orientation, and 30-210 in the sagittal orientation (out of 256). To preserve anatomically plausible images, the augmentation strategies varied depending on orientation: extensive augmentation for transverse slices; only positive zoom and lateral flips for coronal slices; and exclusively positive zoom for sagittal slices. This approach yielded 21.560 transverse slices, 56,420 sagittal slices and 58,800 coronal slices for training. Noisy images were created using a cosine scheduler with 1,000 steps [21]. The model utilized SNR-weighted MSE loss (min-SNR5) for training to improve training stability and sample quality [22]. During inference, EMA weights decay [23] set to $\beta = 0.9995$ was used. The training process was initiated with 10 warm-up epochs, progressively raising the learning rate from 10^{-6} to 10^{-4} at each step with a cosine schedule. Subsequently, a second cosine scheduler reduced it from 10^{-4} to 10^{-6} over a 7 day period. Training was done on *Tensorflow 2.15.1* with four *Nvidia A100* 40GB GPUs using Tensorflow's mirrored strategy and mixed precision. A batch size of 80 per GPU for sagittal and coronal slices, and 40 for transverse slices was used. Training time depended on the orientation of the images, as some information was learned more quickly by the model. The transverse model took 3000 epochs to converge, while the sagittal model needed 3400 epochs and the coronal model 1200 epochs.

To reduce overfitting, validation metrics were evaluated every 200 epochs using a single batch of images. Generating these validation images required 28 minutes for the transverse model and 22 minutes for the sagittal and coronal models. The progression of these metrics is shown in Figure 3.

Test samples were generated on an *Nvidia A100* 40GB GPU, with batch sizes of 80, 120, and 120 for the transverse, sagittal, and coronal models, respectively. In the same order, generating the pCT image of one patient took 20, 22, and 22 minutes without mixed precision.

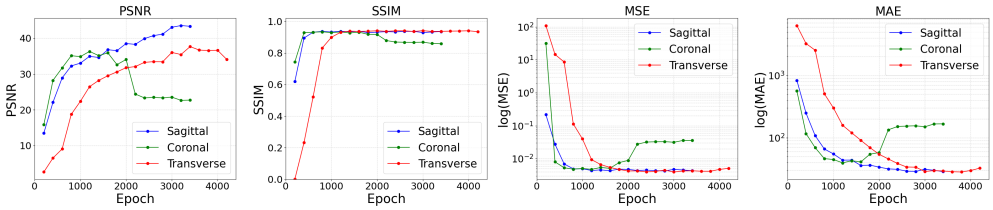


Fig. 3: Progression of the validation metrics through the training.

2.5 Evaluation of PET quantification

Reconstructed PET images were generated using Siemens' *OPTOF* reconstruction algorithm. To obtain sharper corrected PET images, the automatic blur normally applied to the uncorrected PET image was removed, although this modification may slightly affect the results. PET images were reconstructed using two attenuation maps: the real CT available in the dataset (with dimensions $256 \times 256 \times 96$) and the corresponding pCT of identical dimensions, obtained after applying the majority voting. The PET image used for reconstruction is the original image that was not interpolated ($448 \times 448 \times 90$). The final attenuation corrected PET image therefore had the same dimension. The percentage difference between both reconstructions was computed using Equation 3:

$$Err = \frac{I_{\mathrm{pCT}} - I_{\mathrm{CT}}}{I_{\mathrm{CT}}} \times 100\% \tag{3}$$

where I_{CT} denotes the voxel intensity of the PET image corrected using the real CT, and I_{pCT} denotes the voxel intensity of the PET image corrected using the pCT. Ten Regions of Interest (ROIs) were extracted for both radiotracers using atlases provided by *PMOD*. ROIs were selected to target regions in the brain that are studied for neurodegenerative diseases.

3 Results

The generated pCTs are visually compared with the CT and NAC PET AcAc images used to generate the pCTs in Figure 4. Each model's image shows a great generation of the skull and the brain with very little visual difference from the CT. All three models struggle in the sinus region, where predictions of cavities filled with air and cervical bones differ substantially. The transverse model shows lack of consistency in the airway and the back of the neck's shape is inconsistent. The sagittal model shows a smoother back neck generation, but fails to adequately generate the airway and has difficulties generating the cervical bones. The coronal model's image provides a clear and consistent airway, but it appears completely closed at the bottom of the image. The voting image shows a well shaped neck, but the issues from the airway persist as the majority of the models failed to correctly generate it. Patient 34028 is further analyzed in a difference map in Figure 5. Results of the average metrics of all 15 pCT from the test data set are quantitatively compared to their corresponding CT in Table 1.

