GeoMamba: Geometry-Aware State Space Modeling for Image Restoration

Dafeng Zhang, Shizhuo Liu
Samsung R&D Institute China-Beijing (SRC-B)
{dfeng.zhang, shizhuo.liu}@samsung.com

Abstract

State space models (SSMs) provide an efficient route to global image restoration because they model long sequences with linear complexity. Recent restoration architectures further enlarge the effective context by reordering image tokens so that semantically similar pixels become neighbors in the 1D scan. This strategy improves texture aggregation, but it also exposes a mismatch between the geometry of images and the discretization used by SSMs: adjacent tokens in the reordered sequence can be spatially distant, while the recurrent update still treats them as consecutive samples of a continuous process. In addition, existing prompt-based modulation relies on a finite set of discrete prompts, which is poorly matched to the continuous variation of natural textures and offers no explicit compensation for the low-pass tendency of recurrent integration. We propose GeoMamba, a geometry-aware and continuously modulated SSM for image restoration. GeoMamba introduces geometry-adaptive discretization (GAD), which conditions the SSM time-step on the Euclidean distance between consecutive reordered tokens and attenuates history propagation across spatial jumps. It also introduces a continuous dynamic matrix (CDM), which regresses pixel-wise modulation parameters from feature and gradient cues to reduce prompt quantization and preserve high-frequency details. Experiments on lightweight and classical image super-resolution demonstrate consistent improvements over strong Transformer- and Mamba-based baselines, while achieving better or competitive performance in additional denoising and JPEG artifact reduction settings.

1 Introduction

Image restoration aims to recover faithful high-quality images from degraded observations, and its success depends on both accurate global reasoning and precise local detail synthesis. Convolutional networks have long been effective for low-level vision because their local inductive bias matches natural image statistics [8, 25]. However, fixed local kernels make it difficult to exploit long-range self-similarity, especially for repeated structures and textures. Vision Transformers alleviate this limitation through self-attention and have achieved strong restoration performance [15, 4], but their quadratic complexity with respect to the number of tokens makes high-resolution restoration expensive.

State space models (SSMs), represented by Mamba [9], offer an attractive alternative because they propagate global context with linear complexity $\mathcal{O}(N)$. Early Mamba-based restoration models such as MambaIR [11] demonstrate the promise of this direction, but their performance is constrained by rigid scan orders that mix image content according to a fixed spatial traversal. MambaIRv2 [12] makes an important step forward through semantic guided neighboring (SGN), which reorders pixels so that semantically similar textures become adjacent in the 1D sequence. This reordering increases the semantic density of the scan and allows the recurrent state to aggregate visually related evidence over long ranges.

40th Conference on Neural Information Processing Systems (NeurIPS 2026).

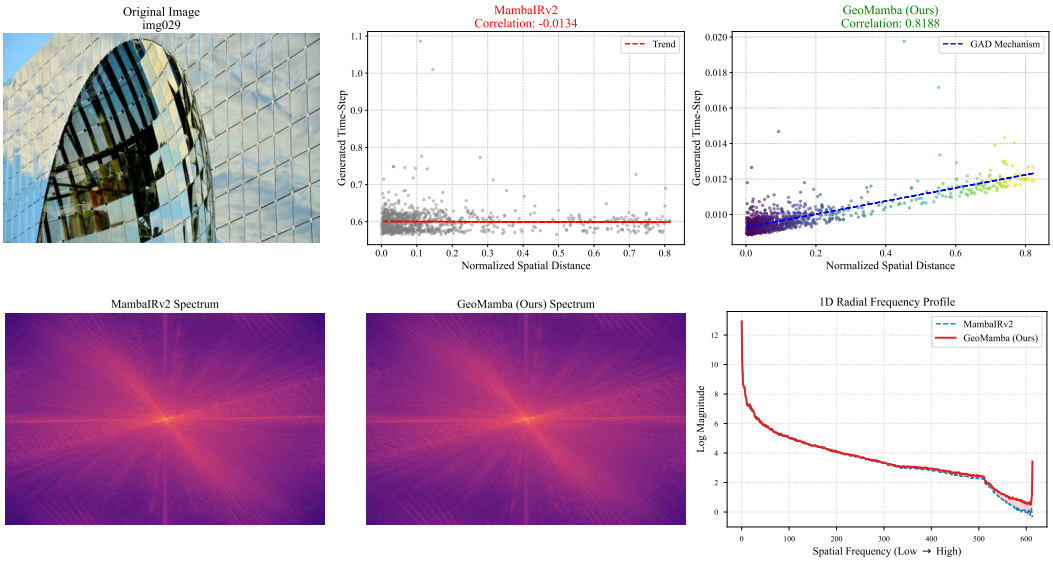

Figure 1: Analysis of geometric correlation and frequency characteristics. **Up:** Given the original image (left), we plot the relationship between the normalized spatial distance of tokens and the generated time-step Δ. Unlike the chaotic distribution in MambaIRv2 (center), our GAD module establishes a strong positive correlation (right), validating its ability to handle discontinuous spatial jumps. **Down:** 2D frequency spectra and the corresponding 1D radial frequency profile. The results demonstrate that our CDM mechanism effectively counteracts the inherent low-pass filtering tendency of SSMs, preserving significantly more high-frequency details.

