Larger Hausdorff Dimension in Scanning Pattern Facilitates Mamba-Based Methods in Low-Light Image Enhancement

Xinhua Wang
Imperial College London
London, UK
mw1821@ic.ac.uk

Caibo Feng
University of Sussex
Brighton, UK
2012190104@pop.zjgsu.edu.cn

Xiangjun Fu
University of California San Diego
San Diego, USA
xif001@ucsd.edu

Chunxiao Liu
Zhejiang Gongshang University
Hangzhou, China
cxliu@mail.zjgsu.edu.cn

Abstract

In the domain of low-light image enhancement, both transformer-based approaches, such as Retinexformer and Mamba-based frameworks, such as MambaLLIE, have demonstrated distinct advantages alongside inherent limitations. Transformer-based methods, in comparison with mamba-based methods, can capture local interactions more effectively, albeit often at a high computational cost. In contrast, Mamba-based techniques provide efficient global information modeling with linear complexity, yet they encounter two significant challenges: (1) inconsistent feature representation at the margins of each scanning row and (2) insufficient capture of fine-grained local interactions. To overcome these challenges, we propose an innovative enhancement to the Mamba framework by increasing the Hausdorff dimension of its scanning pattern through a novel Hilbert Selective Scan mechanism. This mechanism explores the feature space more effectively, capturing intricate fine-scale details and improving overall coverage. As a result, it mitigates information inconsistencies while refining spatial locality to better capture subtle local interactions without sacrificing the model's ability to handle long-range dependencies. Extensive experiments on publicly available benchmarks demonstrate that our approach significantly improves both the quantitative metrics and qualitative visual fidelity of existing Mamba-based low-light image enhancement methods, all while reducing computational resource consumption and shortening inference time. We believe that this refined strategy not only advances the state-of-the-art in low-light image enhancement but also holds promise for broader applications in fields that leverage Mamba-based techniques. Our code is available ***here.***

1 Introduction

Low-light image enhancement is an essential but challenging topic in computer vision. It aims to restore the image degradation with low brightness, improve the visual quality of images captured under low-light conditions, and aid high-level visual tasks (e.g., object detection, face recognition, and action detection).

Numerous traditional methods, such as histogram equalization [3, 14] and Retinex theory [9, 18], have been proposed to enhance low-light images. With the development of deep learning, many learning-based methods [2, 19, 24, 30, 39, 53, 56, 57] have emerged, which improve visual quality by learning the mapping between low-light and normal-light images using end-to-end models or deep Retinex-based models to make them more robust than traditional approaches. Recently, some methods [16, 45] have explored wavelet transformation for low-light image enhancement, integrating wavelet information into convolutional neural networks (CNNs) or transformer models to achieve impressive results. However, CNNs [56] have limitations in modeling long-range dependencies and non-local self-similarity, resulting in challenges in addressing image degradation effectively in low-light scenarios. In addition, the quadratic growth of the complexity of transformer models [2] results in inefficient use of computational resources. Both the CNN model [12] and the transformer model [46] achieve average performance only in both aspects of performance, as indicated by the PSNR scores and inference time.

To overcome these issues, Mamba [7] is designed to model long-range dependencies and enhance the efficiency of training and inference through a selection mechanism and a hardware-aware algorithm. For now, numerous studies have explored the applications of Mamba in computer vision. Vision Mamba [59] introduces the Vim block, incorporating a bidirectional state space model for efficient learning. Meanwhile, VMamba [25] introduces a CSM module to traverse the spatial domain, combining the advantages of CNNs and Visual Transformers (ViTs) while achieving computational efficiency in linear complexity without sacrificing the global receptive field. Additionally, U-Mamba [29] combines U-Net and Mamba models to enhance medical image segmentation. SegMamba [44] utilises Mamba's efficient reasoning and linear scalability to enable fast and accurate processing of large-scale 3D medical data. Additionally, WaveMamba [60], RetinexMamba [1], and MambaLLIE [40] demonstrate marvelous performance of Mamba in the field of low-light image enhancement, all of which generally applies Mamba along with U-Net backbone, significantly reducing the computational complexity compared with Retinexformer [2].

However, several challenges arise when incorporating Mamba with vision tasks. The primary challenge originates from the mismatch between the causal sequential modeling of Mamba and the two-dimensional (2D) data structure of images. Mamba is designed for one-dimensional (1D) causal modeling for sequential signals, which cannot be directly leveraged for modeling two-dimensional image tokens. A simple solution is to use the raster-scan order to convert 2D data into 1D sequences. However, it restricts the receptive field of each location to only the previous locations in the raster-scan order, which means such type of raster scan, in

some cases, fails to capture local interactions when dealing with features with more extensive areas in the images. Moreover, in the raster-scan order, the ending of the current row is followed by the beginning of the next row, while they do not share spatial continuity.

To accurately evaluate the ability of a scanning path to capture complicated patterns in an image, we propose the Hausdorff Dimension Measurement Method. The Hausdorff dimension of a scanning curve directly reflects its inherent complexity and capacity to capture local interactions. A higher Hausdorff dimension indicates that the curve is not simply a smooth, predictable path but rich in intricate detail and subtle fluctuations. This complexity enables the scanning pattern to more effectively perceive local variations, ensuring that fine-scale interactions within the scanned area are not overlooked. In other words, as the Hausdorff dimension increases, it signals a more nuanced and densely packed trajectory that can more effectively capture local features while maintaining the ability of Mamba to understand long-range dependencies.

Besides, we conduct extensive experiments to show that Mamba scanning patterns with higher Hausdorff dimension, Hilbert scan, and Peano scan, inspired by Hilbert curve [13] and Peano curve [33], can effectively prompt the performance of previous Mamba-based low-light image enhancement methods on several paired datasets. Hilbert and Peano scans not only leverage the Hausdorff Dimension of 2 but also map neighboring 2D points to nearby positions in the 1D sequence, which minimizes discontinuities and improves cache performance and data coherence. The visual presentations of different scanning patterns are provided in the supplementary material.

Our main contributions are summarized as follows:

- We propose an innovative method for evaluating scanning paths by employing the Hausdorff dimension. This metric measures the inherent complexity of a scanning curve, thereby assessing its capacity to capture intricate, fine-scale local variations and to effectively navigate complex image structures.
- We proved that scanning patterns with high Hausdorff Dimension (eg, Hilbert scan, and Peano scan) leverage the curve's intrinsic properties to capture local features effectively in larger regions, mapping neighboring 2D points to adjacent positions in the 1D sequence.
- Experimental results on a comprehensive set of benchmark datasets consistently demonstrate the reliability of the measure of Hausdorff Dimension and the effectiveness of Hilbert scan and Peano scan on improving the performance of existing Mamba-based low-light image enhancement methods.