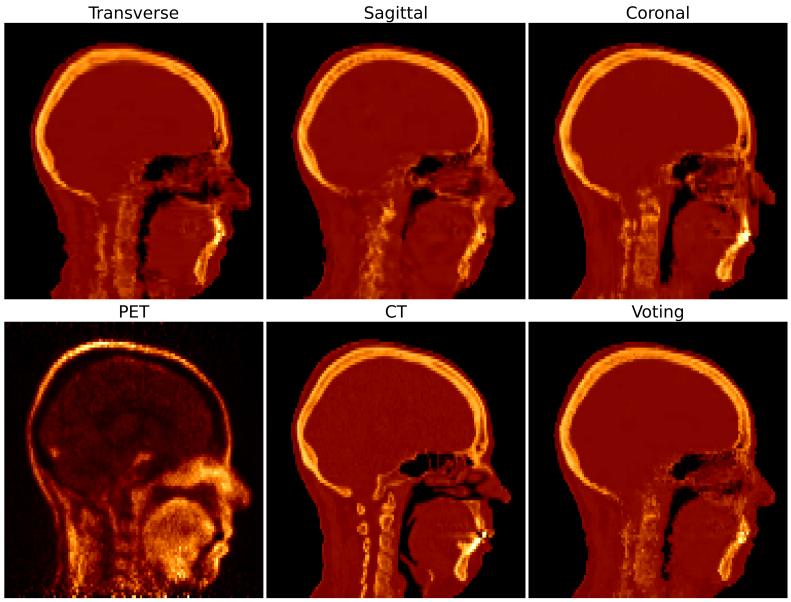


Fig. 4: Sagittal views of the pCTs generated by the Transverse, Coronal and Sagittal models, respectively, *(top)* compared to the NAC PET image *(bottom left)*, ground truth CT *(bottom middle)* and pCT after voting *(bottom right)* using the AcAc tracer in Patient 34028.

Table 1: Average metrics and standard error (STD) of the pCT for the 15 patients of the test data set compared to the corresponding CT for each model.

Method	**Transverse**	**Coronal**	**Sagittal**	**Voting**
PSNR (dB) (↑)	29.9 ± 2.3	29.4 ± 2.1	29.8 ± **2.0**	**30.1** ± 2.2
SSIM (↑)	0.924 ± 0.030	0.916 ± 0.032	0.919 ± 0.030	**0.925** ± **0.029**
MAE (HU) (↓)	31.7 ± 11.9	40.8 ± 10.9	**30.2** ± 10.8	32.1 ± **10.4**
RMSE (HU) (↓)	131.8 ± 33.9	138.7 ± 31.7	133.2 ± **29.5**	**128.5** ± 31.0
Avg Error (HU) (↓)	-0.15 ± 6.83	-12.15 ± 3.86	**0.34** ± 4.02	-3.80 ± **3.76**

* ↑: higher is better, ↓: lower is better

These metrics show that voting with a very loose 50 HU threshold helps the pCT to more closely match the CT. This is shown by a higher PSNR, SSIM, lower MAE, lower RMSE and lower STD of the average error. All models show similar STD on all metrics, showing that predictions are consistent. Without the voting method, all metrics, showing that predictions are consistent. Without the voting method, all three models show very similar results in similarity metrics (SSIM and PSNR). The coronal model has higher error metrics with the average error pointing to an average under estimation of the values by -12 HU.

Figure 6 shows that majority voting can remove artifacts generated inside the brain by the sagittal model along with other small artifacts that cannot be seen visually.

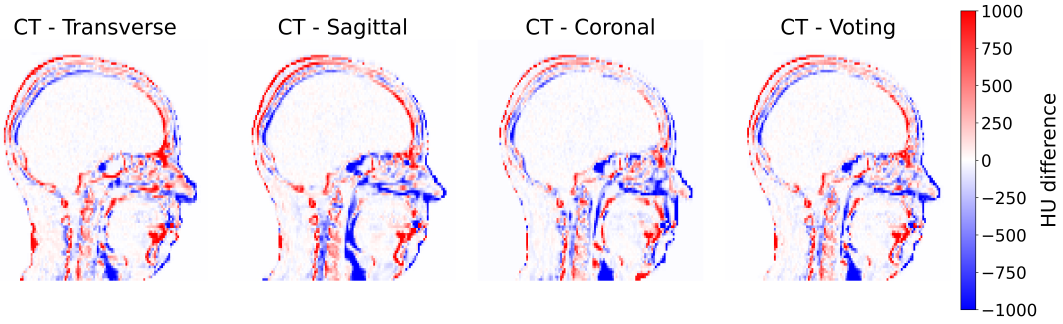


Fig. 5: Difference map of patient 34028 between the CT and pCT for each model's prediction and voting.