The benefit of semantic reordering also raises a question that has received limited attention: what physical meaning should an SSM transition have after the image grid has been rearranged? Standard discretization assumes that adjacent steps form a continuous process, yet SGN can place pixels from distant image locations next to each other in the scan. If the time-step Δ is predicted only from token content, a recurrent state can be carried across a large spatial jump as if the two tokens were physically continuous. We refer to this failure mode as a geometry-discretization mismatch. It weakens geometric causality and may propagate structure from a spatially unrelated region into the current reconstruction. A second limitation comes from how existing Mamba restoration models modulate the observation process. MambaIRv2 uses a discrete prompt bank to adapt the output matrix to different semantics [12]. However, image restoration is a continuous regression problem: edges, fine textures, and compression artifacts vary smoothly across pixels and do not naturally align with a finite set of prompts. Such discrete routing can introduce prompt quantization and abrupt feature transitions. Moreover, recurrent integration in SSMs tends to smooth high-frequency signals [20]; a semantic prompt selected from a codebook does not provide an explicit pixel-level mechanism for compensating this attenuation.

We propose GeoMamba, a geometry-aware and continuously modulated state-space framework for image restoration. GeoMamba first introduces geometry-adaptive discretization (GAD) to make the SSM time-step aware of the Euclidean distance between consecutive tokens in the reordered sequence. When two adjacent sequence elements correspond to nearby image locations, the model behaves like a standard content-adaptive SSM. When the transition spans a large spatial distance, GAD increases the effective time-step so that the memory decay term suppresses unrelated history, acting as a soft reset without abandoning the current input. As shown in Figure 1 (Left), GAD establishes a clear positive relation between spatial transition distance and the predicted time-step, while the baseline exhibits a much less structured response.

GeoMamba further introduces a continuous dynamic matrix (CDM) to replace discrete prompt modulation with pixel-wise continuous modulation. CDM predicts modulation parameters from both local features and gradient magnitude, using the gradient as a contrast-aware prior for high-frequency content. This design gives the output matrix a continuous adaptation space and allows the model to strengthen responses around edges and textures where recurrent smoothing is most harmful. Figure 1

(Right) shows that CDM preserves more high-frequency energy than the baseline, which is consistent with its role in detail recovery.

The main contributions of this work are summarized as follows:

- We identify two limitations in reordering-based Mamba restoration models: geometry-discretization mismatch in the state transition and prompt quantization under low-pass recurrent dynamics.
- We propose geometry-adaptive discretization, which conditions the SSM time-step on physical spatial distance and suppresses unrelated history across discontinuous jumps in the reordered sequence.
- We introduce the continuous dynamic matrix, a contrast-aware pixel-wise modulation mechanism that reduces discrete prompt quantization and compensates high-frequency attenuation.
- Experiments show consistent gains on both lightweight and classic super-resolution, with better or competitive transfer results on Gaussian color denoising and JPEG artifact reduction.

2 Related Work

2.1 CNN-based Image Restoration

Since SRCNN [8], Convolutional Neural Networks (CNNs) have dominated low-level vision tasks for nearly a decade. Early approaches deepened architectures to learn discriminative feature representations. VDSR [13] and EDSR [16] utilized residual learning to train deeper networks, significantly improving reconstruction quality. RCAN [25] introduced channel attention to adaptively rescale features, while RDN [26] combined residual connections with dense blocks for enhanced feature reuse. Recently, NAFNet [3] proposed a simplified baseline by replacing nonlinear activations with gating mechanisms, achieving state-of-the-art performance with reduced computational overhead. Despite these advancements, CNNs suffer from finite receptive fields. Even with dilated convolutions or deep stacking, they struggle to effectively capture long-range dependencies, hindering the utilization of non-local self-similarity for restoring complex textures.

2.2 Vision Transformers for Restoration

Vision Transformers (ViTs) marked a paradigm shift toward global modeling. IPT [2] pioneered applying Transformers to low-level vision, though requiring massive pre-training. SwinIR [15] adapted Swin Transformer [18] for image restoration, using shifted window self-attention to model long-range interactions while maintaining translational invariance. HAT [4] refined this strategy with hybrid attention to activate more pixels. Concurrently, Restormer [21] computed self-attention across the channel dimension, achieving efficiency gains for high-resolution images. While ViTs generally outperform CNNs due to content-dependent weights and larger receptive fields, they face a severe computational bottleneck. Self-attention exhibits quadratic complexity $\mathcal{O}(N^2)$ regarding input resolution. This prohibits deployment in resource-constrained scenarios or for processing very high-resolution inputs, creating a pressing need for architectures balancing global modeling with linear efficiency.