2 Related Work

2.1 Low-light Image Enhancement

Following the development of learning-based image restoration methods [21, 22, 36, 48], LLNet [26] synthesizes paired data by applying gamma adjustment and randomly adding noise to clean images. MBLLEN [28] extracts features at different levels using a multi-scale network structure to achieve better results. LightenNet [20] directly estimates the illumination map based on the input. RetinexNet [39] utilizes the Retinex theory to decompose low-light images into reflection and illumination components. ZeroDCE [10] proposed to transform LLIE into a curve estimation problem and designed a zero-reference learning strategy for training. EnlightenGAN [17] adopted unpaired images for training for the first time by using the generative adverse network as the main framework. KinD++ [56] utilizes a layer-wise decomposition strategy for brightness adjustment and detail reconstruction to suppress noise in the reflection layer further. LLFlow [37] uses conditional normalization flows for low-light image enhancement. Restormer [51] designs a lightweight transformer for image restoration tasks. SNRNet [46] introduces an SNR-aware CNN-transformer hybrid network for low-light image enhancement. MBPNet [54] introduces a multi-branch progressive network for low-light image enhancement. Bread [12] decomposes low-light images into texture and chrominance components and suppresses complex noise through adjustable noise suppression networks. Retinexformer [2] combined the Retinex theory with the design of a one-stage transformer, further refining and optimizing this approach. Diff-Retinex [50] designed a transformer-based decomposition network and adopted generative diffusion networks to reconstruct the results. Overall, they typically applied the Retinex theory directly, which may be limited for low light enhancement problems.

2.2 Mamba-based Methods

Recently, the success of State Space Models (SSMs) [8] has been increasingly recognized as a promising direction in research, which is proposed as a novel alternative to CNNs or Transformers to model long-range dependency. Contemporary SSMs such as Mamba [7] not only establish long-distance dependency relations but also demonstrate linear complexity with respect to input size. Other methods introduce the state space models for visual applications. UMamba [29] proposes a hybrid CNN-SSM architecture to handle the long-range dependencies in biomedical image segmentation. Vision Mamba [59] suggests that pure SSM models can serve as a generic visual backbone. SegMamba [44] introduces a novel 3D medical image segmentation model designed to capture long-range dependencies within volume features at every scale effectively. Swin-UMamba [23] introduces a medical image segmentation model based on Mamba, which utilizes pre-trained models from ImageNet to improve the performance of medical image segmentation tasks. Furthermore, LocalMamba [15] was focused on the local scanning strategy and preservation of local context dependencies. EfficientVMamba [34] designed a lightweight SSMs with an additional convolution branch to learn both global and local representational features. MambaIR [11] employed convolution and channel attention to enhance the capabilities of the Mamba. WaveMamba [60] focuses on applying Wavelet Transform alongside Mamba to mitigate information loss and address the limitation of SSMs, which struggle to model noise effectively. In the case of low-light image enhancement, this insensitivity could lead to the inability to detect or leverage subtle noise patterns that carry important information effectively. RetinexMamba [1] uses SS2D to replace Transformers in capturing long-range dependencies. MambaLLIE [40] introduces a novel global-then-local state space block that integrates a local-enhanced state space module and an implicit

Retinex-aware selective kernel module. This design effectively captures intricate global and local dependencies. However, although a number of applications of the Mamba-based image restoration approach have been proposed, none of them have researched the factors determining the effectiveness of a scanning pattern of Mamba.

3 Preliminaries

3.1 State Space Model

SSMs, such as the structured state space sequence models (S4) [8] and Mamba [7], can be interpreted as continuous linear time-invariant (LTI) systems [41]. Given a one-dimensional input sequence $x(t) \in \mathbb{R}$, these models map it to an output sequence $y(t) \in \mathbb{R}$ by means of a hidden state $h(t) \in \mathbb{R}^m$, where m is the dimension of the hidden state. The entire system is governed by the following linear ordinary differential equations:

$$\begin{aligned} h'(t) &= \mathbf{A}\, h(t) + \mathbf{B}\, x(t), \\ y(t) &= \mathbf{C}\, h(t) + \mathbf{D}\, x(t). \end{aligned} \tag{1}$$

Here, $\mathbf{A} \in \mathbb{R}^{m\times m}$ is the state matrix, $\mathbf{B} \in \mathbb{R}^{m\times 1}$ represents the input projection, $\mathbf{C} \in \mathbb{R}^{1\times m}$ is the output projection, and $\mathbf{D} \in \mathbb{R}$ denotes the feedthrough parameter.

Since these state-space models are inherently continuous, they must be discretized for implementation on a computer. Using the zero-order hold (ZOH) method, the continuous matrices $\mathbf{A}$ and $\mathbf{B}$ are converted into their discrete counterparts $\bar{\mathbf{A}}$ and $\bar{\mathbf{B}}$ as follows:

$$\bar{\mathbf{A}} = \exp(\Delta\,\mathbf{A}), \quad \bar{\mathbf{B}} = (\Delta\,\mathbf{A})^{-1}\left(\exp(\Delta\,\mathbf{A}) - \mathbf{I}\right)\cdot \Delta\,\mathbf{B}, \tag{2}$$

where Δ is the step size. Thus, the discrete formulation becomes:

$$h_t = \bar{\mathbf{A}}\, h_{t-1} + \bar{\mathbf{B}}\, x_t, \quad y_t = \mathbf{C}\, h_t + \mathbf{D}\, x_t. \tag{3}$$

However, this formulation remains static with respect to varying inputs. To overcome this limitation, Mamba [7] introduces selective state-space models, where the parameters dynamically adapt based on the input, which is formulated as:

$$\overline{\mathbf{B}} = f_{\mathbf{B}}(x_t), \quad \overline{\mathbf{C}} = f_{\mathbf{C}}(x_t), \quad \Delta = \vartheta_{\mathbf{A}}\Big(\mathbf{P} + f_{\mathbf{A}}(x_t)\Big), \tag{4}$$

with $f_{\mathbf{B}}(x_t)$, $f_{\mathbf{C}}(x_t)$, and $f_{\mathbf{A}}(x_t)$ being linear functions that expand the input features into the hidden state space. Although SSMs are effective at modeling long sequences, they may struggle to capture complex local details. To address this challenge in visual data, methods such as VMamba [25] and Vim [59] employ specialized location-aware scanning strategies that preserve the two-dimensional structure of images.

3.2 Hausdorff Dimension

The Hausdorff dimension provides a robust and precise method of quantifying the complexity of geometric objects, including fractals, whose dimensions are often non-integer and, therefore, cannot be described by traditional geometric measures [5, 31].

To understand Hausdorff Dimension, we must first introduce the definition of Hausdorff Measure. Let $S \subseteq \mathbb{R}^n$ be a subset, a geometric object. For each real number $d \geq 0$ and every $\epsilon > 0$, define the d-dimensional Hausdorff content $H^d_\epsilon(S)$ by:

$$H^d_\epsilon(S) = \inf\left\{\sum_{i=1}^{\infty}(\mathrm{diam}(U_i))^d : S \subseteq \bigcup_{i=1}^{\infty} U_i,\ \mathrm{diam}(U_i) \leq \epsilon\right\}. \tag{5}$$

where $\mathrm{diam}(U_i)$ is defined as $\mathrm{diam}(U_i) = \sup\{|x-y| : x, y \in U_i\}$ represents the greatest distance between any two points within within subset U_i. U_i are subsets that "cover" the set S. The set S must be fully contained within the union $\bigcup_{i=1}^{\infty} U_i$. Typically, these can be intervals, balls, or cubes. ϵ is a positive real number representing the scale we measure. Smaller ϵ means subsets U_i have smaller sizes. The infimum is then taken over all countable coverings $\{U_i\}$ of S [5] as a measure of the total size of set S.