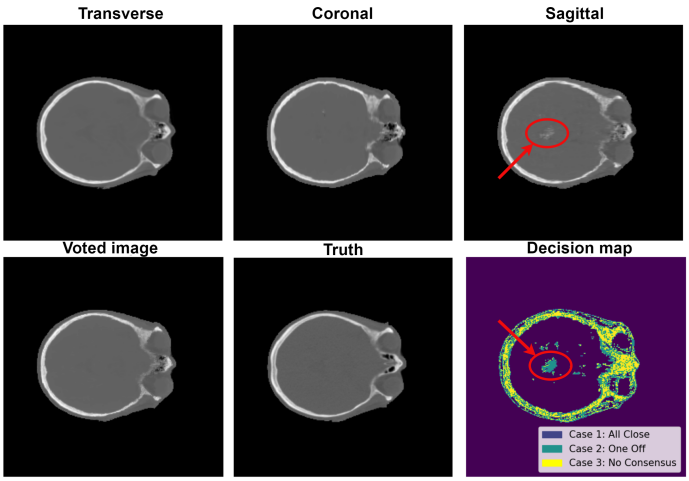


Fig. 6: Majority voting removing artifacts present in the sagittal model's image of patient 34024. Decision map: [Dark blue] All values from different images within threshold, [Green] One value above threshold, [Yellow] Two or more values above threshold.

Preliminary reconstructions were performed on four patients injected with an ^{18}F-FDG tracer, and the activity within each ROIs was calculated. The results are presented in Table 2. The mean error across all 116 ROIs provided by the atlas, which covers most of the brain, was $(1.46 \pm 0.68)\%$.

Figure 7 shows a visual comparison between a PET images reconstructed using a pCT and a real CT. Visually, for both tracers, the two images show close to no difference. The corresponding difference map supports this observation, showing an average pixel error of $(0.58 \pm 0.50)\%$ for the FDG and $(2.92 \pm 1.68)\%$ for the AcAc.

Table 2: ROI difference between CTAC and pCTAC PET images for 4 FDG patients.

Region	Mean ± STD [%]	Volume (ccm)
Amygdala	1.04 ± 0.85	7.49
Caudate	2.06 ± 0.93	31.30
Cingulum Post	1.73 ± 0.92	12.77
Frontal Mid	1.99 ± 1.45	159.47
Frontal Sup	2.39 ± 1.69	122.48
Hippocampus	1.08 ± 0.81	30.05
Parahippocampus	0.90 ± 1.02	33.76
Precuneus	2.13 ± 1.61	108.69
Temporal Mid	0.75 ± 0.45	149.62
Temporal Sup	1.02 ± 0.52	86.99

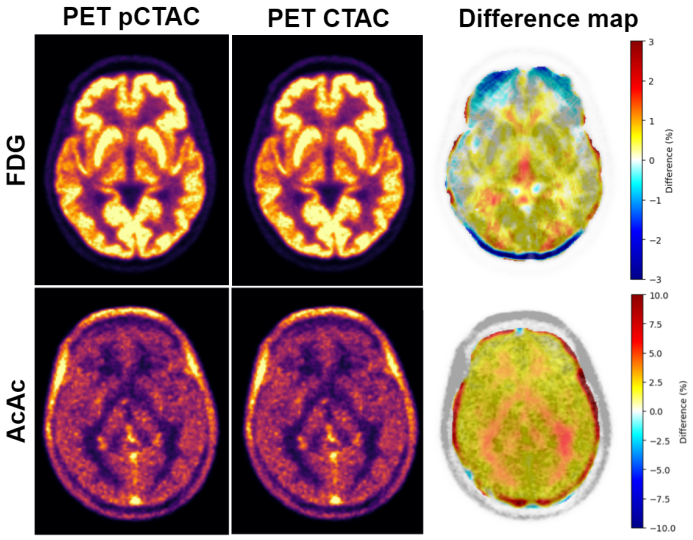


Fig. 7: Visual comparison of reconstructed PET images using a CT vs pCT attenuation map of patient 33609 FDG.

To assess the confidence of the three coronal, sagittal and transverse models, the same test dataset was generated three times per model. A strong confidence model should demonstrate minimal variability in its predictions. Table 3 reports similarity scores (PSNR and SSIM) that are extremely high across all models, showing near identical images. Overall, all scores indicate strong, confident models. The biggest variations occur towards the shoulders of the patient (if they are visible) as can be seen in Figure 8.

Table 3: Model's confidence analysis metrics for 3 generation of the same image for 15 patients.

Metric	Coronal	Transverse	Sagittal
Pairwise PSNR (dB) (↑)	61.2 ± 4.8	68.9 ± 11.3	63.1 ± 8.5
Pairwise SSIM (↑)	0.9998 ± 0.0006	0.9998 ± 0.0008	0.9998 ± 0.0004
Mean Coefficient of Variation (CV) (↓)	0.124 ± 0.464	0.091 ± 0.336	0.345 ± 0.969
Mean voxel-wise std (↓)	0.22 ± 0.23	0.27 ± 0.60	0.10 ± 0.05
95th percentile voxel std (↓)	0.60 ± 0.78	0.27 ± 0.17	0.17 ± 0.03