2.3 State Space Models in Vision

Structured State Space Models (SSMs), particularly S4 [10] and its selective variant Mamba [9], offer a promising solution to the efficiency-accuracy trade-off. By discretizing continuous-time systems, SSMs model long sequences with linear complexity $\mathcal{O}(N)$. To adapt 1D SSMs to 2D images, Vim [30] proposed bidirectional scanning, while VMamba [17] introduced the Cross-Scan Module (CSM) to traverse spatial features in four directions. For image restoration, MambaIR [11] incorporated local enhancement blocks to complement the Mamba layer, addressing local pixel forgetting. Notably, MambaIRv2 [12] proposed the Semantic Guided Neighboring (SGN) mechanism. Instead of raster scanning, SGN reorders pixels based on semantic similarity, clustering related textures to enhance state propagation. However, current Vision Mamba approaches overlook the geometric implications

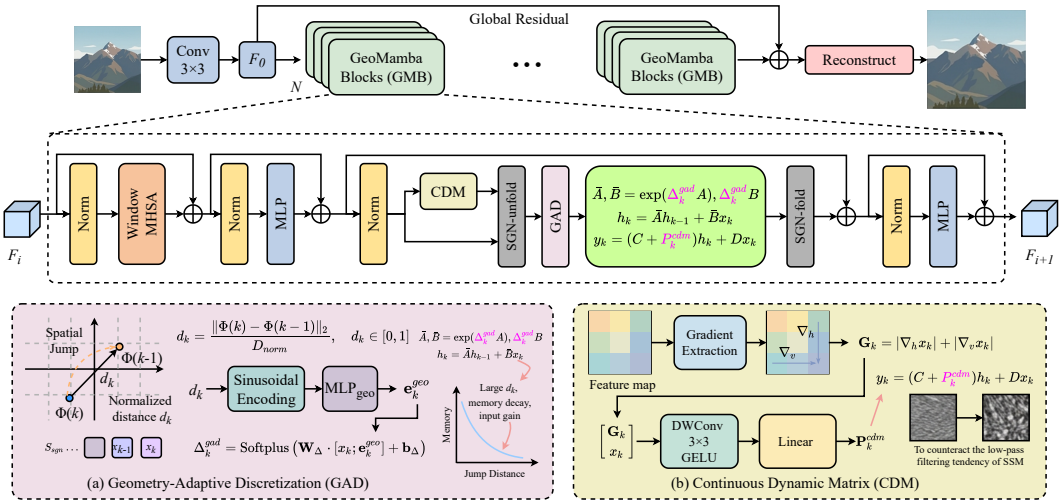

Figure 2: Overall architecture of GeoMamba. Within each GMB, (a) Geometry-Adaptive Discretization (GAD) computes a geometry-aware time-step Δ_k^{gad} from spatial transition distances to handle discontinuous token jumps, while (b) the Continuous Dynamic Matrix (CDM) uses spatial gradients to regress continuous modulation parameters $\mathbf{P}_k^{cdm}$ for high-frequency detail preservation.

of this reordering. Applying standard discretization to spatially discontinuous sequences causes a Geometry-Discretization Mismatch, compromising geometric causality. Furthermore, existing SSMs rely on discrete prompt selection for adaptive modulation. This introduces quantization errors and fails to counteract the inherent low-pass filtering tendency of recurrent state integration, termed the Quantization-Smoothing Dilemma. GeoMamba addresses these fundamental limitations by introducing Geometry-Adaptive Discretization (GAD) and the Continuous Dynamic Matrix (CDM).

3 Methodology

This section presents GeoMamba as a geometry-aware state-space formulation for image restoration. We first describe the overall architecture, then revisit the SSM formulation used by MambaIRv2 [12] to expose two limitations caused by semantic token reordering and discrete prompt modulation. Based on this analysis, we introduce geometry-adaptive discretization (GAD) for spatially consistent state propagation and the continuous dynamic matrix (CDM) for pixel-wise detail-aware modulation.

3.1 Overall Network Architecture

GeoMamba follows the common restoration pipeline of shallow feature extraction, deep feature refinement, and image reconstruction. Given a degraded input image $I_{LQ} \in \mathbb{R}^{H\times W\times C_{in}}$, a 3×3 convolution first maps the image to shallow features F_0. A stack of N GeoMamba blocks (GMBs) then refines these features under a global residual connection, and a reconstruction head produces the restored image I_{HQ}. The key difference from a standard Mamba restoration backbone lies inside each GMB: after semantic guided neighboring (SGN) [12] reorders the 2D features into a 1D sequence, GAD controls how much history should cross each geometric transition, and CDM controls how the hidden state should be read out at each pixel.