Then the d-dimensional Hausdorff measure $H^d(S)$ is defined as:

$$H^d(S) = \lim_{\epsilon\to 0} H^d_\epsilon(S). \tag{6}$$

Given the Hausdorff measure, the Hausdorff dimension $\dim_H(S)$ of the set $S \subseteq \mathbb{R}^n$ is defined as:

$$\dim_H(S) = \inf\{d : H^d(S) = 0\} = \sup\{d : H^d(S) = \infty\}. \tag{7}$$

Referring to equation 5, as $\epsilon \to 0$, the value of $\mathrm{diam}(U_i)$ will be very small, which means when d becomes too large, the value of $H^d(S)$ would become 0. Vice versa, when d becomes too small, the value of $H^d(S)$ becomes infinite. Intuitively, this dimension identifies a critical threshold separating two regimes: dimensions for which the measure is infinite and dimensions for which it collapses to zero [5].

The concept of Hausdorff dimension has wide-ranging implications across mathematics and applied fields, particularly in analyzing the complexity and scaling properties of fractal objects, chaotic dynamics, and geometric complexity in nature and data sciences [5, 31].

4 Methodology

In this section, we first introduce how the Hausdorff Dimension of a scanning pattern of Mamba influences Mamba-based low-light image enhancement methods and then move on to the explanation of the superiority of the Hilbert scan and Peano scan in Mamba-based low-light enhancement methods.

4.1 Hausdorff Dimension's Implications for Vision Mamba

In many vision applications, image processing and feature extraction depend on how well a discrete set of samples approximates a continuous image function. In particular, the scanning pattern used to traverse the spatial domain of an image can have significant implications for the performance of downstream tasks. This section investigates the effect of employing a scanning pattern with a high Hausdorff dimension. This intuitively leads to more uniform and space-filling coverage, thus reducing the worst-case approximation error.

Spatial Domain & Sampling Set: We define Ω as the spatial domain of the original input images:

$$\Omega \subset \mathbb{R}^n,$$

where n denotes the number of spatial dimensions. For example, for a two-dimensional image of height H and width W, we have

$$\Omega = \{(x, y) \mid 0 \leq x < H,\ 0 \leq y < W\}.$$

It should be noted that Ω represents the set of all spatial coordinates on which the image is defined. The image itself can be

described as a function.

$$f : \Omega \to \mathbb{R}^c,$$

where c is the number of channels, e.g., $c = 3$ for an RGB image. Note that Ω does not include any feature content of the image; it merely defines the locations over which the image f is defined.

We define P as the sampling set of the spatial domain Ω:

$$P \subset \Omega,$$

where P is a collection of points from which the scanning algorithm extracts data. While a full ordinary raster scan in Mamba might have $P = \Omega$ in a discrete sense, advanced architectures such as Vision Mamba often employ selective or non-uniform scanning strategies. In such cases, P is a proper subset of Ω, and the distribution of points in P, their density, uniformity, and fractal properties, directly influence the quality of the approximation of the continuous function f, which represents the original image.

Function Spaces and Smoothness Conditions: To rigorously analyze how well the continuous image function $f : \Omega \to \mathbb{R}^c$ is approximated by its samples, we place f within appropriate function spaces and assume certain smoothness conditions. These conditions, such as L^2 integrability, Lipschitz continuity, and Hölder continuity, provide a mathematical framework that allows us to derive error bounds on the approximation quality. In this subsection, we describe these spaces and conditions in detail.

A function f belongs to $L^2(\Omega)$ if

$$\|f\|_{L^2(\Omega)} = \left(\int_\Omega |f(x)|^2 \, dx \right)^{1/2} < \infty. \tag{8}$$

This condition ensures that f has finite energy, a natural requirement in signal processing and machine learning.

A function $f : \Omega \to \mathbb{R}$ is said to be Lipschitz continuous if there exists a constant $L_f > 0$ such that

$$|f(x) - f(y)| \le L_f \, \|x - y\| \quad \forall \, x, y \in \Omega. \tag{9}$$

This condition bounds the rate at which f can change between any two points in Ω, ensuring that the function does not exhibit abrupt transitions.

More generally, f is Hölder continuous with exponent $\alpha \in (0, 1]$ if there exists a constant $C_f > 0$ such that

$$|f(x) - f(y)| \le C_f \, \|x - y\|^\alpha \quad \forall \, x, y \in \Omega. \tag{10}$$

When $\alpha = 1$, Hölder continuity is equivalent to Lipschitz continuity. For $\alpha < 1$, the condition allows for more gradual variations in f. These smoothness properties are essential for establishing rigorous error estimates in function approximation.

Worst-Case Approximation Error from Sparse Sampling: This subsection examines the error incurred when approximating a continuous function from discrete samples, which is central to understanding the performance impact of different scanning patterns.

Dispersion & Worst-Case Approximation Error: The *dispersion* of a sampling set P in Ω is defined as

$$\varepsilon(P, \Omega) = \sup_{x \in \Omega} \min_{p \in P} \|x - p\|. \tag{11}$$

For each point $x \in \Omega$, the term $\min_{p \in P} \|x - p\|$ is the distance from x to its nearest sample in P. Taking the supremum over $x \in \Omega$ yields the largest distance, quantifying the worst-case gap in the sampling coverage. A smaller $\varepsilon(P, \Omega)$ indicates that every point in Ω is close to at least one sample, which is critical for accurate approximation.

Let $\mathcal{F}$ denote the class of functions defined on Ω that belong to $L^2(\Omega)$ and satisfy Lipschitz or Hölder continuity. When approximating a function $f \in \mathcal{F}$ using only its values at points in P, interpolation theory provides a pointwise error bound:

$$|f(x) - \hat{f}(x)| \le K \, \varepsilon(P, \Omega)^\alpha, \quad \forall \, x \in \Omega, \tag{12}$$

where $\hat{f}(x)$ is an interpolant of f from the samples P, and K is a constant dependent on the smoothness constants (e.g., C_f or L_f) [4], and α is the Hölder exponent. Integrating this error over Ω in the L^2 norm yields the worst-case approximation error:

$$E(P, \mathcal{F}) = \sup_{f \in \mathcal{F}} \inf_{\hat{f}} \|f - \hat{f}\|_{L^2(\Omega)} \le K' \, \varepsilon(P, \Omega)^\alpha. \tag{13}$$

This relation indicates that minimizing the dispersion $\varepsilon(P, \Omega)$ is key to reducing the overall approximation error in any $f \in \mathcal{F}$ [4]. If the sampling points P are sparse or irregularly distributed, $\varepsilon(P, \Omega)$ will be larger, leading to a higher worst-case error. Hence, having a scanning pattern that minimizes this worst-case error is critical for faithful approximation.

Scanning Patterns & Hausdorff Dimension: In this subsection, we discuss how the fractal properties of a scanning pattern, as measured by its Hausdorff dimension, impact the dispersion and, consequently, the approximation error.

The Hausdorff dimension $\dim_H(P)$ of a set P is defined using the s-dimensional Hausdorff measure:

$$\mathcal{H}^s(P) = \inf \left\{ \sum_{i=1}^{\infty} (\operatorname{diam}(U_i))^s : P \subseteq \bigcup_{i=1}^{\infty} U_i, \ \operatorname{diam}(U_i) \le \delta \right\}. \tag{14}$$

where $\delta \to 0$, and

$$\dim_H(P) = \inf\{s : \mathcal{H}^s(P) = 0\}.$$

A higher Hausdorff dimension implies that P is more space-filling and provides a more uniform coverage of Ω. This concept has been thoroughly discussed in the literature [6, 31] and plays a key role in determining the uniformity of the sampling.