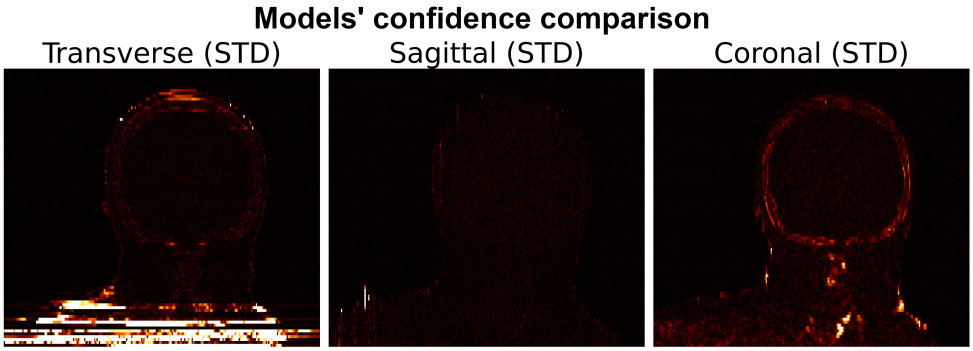


Fig. 8: STD variation by model for three predictions of the same patient (33011 AcAc). Displayed in the coronal view. Range of the displayed STD set from 0 to 5 HU.

The transverse model shows variation across its three predictions at the shoulders, with standard deviations reaching up to 500 HU for the same input image. In contrast, the sagittal model's predictions are almost identical, while the coronal model only shows minor variation, mostly in the bone regions. An example of a transverse PET slice that resulted in a high-variability pCT prediction of the shoulders is shown in Figure 9. It appears very noisy and with little to no recognizable structures. Applying majority voting reduces this variability and improves the prediction.

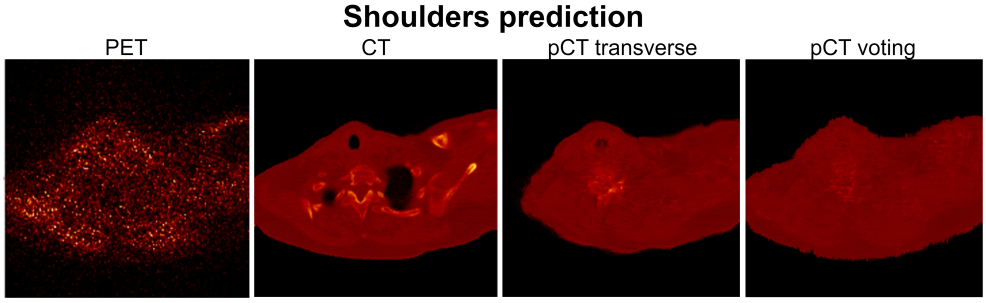


Fig. 9: Generation of a transverse and voting pCT image from a PET image at shoulders level compared to the CT for patient 33011 AcAc.

Figure 10 shows a patient in the test dataset that moved between the CT and PET acquisitions. The patient's head tilted to the right after the CT scan (in green) was completed, as evidenced by the noses not being aligned with the PET image (in red) on the left image. Looking closely at the nose of the PET image, the patient also moved during the PET acquisition as there is some ghosting to the right of the hotspot of the tip of the nose. The image generated by the model, on the right, is well aligned with the PET image as it only relies on the PET image to generate the pCT.

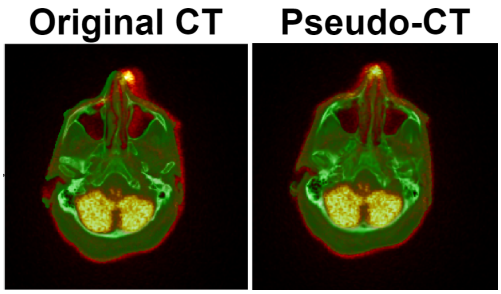


Fig. 10: Overlapping of a CT image (green) with a PET image (red) of patient 34113. [Left] Overlapping with the CT; [Right] Overlapping with the pCT.

Figure 11 shows a CT image with metal streak artifacts from dental implants, which produce over- and under-estimations across the image. The pCT generated by the transverse model substantially reduces these artifacts, but has parts of the chin missing (indicated by the green arrow). The use of majority voting corrects this error.

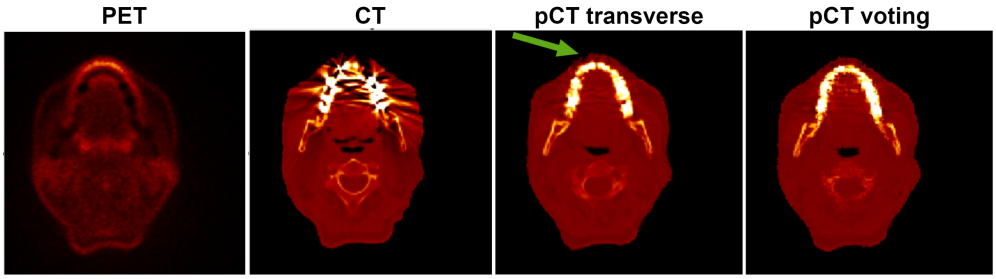


Fig. 11: CT and pCTs of Patient 34024 FDG with metal dental fillings.