3.2 Preliminaries and Problem Formulation

We begin from the discrete SSM update used in MambaIRv2 because it makes the role of token order explicit. Given an input token x_k at step k, the latent state h_k and output y_k are computed as

$$h_k = \overline{\mathbf{A}} h_{k-1} + \overline{\mathbf{B}} x_k, y_k = \mathbf{C} h_k + \mathbf{D} x_k, \tag{1}$$

where $\overline{\mathbf{A}}$ and $\overline{\mathbf{B}}$ are the discretized state transition and input matrices. Under zero-order hold (ZOH) discretization, these matrices are obtained from continuous parameters using a content-dependent

time-step Δ_k:

$$\overline{\mathbf{A}} = \exp(\Delta_k \mathbf{A}), \overline{\mathbf{B}} = (\Delta_k \mathbf{A})^{-1}(\exp(\Delta_k \mathbf{A}) - \mathbf{I})\Delta_k \mathbf{B} \approx \Delta_k \mathbf{B}. \tag{2}$$

MambaIRv2 further introduces the attentive state-space equation (ASE) to adapt the observation matrix. ASE routes an instance-specific prompt $\mathbf{P}_k$ from a discrete prompt pool and adds it to the output matrix:

$$y_k = (\mathbf{C} + \mathbf{P}_k)h_k + \mathbf{D}x_k. \tag{3}$$

MambaIRv2 also uses SGN to reorder 2D image features into a 1D sequence $S_{sgn} = \{x_1, \ldots, x_L\}$ so that similar textures are processed near each other. This design improves semantic aggregation, but it changes the geometric meaning of adjacent sequence steps.

The first limitation is a geometry-discretization mismatch. ZOH discretization assumes that the transition from step $k-1$ to step k corresponds to a continuous process. In a SGN sequence, however, adjacent tokens can be far apart on the image plane. Let $\Phi : \{1, \ldots, L\} \to \mathbb{Z}^2$ map the k-th sequence token to its 2D coordinate $\mathbf{p}_k = (u_k, v_k)$. When $\|\mathbf{p}_k - \mathbf{p}_{k-1}\|_2 \gg 1$, applying a time-step predicted only from content can propagate the previous state into a spatially unrelated region. This breaks geometric causality and may produce non-local structural artifacts.

The second limitation is a quantization-smoothing dilemma in the observation process. A finite prompt pool can only approximate the continuous variation of natural textures, so discrete routing may introduce prompt quantization and abrupt changes in the output matrix. At the same time, the recurrent state update tends to attenuate high-frequency signals. Although semantic prompts adapt the readout to coarse context, they do not explicitly use local contrast or gradients to compensate for texture loss at each pixel.

3.3 Geometry-Adaptive Discretization (GAD)

GAD resolves the geometry-discretization mismatch by making the SSM time-step depend on the physical distance between consecutive reordered tokens. The goal is to preserve normal state propagation within local neighborhoods while suppressing history when the sequence jumps to a distant image location.

We define the spatial transition distance d_k as the normalized Euclidean distance between the current token and its predecessor in the reordered sequence:

$$d_k = \frac{\|\Phi(k) - \Phi(k-1)\|_2}{D_{norm}}, \quad d_k \in [0, 1], \tag{4}$$

where $D_{norm} = \sqrt{H^2 + W^2}$ is the image diagonal. We embed this scalar distance with sinusoidal encoding followed by an MLP, yielding $\mathbf{e}_k^{geo} = \text{MLP}_{geo}(d_k)$. The geometry-adaptive time-step is then predicted from both token content and transition geometry:

$$\Delta_k^{gad} = \text{Softplus}\left(\mathbf{W}_\Delta \cdot [x_k; \mathbf{e}_k^{geo}] + \mathbf{b}_\Delta\right). \tag{5}$$

This formulation gives the recurrence two operating regimes. In the local continuity regime ($d_k \to 0$), Δ_k^{gad} is dominated by content and the model behaves like a standard content-adaptive SSM. In the spatial jump regime ($d_k \gg 0$), the distance embedding encourages a larger time-step, representing a longer physical transition. Because the eigenvalues of the state matrix $\mathbf{A}$ have negative real parts, the transition matrix satisfies

$$\lim_{\Delta \to \infty} \|\exp(\Delta \mathbf{A})\| = 0, \tag{6}$$

$$h_k = \exp(\Delta_k^{gad} \mathbf{A})h_{k-1} + \overline{\mathbf{B}}_k x_k. \tag{7}$$

Therefore, the historical term $\exp(\Delta_k^{gad}\mathbf{A})h_{k-1}$ is attenuated when the sequence crosses a large spatial gap. Under the exact ZOH input matrix in Eq. (2), $\overline{\mathbf{B}}_k$ converges to a bounded steady-state gain as $\exp(\Delta_k^{gad}\mathbf{A}) \to \mathbf{0}$; with the common first-order approximation, we use this result as an interpretation of the intended decay behavior rather than as an additional implementation assumption. GAD thus acts as a soft reset: it weakens dependency on the spatially unrelated predecessor while keeping current-token integration stable.

3.4 Continuous Dynamic Matrix (CDM)

CDM addresses the quantization-smoothing dilemma by replacing discrete prompt selection with continuous pixel-wise modulation. Instead of routing a token to a finite prompt codebook, CDM learns a mapping $\mathcal{F}_\theta : \mathbb{R}^{2C} \to \mathbb{R}^N$ that predicts modulation parameters from local features and their spatial gradients.