Under reasonable uniformity assumptions, if two sampling sets P_1 and P_2 satisfy

$$\dim_H(P_2) > \dim_H(P_1),$$

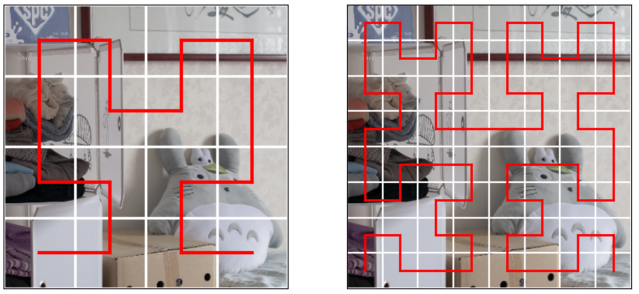
Then generally, according to [4, 6, 31], as scanning patterns with higher Hausdorff Dimensions cover the continuous image function f more uniformly, higher Hausdorff Dimensions lead to lower dispersion value:

$$\varepsilon(P_2, \Omega) \le \varepsilon(P_1, \Omega).$$

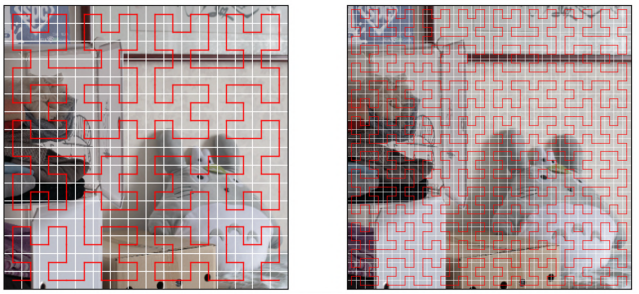
which directly leads to a lower worst-case approximation error.

Consider two scanning methods: ordinary raster scan and Hilbert scan. Raster scan follows a simple row-by-row order. Its continuous parameterization is effectively one-dimensional (Hausdorff dimension of one) even though it visits all discrete points. In contrast, Hilbert scan is a space-filling curve with Hausdorff dimension of two [6, 31], designed to preserve more input features with its fractal nature.

While both methods may visit every pixel in a discrete grid, in scenarios of selective scan in Mamba, the Hilbert curve typically yields a lower dispersion, thereby providing a more robust reconstruction of the underlying image function [6, 31, 32]. The same

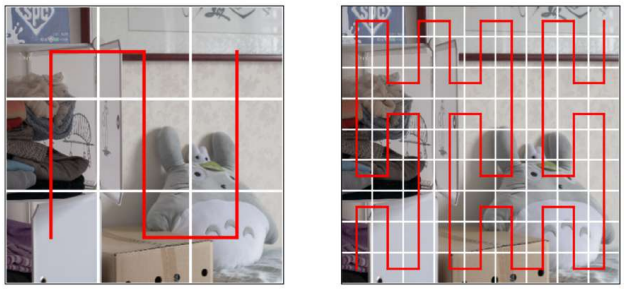


(a) Grid Size $N = 4$ (b) Grid Size $N = 8$

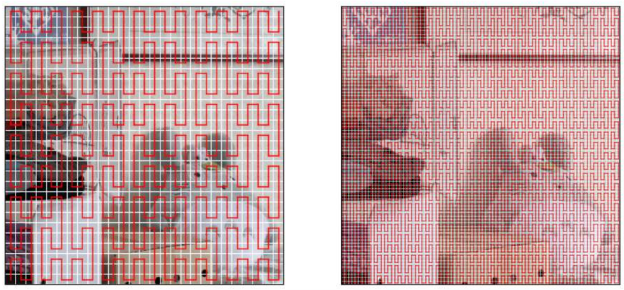


(c) Grid Size $N = 16$ (d) Grid Size $N = 32$

Figure 1: Visualization of self-similar nature of Hilbert curve [13]. Proper rotation and duplication can transform the pattern in (a) into (b), the second-order Hilbert curve. Vice versa for (c) and (d).



(a) Grid Size $N = 3$ (b) Grid Size $N = 9$



(c) Grid Size $N = 27$ (d) Grid Size $N = 81$

Figure 2: Visualization of self-similar nature of Peano curve [33]. Proper rotation and duplication can transform the pattern in (a) into (b), the second-order Peano curve. Vice versa for (c) and (d).

principle can also be applied to evaluate the effectiveness of the Peano scan.

4.2 Hilbert Scan & Peano Scan

In this subsection, we introduce two distinct low-light enhancement methods, HilbertMamba and PeanoMamba, that leverage the superiority of the Hilbert curve [13] and Peano curve [33]. The Hilbert Scan builds on the principles of space-filling curves. It leverages the fractal geometry of the Hilbert curve and Peano curve [13, 31, 33] to address the challenges associated with mapping two-dimensional image data to a one-dimensional sequence. Both methods are based on WaveMamba [60], where we replace its raster-scanning strategy with Hilber-scanning and Peano-scanning strategy, respectively.

HilbertMamba: Based on WaveMamba [60], HilbertMamba leverages the superiority of the Hilbert scan, a scanning strategy inspired by the Hilbert curve [13], which is defined as a continuous mapping:

$$H : [0, 1] \rightarrow [0, 1]^2, \tag{15}$$

that fills the unit square. In our context, the Hilbert Scan reorders the pixels of an $H \times W$ image grid into a one-dimensional sequence, denoted by $I(x, y)$, such that adjacent 2D coordinates are mapped to nearby positions in the sequence. As shown in Figure 1, the Hilbert curve has a self-similar nature, which means the Hilbert curve at higher orders can always be formed by proper rotation and duplication of patterns of lower orders. Formally, for a pixel at spatial location $(x, y) \in \Omega$, the Hilbert Scan assigns an index:

$$I(x, y) = H^{-1}(x, y), \tag{16}$$

thereby preserving the continuity of the spatial domain and ensuring that local neighborhoods remain coherent after transformation.

PeanoMamba: PeanoMamba utilizes the Peano scan, inspired by the Peano curve [33], another space-filling curve, offering a distinct fractal approach compared to the Hilbert curve. The Peano curve similarly provides a continuous mapping:

$$P : [0, 1] \rightarrow [0, 1]^2, \tag{17}$$

and ensures comprehensive coverage of the image domain through a unique traversal pattern. The Peano scanning patterns with different grid sizes are demonstrated in 2. Analogous to HilbertMamba, PeanoMamba transforms a two-dimensional image into a one-dimensional sequence, preserving local spatial coherence through the indexing:

$$I(x, y) = P^{-1}(x, y). \tag{18}$$

Peano scan's distinct fractal structure effectively maintains both fine-grained local details and global dependencies, contributing uniquely to low-light image enhancement.

Superiority of HilbertMamba & PeanoMamba: The superiority of HilbertMamba and PeanoMamba becomes solid when addressing the two challenges in low-light image enhancement: 1) preserving fine-grained local interactions and 2) maintaining global dependency modeling.

Their advantages include: 1) The continuity of the Hilbert and Peano scans ensures that subtle local image features, such as textures and edges, are sampled with high fidelity. This is particularly important in low-light conditions, where minor details significantly impact the perceived visual quality. 2)By providing a more uniform sampling of the spatial domain Ω, Hilbert Scan lowers the

Table 1: Quantitative comparisons of different methods on LOLv1 [39], LOLv2-real [49] and LOLv2-synthetic [49]. The best and second-best results are highlighted in bold and underlined, respectively. Note that we download the pre-trained models from the authors' websites.