4 Discussion

Only four reconstructions with a single radiotracer were shown quantitatively in this current article because it is still a work in progress. All reconstructions with both tracers will be included in the final version of the article.

The bed and other external objects (e.g., oxygen tubes) were segmented out of the CT images so the model could focus only on relevant anatomical information. This segmentation was not applied to the PET images, since doing so would create artificial discontinuities (regions set to zero) that would disrupt the natural noise pattern of PET images. These discrepancies may increase the risk of the model learning irrelevant artifacts. For instance, these inconsistencies between the PET and CT images might introduce attenuation patterns that do not appear in the corresponding CT image. Although training without segmentation of external objects was not tested,

the results obtained using segmented CT images suggest that the model effectively learns to disregard these inconsistencies. Instead, the model focuses on the relevant element of the image: the patient's head.

The choice to use a threshold value for voting of 50 HU is meant to increase the accuracy for the brain area. Most commonly found matter inside this region ranges approximately from 0 (cerebrospinal fluid) to 45 HU (gray matter). Predicted values outside this range can be considered aberration and should not be included in the final image. Voting has little effect on regions with higher values, like bones, where simple fluctuations in these regions can be more than the defined threshold, leading to the model simply averaging these areas in most cases.

Training took a different number of epochs for each model, as seen in Figure 3, with the most noticeable difference coming from the coronal model that took about 2000 less epochs to converge compared to the sagittal and transverse models. There is no clear explanation for this behaviour. All models had the exact same training parameters and seeds, the only difference being the orientation of the data fed to the models. One argument can be that data in the coronal view is more repetitive: A human head is longer (coronal) than it is large (sagittal), resulting in more anatomical information seen per epoch. This, added with the fact that the coronal shape is quite repetitive between image slices could be the reason of this faster training yielding similar results to the other models.

This is an important consideration when looking at the computational resources required to train such models. In our case, the training time to convergence (no validation calculations, purely time per epoch time number of epoch) was 56.6 hours for the coronal model, 112.5 hours for the transverse model, and 153 hours for the sagittal model. This shows a large disparity in training costs. With limited resources, training a single coronal model is far more efficient, as it achieves comparable results while requiring only a third of the computation time. Figure 4 reveals that diffusion models have their greatest inaccuracies in the nose and sinuses. These regions contain air and thin, low-uptake tissue with minimal detail in the input PET image, making these features difficult to generate accurately. This also applies to cervical bones that show low uptake for both FDG and AcAc tracers, making it also difficult to generate the corresponding image. These regions are also unique to each patient, where everyone has varying sinus [24] and cervical bone [25] shapes. The learning of these features therefore becomes very challenging for the model that only has access to the PET image. The voting image contains features of all three images, but without any significant visual changes. Images from both radiotracers yielded similar results.

The voting method offers many advantages, such as improving most metrics and effectively removing artifacts produced by individual models, as shown in Figure 6. We compared it with two alternatives: simple averaging and taking the median value. While both averaging and median methods achieved slightly better quantitative metrics, they come with drawbacks. Averaging provides good visual results, but does not remove artifacts as it only attenuates them, not solving the artifact problem. The median method can remove artifacts, but it discards valuable information from other models, leading to visual differences in some regions. Voting combines the strengths of both approaches, producing high-quality images that are both visually convincing and artifact-free.

Typically, in clinical images, all uncorrected PET images are systematically blurred using a heavy gaussian filter before applying the attenuation map. This produces the blurred PET images commonly used in the clinic, but for demonstration purposes we decided to skip this filtering step. Gaussian smoothing reduces high-frequency differences between the images, which are typically the main contributors to error metrics. As a result, omitting the filter tends to worsen these metrics. Nevertheless, it provides sharper reconstructed PET images, which is more valuable for us, as it allows precise identification of where the errors occur inside the images.

Visually, looking at Figure 5, all the models show a strong performance inside the brain region where maximal errors can reach up to 50 to 100 HU. Most errors come from the bones that have a lower uptake value and, therefore have less information for a prediction. Table 1 shows the coronal model is slightly worse in terms of errors. This can be attributed to the smaller training time resulting in a shorter convergence point for the model. Indeed, the validation metrics were taken every 200 epochs and looking at Figure 3, there is a steep drop in metrics quality between epoch 1200 (best epoch) and 1400. This shows that the true convergence point may lie at epoch 1200 ± 200. This point also applies to the other two models, but the drop in metrics quality before and after convergence is more subtle than for the coronal model.

This issue could have been avoided with more frequent validation, but we initially did not expect the coronal model to converge so quickly. However, increasing validation frequency comes at a cost: generating a single batch of validation images takes 22 or 28 minutes, depending on the model. For context, the transverse model required 9.8 hours just to generate validation samples over its 4200 training epochs. More frequent validation steps could be used, but would significantly extend total training time and resource usage.