We use gradient magnitude as a contrast-aware prior because edges and textures are the regions most vulnerable to recurrent smoothing. For each token feature x_k, we compute horizontal and vertical first-order differences:

$$\mathbf{G}_k = |\nabla_h x_k| + |\nabla_v x_k|. \tag{8}$$

The resulting $\mathbf{G}_k$ acts as a dynamic indicator of local high-frequency content. Large gradient responses mark regions where the readout should preserve sharper structures instead of relying only on the smoothed recurrent state.

Given the concatenated feature-gradient representation $\mathbf{Z}_k = [x_k; \mathbf{G}_k] \in \mathbb{R}^{2C}$, a lightweight hyper-network predicts the continuous modulation vector. It first performs depth-wise contextual interaction and then applies a point-wise projection:

$$\mathbf{P}_k^{cdm} = \mathbf{W}_{proj} \cdot \sigma\left(\text{DWConv}_{3\times3}(\mathbf{Z}_k)\right), \tag{9}$$

where σ denotes GELU. The predicted vector modulates the static output matrix $\mathbf{C}_{static}$ when computing the final observation:

$$y_k = (\mathbf{C}_{static} + \alpha \cdot \mathbf{P}_k^{cdm}) h_k + \mathbf{D} x_k, \tag{10}$$

where α is a learnable scaling factor. This formulation provides two advantages over discrete prompts. First, the modulation space is continuous, so nearby textures can receive nearby readout parameters rather than being forced into separate prompt bins. Second, conditioning on $\mathbf{G}_k$ gives the model an explicit mechanism for detail-aware frequency compensation, allowing sharper readout around edges and textures.

4 Experiments

We organize the experiments to verify effectiveness, generality, and component causality. Specifically, we compare GeoMamba with strong CNN, Transformer, and Mamba baselines on lightweight and classic SR, include representative transfer checks on denoising and JPEG CAR, and ablate GAD and CDM under a matched protocol.

4.1 Experimental Protocol

We follow the training and evaluation protocol of MambaIRv2 [12] for fair comparison. For SR, lightweight models are trained on DIV2K and classic models are trained on DF2K. We evaluate PSNR and SSIM on the Y channel of YCbCr images using Set5, Set14, BSDS100, Urban100, and Manga109 under $2\times$, $3\times$, and $4\times$ upsampling. For Gaussian color denoising and JPEG CAR, we train on DIV2K, Flickr2K, BSDS500, and WED, and evaluate denoising on CBSD68, Kodak24, McMaster, and Urban100, and JPEG CAR on Classic5 and LIVE1. The cross-task results are reported in Appendix B. Training also follows the MambaIRv2 setting. We use horizontal flipping and random rotations of 90°, 180°, and 270° for data augmentation. Training patches are cropped to 64×64 for SR and 128×128 for denoising. The batch size is 32 for SR and 8 for denoising and JPEG CAR. We use Adam with $\beta_1 = 0.9$ and $\beta_2 = 0.999$, train SR models with the ℓ_1 loss, and train denoising/JPEG CAR models with the Charbonnier loss. The initial learning rate is 2×10^{-4} and is halved at the same milestones as MambaIRv2. For SR, the $3\times$ and $4\times$ models are initialized from the pretrained $2\times$ model, and both the learning rate and total training iterations are halved to reduce training time.

4.2 Lightweight Image Super-Resolution

Table 1 evaluates the parameter-constrained setting, where useful gains should not arise from simply increasing model size. GeoMamba achieves the best performance across the reported datasets and scales while using fewer parameters than MambaIRv2-light. At $2\times$ scale, GeoMamba improves over MambaIRv2-light by 0.04 dB on Set5, 0.04 dB on Set14, 0.01 dB on BSDS100, 0.04 dB on

Table 1: Quantitative comparison on *lightweight image super-resolution* with state-of-the-art methods. The best and the second best results are in red and blue.