Methods	LOLv1			LOLv2-real			LOLv2-synthetic		
	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RetinexNet [39]	16.77	0.419	0.376	16.10	0.401	0.437	17.14	0.761	0.204
MBLLEN [28]	17.86	0.727	0.153	17.78	0.694	0.193	16.10	0.696	0.195
ZeroDCE [10]	14.86	0.559	0.237	18.06	0.574	0.216	17.76	0.816	0.126
MIRNet [52]	24.14	0.840	0.093	20.02	0.820	0.233	15.76	0.735	0.189
EnlightenGAN [17]	17.48	0.651	0.226	18.64	0.675	0.219	16.57	0.775	0.170
KinD++ [56]	21.80	0.834	0.108	22.21	0.843	0.122	19.26	0.806	0.180
LLFlow [37]	19.34	0.840	0.095	24.15	0.864	**0.060**	16.89	0.801	0.166
URetinexNet [42]	20.14	0.823	0.089	19.78	0.843	0.088	18.77	0.824	0.141
SNRNet [46]	24.61	0.842	0.107	21.48	0.849	0.109	24.14	0.908	0.090
Bread [12]	20.62	0.834	0.108	23.69	0.861	0.101	15.97	0.748	0.204
LANet [47]	21.74	0.820	0.101	25.30	0.859	0.095	16.99	0.743	0.241
FourLLIE [35]	20.03	0.820	0.088	22.35	0.847	0.071	24.65	0.910	0.047
Retinexformer [2]	23.50	0.831	0.092	22.79	0.840	0.110	25.39	0.929	0.042
RetinexMamba [1]	23.34	0.833	0.089	22.45	0.844	0.118	25.31	0.9291	0.041
MambaLLIE [40]	21.29	0.824	0.091	22.95	0.847	0.105	24.61	0.9293	0.043
WaveMamba [60]	23.01	0.835	0.1325	29.04	0.908	0.089	24.63	0.923	0.074
PeanoMamba (Ours)	23.99	0.857	0.101	30.71	**0.918**	0.0778	25.32	0.938	0.046
HilbertMamba (Ours)	**24.85**	**0.862**	**0.117**	**31.04**	0.915	0.086	**25.74**	**0.934**	**0.050**

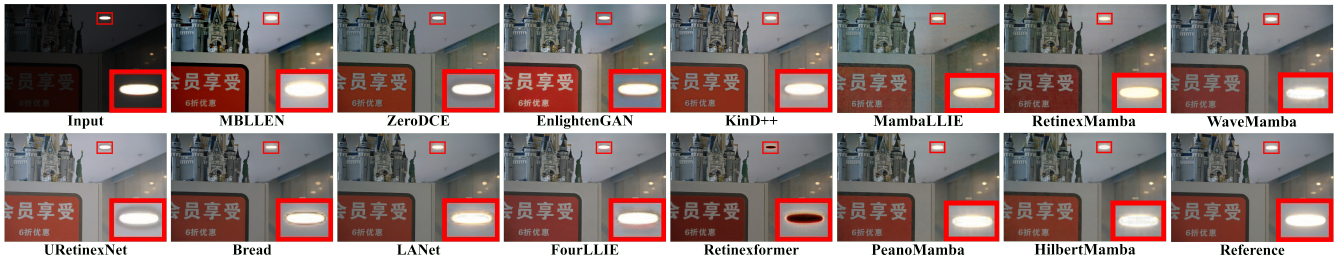


Figure 3: Visual comparisons of the enhanced results by different methods on LOLv2-real.

dispersion $\varepsilon(P, \Omega)$. Under the assumption that the underlying image function f is Hölder continuous with exponent α, the worst-case error bound

$$E(P, \mathcal{F}) \leq K' \, \varepsilon(P, \Omega)^{\alpha},$$

is directly minimized, resulting in higher fidelity in the reconstructed features.

In summary, the Hilbert and Peano scans, by their space-filling and fractal properties, offer mathematically rigorous strategies that significantly enhance the performance of Vision Mamba.

We also introduce a spatial dispersion metric that quantifies discontinuities in scanning patterns by jointly considering the magnitude and frequency of index jumps. For full details and derivations, please refer to the supplementary material.

5 Experiment

5.1 Experiment Settings

Datasets: We trained several Mamba-based low-light image enhancement methods [1, 40, 60] on two datasets: LOLv1 [39] and LOLv2 [49]. LOLv1 contains 500 real-world low/normal-light image pairs, with 485 pairs for training and 15 for testing. LOLv2 is divided into LOLv2-real and LOLv2-synthetic subsets. LOLv2-real contains 689 paired images for training and 100 pairs for testing, collected by adjusting the exposure time and ISO. LOLv2-synthetic contains 900 paired images for training and 100 pairs for testing.

Evaluation metrics: To evaluate the performance of different methods and validate the effectiveness of the proposed method, we adopt full-reference image quality evaluation metrics to evaluate various low-light image enhancement approaches. We employ peak signal-to-noise ratio (PSNR), structural similarity (SSIM) [38], and learned perceptual image patch similarity (LPIPS) [55] as evaluation metrics to assess the model's performance. Higher PSNR and SSIM values and lower LPIPS scores generally indicate more significant similarity between two images.

Methods Comparisons: To demonstrate the superiority of the Mamba-based low-light image enhancement methods reinforced by scanning patterns with Hausdorff dimensions equal to 2, PeanoMamba and HilbertMamba, we compare these prompted baseline methods with a great variety of state-of-the-art methods both quantitatively and qualitatively, including RetinexNet [39], MBLLEN [28], ZeroDCE [10], MIRNet [52], EnlightenGAN [17], KinD++ [56], LLFlow [37], URetinexNet [42], SNRNet [46], Bread [12], LANet

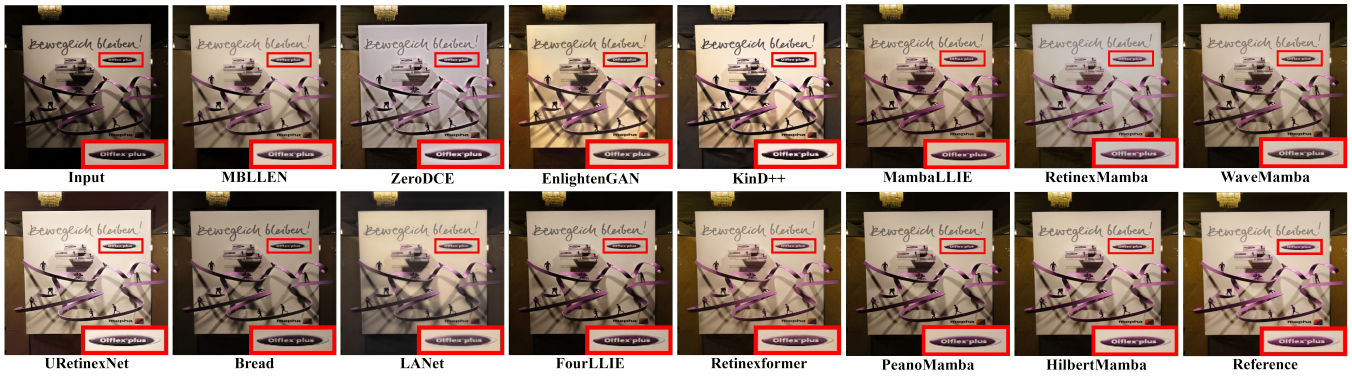


Figure 4: Visual comparisons of the enhanced results by different methods on LOLv2-synthetic.