Even with a threshold set to 50 HU (a value determined based on the brain's values), voting helps the overall image scores to an extent where is outperforms the other models in PSNR, SSIM, MAE STD, RMSE and average STD error. Although it was discussed that voting might have minimal impact on bones or structures with higher HU values, two models can still agree on the value while the third model provides a value above the threshold. This scenario is less likely due to the amplitude of the values varying much, as mentioned, but can still help generating an improved image - In the worst case, it will simply average all values, so if one model really messes up a prediction, the other two models will attenuate the repercussions of this wrong prediction.

Results in Table 2 show very small difference between the PET images reconstructed with a CT versus a pCT for FDG. AcAc tracer shows a higher average error at figure 7, because it has lower activity inside the brain, therefore, the statistical error is increased and gives these higher error numbers, even though the images look visually like the same quality. In Figure 7, it can be seen that the main errors occur at bone/air interfaces where significant density changes take place. Precisely identifying where these structures end can be difficult due to partial volume effects or patient movement during PET imaging, which explains the pseudo-CT's struggles in these areas. This shows, as the hippocampus (1.08 ± 0.81) and parahippocampus (0.90 ± 1.02) that are deep inside the brain have the lowest error, while the frontal mid (1.99 ± 1.45) and sup (2.39 ± 1.69) are on the edge of the brain and show higher error.

Due to the stochastic nature of DDPM models, each generated prediction differs slightly, because random noise is used for each image generation. Verifying whether the model can produce consistent outputs from different noise initializations is essential, as it reflects the model's confidence. An uncertain model will generate varying images with the same input condition, but a confident model will consistently produce the same result. As shown in Figure 8, the transverse model struggles to consistently reconstruct the shoulders. This happens because the PET image used to generate the transverse pCT, seen in Figure 9, is very noisy: the shoulders show low tracer uptake and are further affected by edge-of-field noise amplification. As a result, the image contains very little usable information for the model to generate a patient-specific reconstruction. In contrast, the coronal and sagittal models are less affected, since edge-of-field noise amplification primarily occurs in the axial FOV. This would suggest that only the coronal or sagittal model should be used to generate the bottom transverse slices as they generate more confident results.

The left image in Figure 10 shows a patient who moved between the CT and PET scans. This is a common clinical challenge where the CT is acquired in just a few seconds, while the PET scan lasts several minutes. Patient movement in between can compromise attenuation correction, potentially leading to poorer image quality and misdiagnosis. Generating a pCT image from PET alone completely avoids this issue, as the pCT is always perfectly aligned with the PET, as illustrated by the right image in Figure 10. We could even argue that in case the case where a patient moved, the pCT is more realistic than the CT for attenuation correction.

Even small patient movements between scans can complicate model training. In most cases, the CT is well aligned with the PET, but when misalignment occurs, the model is penalized for producing a pCT image aligned with the PET image, which can degrade training performance. In addition, such misalignment affects evaluation metrics, where in Table 1, the CT-PET mismatch is not accounted for, meaning the reported metrics are pessimistic and would improve if all data were perfectly aligned. Furthermore, all metrics except SSIM are evaluated pixel by pixel meaning that a slight movement will impact negatively the metrics. We assumed that majority of patients inside the dataset moved a negligible amount and we can confirm this because all pCT predictions look correctly aligned with the PET.

As shown in Figure 11, some patients in the dataset had dental fillings, which produced metal streak artifacts [26]. These artifacts create over- and underestimations in the transverse view and can mislead the model, which may attempt to reproduce them. Fortunately, because many patients in the training dataset did not show such artifacts, the model can learn to generate artifact-free images. This is something conventional CT cannot reliably achieve, highlighting a potential advantage of this AI-based approach. Furthermore, training on a dataset that entirely excludes patients with dental fillings might prevent the model from ever generating artifacts in such cases, simply because it never learned them. In the current setup, the model is penalized whenever it generates a clean image while the ground truth still contains artifacts.

Many limitations were encountered in this project. First, the dataset didn't contain any particular conditions or pathologies. An image could have something the model has never seen such as a large tumors, metal plates, missing skull parts, neuro-degenerative diseases, etc. It is unknown how the model would react in these

scenarios and to eventually deploy this model, these concerns must be addressed.

Although majority voting shows great potential, its application in this work was rather simplistic. We acknowledge that more complex implementation may produce better results with the three images provided by the models.

Additionally, data obtained from different PET scanners often show considerable differences in resolution, noise levels, contrast, and overall image texture. Since neural networks are highly sensitive to such variations, a model trained on data from a single scanner may capture some scanner-specific characteristics instead of true anatomical or physiological information. When this model is tested on images from another scanner, the mismatch can result in degraded performance, reduced accuracy, visible artifacts, or limited generalization ability.

Furthermore, patient movement introduces an additional challenge where it causes variability in the ground truth that the model is trained to learn. One possible solution is to use an image registration algorithm to align all scans, but this approach can also introduce bias into the dataset from the algorithm itself.