Method	scale	#param	MACs	Set5		Set14		BSDS100		Urban100		Manga109	
				PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
CARN [1]	2×	1,592K	222.8G	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	38.36	0.9765
LatticeNet [19]	2×	756K	169.5G	38.13	0.9610	33.78	0.9193	32.25	0.9005	32.43	0.9302	-	-
SwinIR-light [15]	2×	910K	244.2G	38.14	0.9611	33.86	0.9206	32.31	0.9012	32.76	0.9340	39.12	0.9783
MambaIR-light [11]	2×	905K	334.2G	38.13	0.9610	33.95	0.9208	32.31	0.9013	32.85	0.9349	39.20	0.9782
ELAN [24]	2×	621K	203.1G	38.17	0.9611	33.94	0.9207	32.30	0.9012	32.76	0.9340	39.11	0.9782
SRFormer-light [29]	2×	853K	236.3G	38.23	0.9613	33.94	0.9209	32.36	0.9019	32.91	0.9353	39.28	0.9785
MambaIRv2-light [12]	2×	774K	286.3G	38.26	0.9615	34.09	0.9221	32.36	0.9019	33.26	0.9378	39.35	0.9785
GeoMamba	2×	757K	290.5G	38.30	0.9617	34.13	0.9221	32.37	0.9020	33.30	0.9380	39.44	0.9786
CARN [1]	3×	1,592K	118.8G	34.29	0.9255	30.29	0.8407	29.06	0.8034	28.06	0.8493	33.50	0.9440
LatticeNet [19]	3×	765K	76.3G	34.53	0.9281	30.39	0.8424	29.15	0.8059	28.33	0.8538	-	-
SwinIR-light [15]	3×	918K	111.2G	34.62	0.9289	30.54	0.8463	29.20	0.8082	28.66	0.8624	33.98	0.9478
MambaIR-light	3×	913K	148.5G	34.63	0.9288	30.54	0.8459	29.23	0.8084	28.70	0.8631	34.12	0.9479
ELAN [24]	3×	629K	90.1G	34.61	0.9288	30.55	0.8463	29.21	0.8081	28.69	0.8624	34.00	0.9478
SRformer-light [29]	3×	861K	105.4G	34.67	0.9296	30.57	0.8469	29.26	0.8099	28.81	0.8655	34.19	0.9489
MambaIRv2-light [12]	3×	781K	126.7G	34.71	0.9298	30.68	0.8483	29.26	0.8098	29.01	0.8689	34.41	0.9497
GeoMamba	3×	764K	128.5G	34.76	0.9300	30.68	0.8485	29.28	0.8102	29.05	0.8700	34.47	0.9501
CARN [1]	4×	1,592K	90.9G	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.47	0.9084
LatticeNet [19]	4×	777K	43.6G	32.30	0.8962	28.68	0.7830	27.62	0.7367	26.25	0.7873	-	-
SwinIR-light [15]	4×	930K	63.6G	32.44	0.8976	28.77	0.7858	27.69	0.7406	26.47	0.7980	30.92	0.9151
MambaIR-light [11]	4×	924K	84.6G	32.42	0.8977	28.74	0.7847	27.68	0.7400	26.52	0.7983	30.94	0.9135
ELAN [24]	4×	640K	54.1G	32.43	0.8975	28.78	0.7858	27.69	0.7406	26.54	0.7982	30.92	0.9150
SRformer-light [29]	4×	873K	62.8G	32.51	0.8988	28.82	0.7872	27.73	0.7422	26.67	0.8032	31.17	0.9165
MambaIRv2-light [12]	4×	790K	75.6G	32.51	0.8992	28.84	0.7878	27.75	0.7426	26.82	0.8079	31.24	0.9182
GeoMamba	4×	773K	76.7G	32.55	0.8998	28.85	0.7882	27.76	0.7432	26.82	0.8084	31.28	0.9187

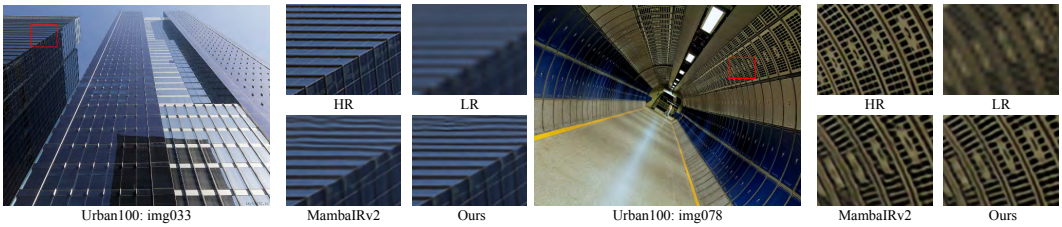

Figure 3: Qualitative comparison on 4× lightweight image SR.

Urban100, and 0.09 dB on Manga109. The Manga109 result is especially informative because its repeated strokes and line structures are sensitive to incorrect long-range memory propagation and high-frequency attenuation. At 3×, GeoMamba maintains positive gains on all datasets, including 0.05 dB on Set5, 0.04 dB on Urban100, and 0.06 dB on Manga109. At 4×, where the inverse problem is more ill-posed, GeoMamba still improves Set5, Set14, BSDS100, and Manga109, and matches the best Urban100 PSNR while increasing Urban100 SSIM from 0.8079 to 0.8084.

The cost profile supports the lightweight design goal. GeoMamba has 757K, 764K, and 773K parameters at 2×, 3×, and 4×, respectively, which is 17K fewer than MambaIRv2-light at each scale. Its MACs are only slightly higher than MambaIRv2-light, increasing from 286.3G to 290.5G at 2×, from 126.7G to 128.5G at 3×, and from 75.6G to 76.7G at 4×. Thus, the consistent accuracy gains are obtained with nearly unchanged computation and a smaller parameter count. This is important for interpreting the result: GAD and CDM improve how the state is propagated and read out, rather than relying on a larger backbone.