[47], FourLLIE [35], Retinexformer [2]. For fair comparison, all codes are downloaded from the authors' GitHub repositories, and all comparison results are gained from retraining based on their recommended experimental configurations.

Scanning Paths Comparisons: To demonstrate the superiority of the scanning patterns with Hausdorff dimensions equal to 2 in Mamba-based low-light image enhancement methods, we compare our proposed Hilbert Scan and Peano Scan with other mainstream types of scans: raster scan [59], continuous scan [58], local scan [15], and tree scan [43]. We implement these scans on three different Mamba-based low-light image enhancement methods: RetinexMamba [1], MambaLLIE [40], and WaveMamba [60], respectively, to show the generalizability of the Hilbert Scan and Peano Scan in Mamba-based low-light image enhancement methods. For fair comparison, all codes are downloaded from the authors' websites, and all comparison results are based on their recommended experimental configurations.

Implementation details: We implement our HilbertMamba and PeanoMamba models using PyTorch and train it for 500000 iterations on an NVIDIA GeForce RTX 4090D GPU. The AdamW [27] optimizer ($\beta_1 = 0.9, \beta_2 = 0.99$) is adopted for optimization. initial learning rate 5×10^{-4} gradually reduced to 1×10^{-7} with the cosine annealing. During training, input images are cropped to 256×256 pixels to serve as training samples, with a batch size of 32. For the augmented training data, we use random rotations of 90, 180, 270, random flips, and random cropping to 256×256 size. To constrain the training of HilbertMamba and PeanoMamba, we use the L_1 loss function.

5.2 Comparison with State-of-the-Art Methods

Results on paired datasets: Table 1 displays the quantitative results obtained from comparison methods, where it can be observed that our proposed methods outperform others in most cases, nearly securing the second-best results where they fall short. When compared with the recent transformer-based approaches, SNR [46] and Retinexformer [2], our method achieves 1.11, 5.94, and 3.10 dB improvement on LOLv1, LOLv2-real, and LOLv2-synthetic datasets. Especially on LOLv2-real, the improvement is over 5 dB, as shown in Table 1. Compared with the recent Fourier-based method, FourLLIE [35], our WaveletMamba yields 5.23, 6.38, and 4.12 dB on the three benchmarks. When compared with the recent CNN-based approaches, Bread [12] and LANet [47], our WaveletMamba gains 4.52, 3.43, and 11.78 dB on the three datasets in Table 1. Particularly, our method yields the best visually appealing results in real-world images, as shown in Fig. 3 and Fig. 4, where our method effectively suppresses noise and restores image details, resulting in visuals that closely resemble the original scene. Please zoom in for a better view. The visual results on LOLv1 are provided in the supplementary material. All these results suggest the outstanding effectiveness and efficiency advantage of our HilbertMamba and PeanoMamba.

5.3 Comparison with Different Scanning Paths

Results with Different Scanning Patterns: Table 2 summarizes the performance of RetinexMamba [1], MambaLLIE [40], and WaveMamba [60] on the LOLv1, LOLv2-real, and LOLv2-synthetic datasets using various 2D scanning paths. Our results clearly show that the dimension-2 scans, PS2D (Peano) and HS2D (Hilbert), consistently outperform the dimension-1 scans (SS2D, CS2D, LS2D, TS2D) in terms of PSNR, SSIM, and LPIPS.

For example, RetinexMamba with PS2D yields PSNRs of 23.8 dB (LOLv1), 29.25 dB (LOLv2-real), and 25.54 dB (LOLv2-synthetic). Similarly, MambaLLIE achieves PSNRs of 22.62 dB and 23.36 dB on LOLv1 and 29.76 dB and 29.99 dB on LOLv2-real using PS2D and HS2D, respectively. Notably, WaveMamba with HS2D reaches the highest PSNRs of 24.85 dB on LOLv1 and 31.04 dB on LOLv2-real, surpassing its performance with any dimension-1 scan. The qualitative comparisons implemented on the method WaveMamba [60] are shown in Fig 5, Fig 6 and Fig 7. The qualitative comparisons of the other two methods are provided in the supplementary material.

The dimension-1 scans include the normal raster selective scan (SS2D) [59], continuous scan (CS2D) [58], local scan (LS2D) [15], and tree scan (TS2D) [43]. Visualizations for SS2D, CS2D, and LS2D are provided in the supplementary material, while TS2D, being input-dependent, is detailed in [43].

6 Conclusion

In this paper, we rigorously demonstrate both mathematically and empirically that the performance of Mamba-based low-light image enhancement methods can be significantly improved by adopting scanning patterns with higher Hausdorff dimensions. By introducing HilbertMamba and PeanoMamba, our work pioneers a novel scanning paradigm that transforms the way state space models (SSMs) handle two-dimensional image data. Leveraging the fractal and space-filling properties of Hilbert and Peano curves, our

Table 2: Quantitative comparisons of different scans implemented on Mamba-based low-light image enhancement methods, Retinexmamba[1], MambaLLIE[40], and WaveMamba[60] on LOLv1, LOLv2-real, and LOLv2-synthetic. SS2D[25] represents raster scan, CS2D[58] represents continuous scan, LS2D[15] represents local scan, TS2D[43] represents tree scan, PS2D represents Peano scan and HS2D represents Hilbert scan.

Methods	Dimensions	Scans	LOLv1			LOLv2-real			LOLv2-synthetic		
			PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓	PSNR↑	SSIM↑	LPIPS↓
RetinexMamba	1	SS2D	23.34	**0.8328**	**0.0893**	22.45	0.8444	0.1182	25.31	0.9291	0.0407
	1	CS2D	23.43	0.8212	0.0967	26.75	0.8656	0.0768	25.27	0.9274	0.0428
	1	LS2D	23.23	0.8165	0.0985	25.92	0.8581	0.0786	24.11	0.9116	0.0533
	1	TS2D	22.74	0.8175	0.0991	25.97	0.8621	0.0793	25.34	0.9259	0.0447
	2	*PS2D*	**23.80**	0.8226	0.0926	**29.25**	**0.8765**	**0.0677**	25.54	**0.9320**	**0.0382**
	2	*HS2D*	23.64	0.8147	0.1034	26.91	0.8614	0.0800	**25.59**	0.9294	0.0403
MambaLLIE	1	SS2D	21.29	0.8236	0.0907	22.95	0.8466	0.1047	21.29	0.8236	0.0907
	1	CS2D	21.96	0.8119	0.1011	26.17	0.8636	0.0811	23.59	0.9228	0.0462
	1	LS2D	22.52	0.8197	0.0970	28.85	0.8760	0.0721	23.42	0.9191	0.0489
	1	TS2D	18.79	0.7867	0.1329	21.32	0.8349	0.1080	22.37	0.9038	0.0648
	2	*PS2D*	22.62	0.8275	0.0899	29.76	**0.8874**	**0.0647**	24.74	0.9316	**0.0385**
	2	*HS2D*	**23.36**	**0.8344**	**0.0870**	**29.99**	0.8863	0.0668	**24.90**	**0.9324**	0.0398
WaveMamba	1	SS2D	23.36	0.8544	0.1164	29.04	0.9080	0.0895	24.63	0.9227	0.0737
	1	CS2D	23.40	0.8515	0.1223	28.83	0.9080	0.0933	24.29	0.9212	0.0711
	1	LS2D	23.92	0.8600	0.1123	29.13	0.9048	0.0879	24.49	0.9260	0.0640
	1	TS2D	23.79	0.8568	0.1199	28.39	0.9009	0.0922	24.92	0.9293	0.0615
	2	*PS2D*	23.99	0.8571	**0.1097**	30.71	**0.9175**	**0.0778**	25.32	**0.9382**	**0.0458**
	2	*HS2D*	**24.85**	**0.8617**	0.1173	**31.04**	0.9148	0.0855	**25.74**	0.9369	0.0498

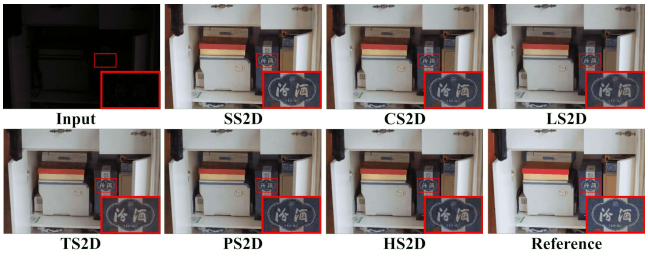


Figure 5: Visual comparisons of results of different scanning paths (based on WaveMamba[60]) on LOLv1.