5 Conclusion

This work introduced a new approach using Multiview Denoising Diffusion Probabilistic Models to generate convincing pCT images for attenuation correction of PET images . The use of majority voting with a simple threshold of 50 HU improves almost all metrics of the pCT while removing artifacts that may have been generated by one of the transverse, coronal or sagittal models.

With majority voting, PSNR improved from (29.93 $\pm$ 2.26) dB from the best model (transverse) to (30.13 $\pm$ 2.16) dB for the voting and RMSE was reduced from (131.79 $\pm$ 33.93) HU (transverse) to (128.46 $\pm$ 31.00) HU. The nasal and sinus areas present difficulties for the model, as PET images provide limited information about these anatomically complex regions, but the model showed strong performance in the brain region.

We demonstrated that this approach can successfully generate accurate, well aligned and artifact free deep learning-based pCT images from non-attenuation-corrected PET data, allowing attenuation correction of brain PET scans acquired with two different radiotracers showing distinct uptake distributions. By eliminating the need for additional imaging modalities and avoiding extra radiation exposure, this method has the potential to make brain imagingsafer, more cost-effective and free from misalignment errors.

Declarations

Ethics approval: Data availability: The dataset used and analyzed in this study is available from the corresponding author on reasonable request.
Funding: Competing interests: RL is both with the Université de Sherbrooke and a co-founder and Chief Scientific Officer of IR&T Inc., which funded part of this work

through the Acuity-QC Consortium. SC consults for and has received research funding and research materials from Nestlé Health Science. He also consults for Cerecin. Other authors have no competing interests to declare relevant to the content of this article.

Author contributions:

ASG: Conceptualization, DDPM implementation, Software development, Data analysis, Writing original draft, review & editing;
MT: Supervision, Data analysis, Review & editing;
CT, ÉA: Technical assistance, Review & editing;
ÉC, SC: Data procurement, Review & editing;
RL: Funding acquisition, Supervision, Review & editing;
JBM: Funding acquisition, Conceptualization, Supervision, Review & editing.
All authors gave their approval for the final version of the manuscript.

Acknowledgements: Not applicable.

References

[1] Wanek, T., Schweifer, A., Mair, C., Unterrainer, M., Schmid, A.I., Stadlbauer, A., al.: Impact of attenuation correction on quantification accuracy in preclinical whole-body PET images. Front Phys **8** (2020) https://doi.org/10.3389/fphy.2020.00123

[2] Burger, C., Goerres, G., Schnyder, P., Buck, A., Schulthess, G.K., Flueckiger, F.: PET attenuation coefficients from CT images: experimental evaluation of the transformation of CT into PET 511-keV attenuation coefficients. Eur J Nucl Med Mol Imaging **29**(7), 922–927 (2002) https://doi.org/10.1007/s00259-002-0796-3

[3] Keereman, V., Fierens, Y., Berghe, T.V., Vandenberghe, S.: Challenges and current methods for attenuation correction in PET/MR. Magn Reson Mater Phy **26**(1), 81–98 (2013) https://doi.org/10.1007/s10334-012-0334-7

[4] Ritt, P., Sanders, J.: Spect/ct technology. Clin Transl Imaging **2**, 445–457 (2014) https://doi.org/10.1007/s40336-014-0086-7

[5] Sekine, T., Buck, A., Delso, G., Voert, E.E.G.W., Huellner, M., Veit-Haibach, P., Warnock, G.: Evaluation of atlas-based attenuation correction for integrated pet/mr in human brain: Application of a head atlas and comparison to true ct-based attenuation correction. Journal of Nuclear Medicine **57**(2), 215–220 (2016) https://doi.org/10.2967/jnumed.115.159228

[6] Kim, J.H., Lee, J.S., Song, I.C., Lee, D.S.: Comparison of segmentation-based attenuation correction methods for pet/mri: evaluation of bone and liver standardized uptake value with oncologic pet/ct data. Journal of Nuclear Medicine **53**(12), 1878–1882 (2012) https://doi.org/10.2967/jnumed.112.104109

[7] Sekine, T., Voert, E.E.G.W., Warnock, G., Buck, A., Huellner, M., Veit-Haibach, P., Delso, G.: Clinical evaluation of zero-echo-time attenuation correction for brain 18f-fdg pet/mri: Comparison with atlas attenuation correction. Journal of Nuclear Medicine **57**(12), 1927–1932 (2016) https://doi.org/10.2967/jnumed.116.175398

[8] G. Krokos, J.D. J. MacKewn, Marsden, P.: A review of pet attenuation correction methods for pet-mr. EJNMMI Physics **10**, 52 (2023) https://doi.org/10.1186/s40658-023-00569-0

[9] Bailey, D.L.: Transmission scanning in emission tomography. Eur J Nucl Med **25**(7), 774–787 (1998) https://doi.org/10.1007/s002590050282

[10] Raymond, C., Jurkiewicz, M.T., Orunmuyi, A., Liu, L., Dada, M.O., Ladefoged, C.N., Teuho, J., Anazodo, U.C.: The performance of machine learning approaches for attenuation correction of pet in neuroimaging: A meta-analysis. Journal of Neuroradiology **50**(3), 315–326 (2023) https://doi.org/10.1016/j.neurad.2023.01.157