4.3 Classic Image Super-Resolution

Table 2 tests whether the proposed formulation remains effective when model capacity is scaled up. GeoMamba consistently ranks first across the classic SR benchmarks and improves over strong CNN, Transformer, and Mamba baselines. Compared with MambaIRv2-B, GeoMamba improves PSNR on all five datasets at every scale. At 2×, it gains 0.03 dB on Set5, 0.06 dB on Set14, 0.01 dB on BSDS100, 0.06 dB on Urban100, and 0.05 dB on Manga109. At 3×, the gains remain positive on every dataset, including 0.05 dB on Urban100. At 4×, the advantage becomes clearer on challenging

Table 2: Quantitative comparison on *classic image super-resolution* with state-of-the-art methods.

Method	scale	#param	Set5		Set14		BSDS100		Urban100		Manga109	
			PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
EDSR [16]	2×	42.6M	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773
RCAN [25]	2×	15.4M	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384	39.44	0.9786
SAN [7]	2×	15.7M	38.31	0.9620	34.07	0.9213	32.42	0.9028	33.10	0.9370	39.32	0.9792
IPT [2]	2×	115M	38.37	-	34.43	-	32.48	-	33.76	-	-	-
SwinIR [15]	2×	11.8M	38.42	0.9623	34.46	0.9250	32.53	0.9041	33.81	0.9427	39.92	0.9797
EDT [14]	2×	11.5M	38.45	0.9624	34.57	0.9258	32.52	0.9041	33.80	0.9425	39.93	0.9800
MambaIR [11]	2×	20.4M	38.57	0.9627	34.67	0.9261	32.58	0.9048	34.15	0.9446	40.28	0.9806
CAT-A [5]	2×	16.5M	38.51	0.9626	34.78	0.9265	32.59	0.9047	34.26	0.9440	40.10	0.9805
DAT [6]	2×	14.8M	38.58	0.9629	34.81	0.9297	32.61	0.9051	34.37	0.9458	40.33	0.9807
HAT [4]	2×	20.6M	38.63	0.9630	34.86	0.9274	32.62	0.9053	34.45	0.9466	40.26	0.9809
MambaIRv2-S [12]	2×	9.6M	38.53	0.9627	34.62	0.9256	32.59	0.9048	34.24	0.9454	40.27	0.9808
MambaIRv2-B [12]	2×	22.9M	38.65	0.9631	34.89	0.9275	32.62	0.9053	34.49	0.9468	40.42	0.9810
GeoMamba	2×	23.1M	38.68	0.9633	34.95	0.9278	32.63	0.9053	34.55	0.9470	40.47	0.9812
EDSR [16]	3×	42.6M	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476
RCAN [25]	3×	15.4M	34.74	0.9299	30.65	0.8482	29.32	0.8111	29.09	0.8702	34.44	0.9499
SAN [7]	3×	15.7M	34.75	0.9300	30.59	0.8476	29.33	0.8112	28.93	0.8671	34.30	0.9494
IPT [2]	3×	115M	34.81	-	30.85	-	29.38	-	29.49	-	-	-
SwinIR [15]	3×	11.8M	34.97	0.9318	30.93	0.8534	29.46	0.8145	29.75	0.8826	35.12	0.9537
EDT [14]	3×	11.5M	34.97	0.9316	30.89	0.8527	29.44	0.8142	29.72	0.8814	35.13	0.9534
MambaIR [11]	3×	20.4M	35.08	0.9323	30.99	0.8536	29.51	0.8157	29.93	0.8841	35.43	0.9546
CAT-A [5]	3×	16.5M	35.06	0.9326	31.04	0.8538	29.52	0.8160	30.12	0.8862	35.80	0.9546
DAT [6]	3×	14.8M	35.16	0.9331	31.11	0.8550	29.55	0.8169	30.18	0.8886	35.59	0.9554
HAT [4]	3×	20.6M	35.07	0.9329	31.08	0.8555	29.54	0.8167	30.23	0.8896	35.53	0.9552
MambaIRv2-S [12]	3×	9.8M	35.09	0.9326	31.07	0.8547	29.51	0.8157	30.08	0.8871	35.44	0.9549
MambaIRv2-B [12]	3×	23.1M	35.18	0.9334	31.12	0.8557	29.55	0.8169	30.28	0.8905	35.61	0.9556
GeoMamba	3×	23.3M	35.21	0.9337	31.14	0.8558	29.56	0.8170	30.33	0.8913	35.62	0.9557
EDSR [16]	4×	43.0M	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
RCAN [25]	4×	15.6M	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	31.22	0.9173
SAN [7]	4×	15.7M	32.64	0.9003	28.92	0.7888	27.78	0.7436	26.79	0.8068	31.18	0.9169
IPT [2]	4×	116M	32.64	-	29.01	-	27.82	-	27.26	-	-	-
SwinIR [15]	4×	11.9M	32.92	0.9044	29.09	0.7950	27.92	0.7489	27.45	0.8254	32.03	0.9260
EDT [14]	4×	11.6M	32.82	0.9031	29.09	0.7939	27.91	0.7483	27.46	0.8246	32.05	0.9254
MambaIR [11]	4×	20.4M	33.03	0.9046	29.20	0.7961	27.98	0.7503	27.68	0.8287	32.32	0.9272
CAT-A [5]	4×	16.6M	33.08	0.9052	29.18	0.7960	27.99	0.7510	27.89	0.8339	32.39	0.9285
DAT [6]	4×	14.8M	33.08	0.9055	29.23	0.7973	28.00	0.7515	27.87	0.8343	32.51	0.9291
HAT [4]	4×	20.8M	33.04	0.9056	29.23	0.7973	28.00	0.7517	27.97	0.8368	32.48	0.9292
MambaIRv2-S [12]	4×	9.8M	32.99	0.9037	29.23	0.7965	27.97	0.7502	27.73	0.8307	32.33	0.9276
MambaIRv2-B [12]	4×	23.1M	33.14	0.9057	29.23	0.7975	28.00	0.7511	27.89	0.8344	32.57	0.9295
GeoMamba	4×	23.2M	33.19	0.9066	29.29	0.7984	28.03	0.7520	28.02	0.8376	32.61	0.9301