Figure 6: Visual comparisons of results of different scanning paths (based on WaveMamba[60]) on LOLv2-real.

approach reorders image pixels into a one-dimensional sequence in a manner that preserves delicate local features and enhances the reconstruction of fine-scale details under challenging low-light conditions. Our methodology effectively overcomes the inherent limitations of conventional raster scanning by reducing the approximation errors typical in traditional mapping strategies. This

Figure 7: Visual comparisons of results of different scanning paths (based on WaveMamba[60]) on LOLv2-synthetic.

allows SSMs to capture long-range dependencies while faithfully reconstructing local textures and edges critical for perceptual image quality. Extensive experiments on standard low-light benchmarks validate the theoretical benefits of high Hausdorff dimension scanning patterns, with our methods consistently achieving superior quantitative metrics and visually pleasing enhancement results compared to state-of-the-art approaches. In summary, by bridging the gap between fractal geometry and state space modeling, Hilbert-Mamba and PeanoMamba not only advance the state-of-the-art in low-light image enhancement but also offer a versatile framework that could reshape how we approach sampling and representation in a wide range of vision applications.

References

[1] Jiesong Bai, Yuhao Yin, Qiyuan He, Yuanxian Li, and Xiaofeng Zhang. 2024. Retinexmamba: Retinex-based Mamba for Low-light Image Enhancement. arXiv:2405.03349 [cs.CV] https://arxiv.org/abs/2405.03349

[2] Yuanhao Cai, Hao Bian, Jing Lin, Haoqian Wang, Radu Timofte, and Yulun Zhang. 2023. Retinexformer: One-stage retinex-based transformer for low-light image enhancement. In *ICCV*. 12504–12513.

[3] Heng-Da Cheng and XJ Shi. 2004. A simple and effective histogram equalization approach to image enhancement. *Digit. Signal Process.* 14, 2 (2004), 158–170.

[4] Ronald A. DeVore and George G. Lorentz. 1993. *Constructive Approximation*. Springer.

[5] Gerald A. Edgar. 2007. *Measure, Topology, and Fractal Geometry*. Springer.

[6] Kenneth J. Falconer. 2003. *Fractal Geometry: Mathematical Foundations and Applications*. John Wiley & Sons.

[7] Albert Gu and Tri Dao. 2023. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752* (2023).

[8] Albert Gu, Karan Goel, and Christopher Ré. 2021. Efficiently modeling long sequences with structured state spaces. *arXiv preprint arXiv:2111.00396* (2021).

[9] Zhihao Gu, Fang Li, Faming Fang, and Guixu Zhang. 2019. A novel retinex-based fractional-order variational model for images with severely low light. *IEEE TIP* 29 (2019), 3239–3253.

[10] Chunle Guo, Chongyi Li, Jichang Guo, Chen Change Loy, Junhui Hou, Sam Kwong, and Runmin Cong. 2020. Zero-reference deep curve estimation for low-light image enhancement. In *CVPR*. 1780–1789.

[11] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. 2024. MambaIR: A Simple Baseline for Image Restoration with State-Space Model. *arXiv preprint arXiv:2402.15648* (2024).

[12] Xiaojie Guo and Qiming Hu. 2023. Low-light image enhancement via breaking down the darkness. *IJCV* 131, 1 (2023), 48–66.

[13] David Hilbert. 1891. Ueber die stetige Abbildung einer Linie auf ein Flächenstück. *Math. Ann.* 38 (1891), 459–460.

[14] Shih-Chia Huang, Fan-Chieh Cheng, and Yi-Sheng Chiu. 2012. Efficient contrast enhancement using adaptive gamma correction with weighting distribution. *IEEE TIP* 22, 3 (2012), 1032–1041.

[15] Tao Huang, Xiaohuan Pei, Shan You, Fei Wang, Chen Qian, and Chang Xu. 2024. Localmamba: Visual state space model with windowed selective scan. *arXiv preprint arXiv:2403.09338* (2024).

[16] Hai Jiang, Ao Luo, Haoqiang Fan, Songchen Han, and Shuaicheng Liu. 2023. Low-light image enhancement with wavelet-based diffusion models. *ACM Trans. Graph.* 42, 6 (2023), 1–14.

[17] Yifan Jiang, Xinyu Gong, Ding Liu, Yu Cheng, Chen Fang, Xiaohui Shen, Jianchao Yang, Pan Zhou, and Zhangyang Wang. 2021. EnlightenGAN: Deep light enhancement without paired supervision. *IEEE TIP* 30 (2021), 2340–2349.

[18] Daniel J Jobson, Zia-ur Rahman, and Glenn A Woodell. 1997. A multiscale retinex for bridging the gap between color images and the human observation of scenes. *IEEE TIP* 6, 7 (1997), 965–976.

[19] Chongyi Li, Chunle Guo, Linghao Han, Jun Jiang, Ming-Ming Cheng, Jinwei Gu, and Chen Change Loy. 2021. Low-light image and video enhancement using deep learning: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* 44, 12 (2021), 9396–9416.

[20] Chongyi Li, Jichang Guo, Fatih Porikli, and Yanwei Pang. 2018. LightenNet: A convolutional neural network for weakly illuminated image enhancement. *Pattern Recognit. Lett.* 104 (2018), 15–22.

[21] Dong Liang, Ling Li, Mingqiang Wei, Shuo Yang, Liyan Zhang, Wenhan Yang, Yun Du, and Huiyu Zhou. 2022. Semantically contrastive learning for low-light image enhancement. In *AAAI*, Vol. 36. 1555–1563.

[22] Seokjae Lim and Wonjun Kim. 2020. DSLR: Deep stacked Laplacian restorer for low-light image enhancement. *IEEE TMM* 23 (2020), 4272–4284.

[23] Jiarun Liu, Hao Yang, Hong-Yu Zhou, Yan Xi, Lequan Yu, Yizhou Yu, Yong Liang, Guangming Shi, Shaoting Zhang, Hairong Zheng, et al. 2024. Swin-UMamba: Mamba-based unet with imagenet-based pretraining. *arXiv preprint arXiv:2402.03302* (2024).

[24] Risheng Liu, Long Ma, Jiaao Zhang, Xin Fan, and Zhongxuan Luo. 2021. Retinex-inspired unrolling with cooperative prior architecture search for low-light image enhancement. In *CVPR*. 10561–10570.

[25] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, and Yunfan Liu. 2024. VMamba: Visual state space model. *arXiv preprint arXiv:2401.10166* (2024).