[11] Lee, J.S.: A review of deep-learning-based approaches for attenuation correction in positron emission tomography. IEEE Trans Radiat Plasma Med Sci **5**(2), 160–177 (2021) https://doi.org/10.1109/TRPMS.2020.3009269

[12] Liu, F., Jang, H.-Y., Kijowski, R., Bradshaw, T., McMillan, A.B.: A deep learning approach for 18f-fdg pet attenuation correction. EJNMMI Physics **5**(1), 24 (2018) https://doi.org/10.1186/s40658-018-0225-8

[13] Armanious, K., Küstner, T., Reimold, M., Nikolaou, K., La Fougère, C., Yang, B., Gatidis, S.: Independent brain 18f-fdg pet attenuation correction using a deep learning approach with generative adversarial networks. Hellenic Journal of Nuclear Medicine **22**(3), 179–186 (2019) https://doi.org/10.1967/s002449911053

[14] Jafaritadi, M., Anaya, E., Chinn, G., Levin, C.S.: Multi-modal generative vision transformer for attenuation and scatter correction in brain pet. In: 2023 IEEE Nuclear Science Symposium, Medical Imaging Conference and International Symposium on Room-Temperature Semiconductor Detectors (NSS MIC RTSD), pp. 1–1 (2023). https://doi.org/10.1109/NSSMICRTSD49126.2023.10338712

[15] Chen, T., Hou, J., Zhou, Y., Xie, H., Chen, X., Liu, Q., Guo, X., Xia, M., Duncan, J.S., Liu, C., Zhou, B.: 2.5D Multi-view Averaging Diffusion Model for 3D Medical Image Translation: Application to Low-count PET Reconstruction with CT-less Attenuation Correction (2024). https://arxiv.org/abs/2406.08374

[16] Ho, J., Jain, A., Abbeel, P.: Denoising diffusion probabilistic models. In: Advances in Neural Information Processing Systems (NeurIPS), vol. 33, pp. 6840–6851 (2020)

[17] Kossale, Y., Airaj, M., Darouichi, A.: Mode collapse in generative adversarial networks: An overview. In: 2022 8th International Conference on Optimization and Applications (ICOA), pp. 1–6 (2022). https://doi.org/10.1109/ICOA55659.2022.9934291

[18] Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., Glocker, B., Rueckert, D.: Attention U-Net: Learning Where to Look for the Pancreas (2018). https://arxiv.org/abs/1804.03999

[19] Bentourkia, M., Tremblay, S., Pifferi, F., Rousseau, J., Lecomte, R., Cunnane, S.: Pet study of 11c-acetoacetate kinetics in rat brain during dietary treatments affecting ketosis. American Journal of Physiology-Endocrinology and Metabolism **296**(4), 796–801 (2009) https://doi.org/10.1152/ajpendo.90644.2008

[20] St-Pierre, V., Richard, G., Croteau, E., Fortier, M., Vandenberghe, C., Carpentier, A.C., Cuenoud, B., Cunnane, S.C.: Cardiorenal ketone metabolism in healthy humans assessed by ^{11}C-acetoacetate pet: effect of d-β-hydroxybutyrate, a meal, and age. Frontiers in Physiology **15**, 1443781 (2024) https://doi.org/10.3389/fphys.2024.1443781

[21] Nichol, A., Dhariwal, P.: Improved Denoising Diffusion Probabilistic Models (2021). https://arxiv.org/abs/2102.09672

[22] Hang, T., Gu, S., Li, C., Bao, J., Chen, D., Hu, H., Geng, X., Guo, B.: Efficient Diffusion Training via Min-SNR Weighting Strategy (2024). https://arxiv.org/abs/2303.09556

[23] Karras, T., Aittala, M., Lehtinen, J., Hellsten, J., Aila, T., Laine, S.: Analyzing

and Improving the Training Dynamics of Diffusion Models (2024). https://arxiv.org/abs/2312.02696

[24] Ağcayazı, S., Salcan, I., Erşahan, A., Seçkin, E.: Sinonasal anatomic variations in the adult population: Ct examination of 1200 patients. Nigerian Journal of Clinical Practice **27**(8), 990–994 (2024) https://doi.org/10.4103/njcp.njcp_275_24

[25] Palancar, C.A., García-Martínez, D., Cáceres-Monllor, D., Perea-Pérez, B., Ferreira, M.T., Bastir, M.: Geometric morphometrics of the human cervical vertebrae: sexual and population variations. Journal of Anthropological Sciences (JASS) **99**, 97–116 (2021) https://doi.org/10.4436/JASS.99015

[26] Barrett, J.F., Keat, N.: Artifacts in CT: recognition and avoidance. Radiographics **24**(6), 1679–1691 (2004) https://doi.org/10.1148/rg.246045065