[26] Kin Gwn Lore, Adedotun Akintayo, and Soumik Sarkar. 2017. LLNet: A deep autoencoder approach to natural low-light image enhancement. *PR* 61 (2017), 650–662.

[27] Ilya Loshchilov and Frank Hutter. 2017. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101* (2017).

[28] Feifan Lv, Feng Lu, Jianhua Wu, and Chongsoon Lim. 2018. MBLLEN: Low-light image/video enhancement using CNNs.. In *BMVC*, Vol. 220. Northumbria University, 4.

[29] Jun Ma, Feifei Li, and Bo Wang. 2024. U-Mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722* (2024).

[30] Long Ma, Tengyu Ma, Risheng Liu, Xin Fan, and Zhongxuan Luo. 2022. Toward fast, flexible, and robust low-light image enhancement. In *CVPR*. 5637–5646.

[31] Benoit B. Mandelbrot. 1982. *The Fractal Geometry of Nature*. W.H. Freeman and Company.

[32] Jiří Matoušek. 1999. *Geometric Discrepancy: An Illustrated Guide*. Springer.

[33] Giuseppe Peano. 1890. Sur une courbe, qui remplit toute une aire. *Math. Ann.* 36, 1 (1890), 157–160.

[34] Xiaohuan Pei, Tao Huang, and Chang Xu. 2024. Efficientvmamba: Atrous selective scan for light weight visual mamba. *arXiv preprint arXiv:2403.09977* (2024).

[35] Chenxi Wang, Hongjun Wu, and Zhi Jin. 2023. FourLLIE: Boosting low-light image enhancement by fourier frequency information. In *ACM Int. Conf. Multimedia*. 7459–7469.

[36] Wenjing Wang, Chen Wei, Wenhan Yang, and Jiaying Liu. 2018. GLADNet: Low-light enhancement network with global awareness. In *IEEE Conf. Autom. Face Gesture Recognit. (FG)*. IEEE, 751–755.

[37] Yufei Wang, Renjie Wan, Wenhan Yang, Haoliang Li, Lap-Pui Chau, and Alex Kot. 2022. Low-light image enhancement with normalizing flow. In *AAAI*, Vol. 36. 2604–2612.

[38] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. 2004. Image quality assessment: from error visibility to structural similarity. *IEEE TIP* 13, 4 (2004), 600–612.

[39] Chen Wei, Wenjing Wang, Wenhan Yang, and Jiaying Liu. 2018. Deep retinex decomposition for low-light enhancement. *arXiv preprint arXiv:1808.04560* (2018).

[40] Jiangwei Weng, Zhiqiang Yan, Ying Tai, Jianjun Qian, Jian Yang, and Jun Li. 2024. MambaLLIE: Implicit Retinex-Aware Low Light Enhancement with Global-then-Local State Space. *NeurIPS 2024* (2024).

[41] Jan C Willems. 1986. From time series to linear system—Part I. Finite dimensional linear time invariant systems. *Automatica* 22 (1986), 561–580.

[42] Wenhui Wu, Jian Weng, Pingping Zhang, Xu Wang, Wenhan Yang, and Jianmin Jiang. 2022. URetinex-Net: Retinex-based deep unfolding network for low-light image enhancement. In *CVPR*. 5901–5910.

[43] Yicheng Xiao, Lin Song, Shaoli Huang, Jiangshan Wang, Siyu Song, Yixiao Ge, Xiu Li, and Ying Shan. 2024. GrootVL: Tree Topology is All You Need in State Space Model. *arXiv preprint arXiv:2406.02395* (2024).

[44] Zhaohu Xing, Tian Ye, Yijun Yang, Guang Liu, and Lei Zhu. 2024. SegMamba: Long-range Sequential Modeling Mamba For 3D Medical Image Segmentation. *arXiv preprint arXiv:2401.13560* (2024).

[45] Jingzhao Xu, Mengke Yuan, Dong-Ming Yan, and Tieru Wu. 2022. Illumination guided attentive wavelet network for low-light image enhancement. *IEEE TMM* 25 (2022), 6258–6271.

[46] Xiaogang Xu, Ruixing Wang, Chi-Wing Fu, and Jiaya Jia. 2022. SNR-aware low-light image enhancement. In *CVPR*. 17714–17724.

[47] Kai-Fu Yang, Cheng Cheng, Shi-Xuan Zhao, Hong-Mei Yan, Xian-Shi Zhang, and Yong-Jie Li. 2023. Learning to adapt to light. *IJCV* 131, 4 (2023), 1022–1041.

[48] Wenhan Yang, Shiqi Wang, Yuming Fang, Yue Wang, and Jiaying Liu. 2020. From fidelity to perceptual quality: A semi-supervised approach for low-light image enhancement. In *CVPR*. 3063–3072.

[49] Wenhan Yang, Wenjing Wang, Haofeng Huang, Shiqi Wang, and Jiaying Liu. 2021. Sparse gradient regularized deep retinex network for robust low-light image enhancement. *IEEE TIP* 30 (2021), 2072–2086.

[50] Xunpeng Yi, Han Xu, Hao Zhang, Linfeng Tang, and Jiayi Ma. 2023. Diff-Retinex: Rethinking Low-light Image Enhancement with A Generative Diffusion Model. In *ICCV*. IEEE, 12268–12277.

[51] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. 2022. Restormer: Efficient transformer for high-resolution image restoration. In *CVPR*. 5728–5739.

[52] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2020. Learning enriched features for real image restoration and enhancement. In *ECCV*. Springer, 492–511.

[53] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, Ming-Hsuan Yang, and Ling Shao. 2022. Learning enriched features for fast image restoration and enhancement. *IEEE Trans. Pattern Anal. Mach. Intell.* 45, 2 (2022), 1934–1948.

[54] Kaibing Zhang, Cheng Yuan, Jie Li, Xinbo Gao, and Minqi Li. 2023. Multi-branch and progressive network for low-light image enhancement. *IEEE TIP* 32 (2023), 2295–2308.

[55] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. 2018. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*. 586–595.

[56] Yonghua Zhang, Xiaojie Guo, Jiayi Ma, Wei Liu, and Jiawan Zhang. 2021. Beyond brightening low-light images. *IJCV* 129 (2021), 1013–1037.

[57] Yonghua Zhang, Jiawan Zhang, and Xiaojie Guo. 2019. Kindling the darkness: A practical low-light image enhancer. In *ACM Int. Conf. Multimedia*. 1632–1640.

[58] Weilian Zhou, Sei-Ichiro Kamata, Haipeng Wang, Man-Sing Wong, Huiying, and Hou. 2024. Mamba-in-Mamba: Centralized Mamba-Cross-Scan in Tokenized

Mamba Model for Hyperspectral Image Classification. arXiv:2405.12003 [cs.CV] https://arxiv.org/abs/2405.12003

[59] Lianghui Zhu, Bencheng Liao, Qian Zhang, Xinlong Wang, Wenyu Liu, and Xinggang Wang. 2024. Vision mamba: Efficient visual representation learning with bidirectional state space model. *arXiv preprint arXiv:2401.09417* (2024).

[60] Wenbin Zou, Hongxia Gao, Weipeng Yang, and Tongtong Liu. 2024. Wave-Mamba: Wavelet State Space Model for Ultra-High-Definition Low-Light Image Enhancement. In *ACM Multimedia 2024*. https://openreview.net/forum?id=oQahsz6vWe