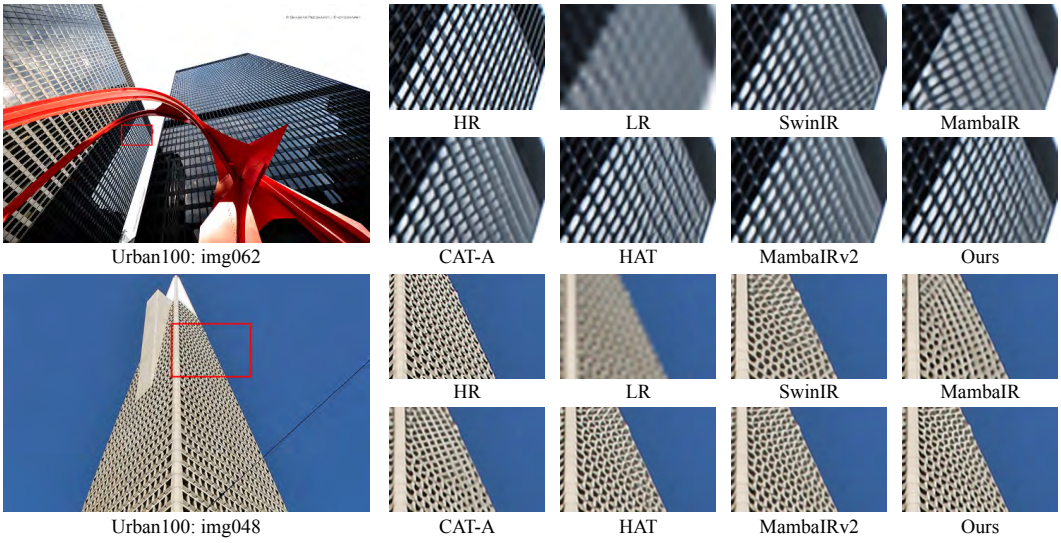

Figure 4: Qualitative comparison of our GeoMamba with different methods on 4× classic image SR.

structured scenes: GeoMamba reaches 28.02 dB on Urban100, outperforming MambaIRv2-B by 0.13 dB and HAT by 0.05 dB.

The distribution of improvements matches the motivation of GeoMamba. Set5 and BSDS100 include many smoother natural regions, where strong methods already cluster tightly; GeoMamba still gives

Table 3: Ablation study on the proposed components of GeoMamba. GAD denotes the geometry-adaptive discretization, and CDM denotes the continuous dynamic matrix.

GAD	CDM	Set5		Set14		BSDS100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
✗	✗	38.23	0.9614	33.97	0.9214	32.34	0.9017	33.07	0.9365	39.35	0.9785
✓	✗	38.26	0.9616	33.98	0.9218	32.35	0.9019	33.15	0.9371	39.41	0.9787
✗	✓	38.27	0.9616	34.05	0.9220	32.36	0.9019	33.15	0.9373	39.37	0.9785
✓	✓	38.30	0.9617	34.13	0.9221	32.37	0.9020	33.30	0.9380	39.44	0.9786

small but stable gains in these saturated settings. Larger gains appear on Urban100 and Manga109, which contain repeated windows, grid patterns, text strokes, and cartoon contours. These cases are exactly where semantic reordering is valuable but risky: semantically similar patches may be far apart in the image, and a recurrent state can carry misleading history if the transition ignores geometry. The consistent gains in both lightweight and classic SR show that geometry-aware discretization and continuous modulation remain beneficial across different capacity regimes.

4.4 Qualitative Results and Cross-Task Generalization

Figure 4 and the additional visual comparisons in Appendix E complement the distortion metrics with local evidence. GeoMamba reconstructs sharper boundaries and cleaner repetitive structures than the compared baselines, particularly around line patterns and high-contrast edges. This qualitative behavior is consistent with the two mechanisms introduced in the method: GAD reduces the influence of spatially unrelated predecessors after semantic reordering, and CDM strengthens detail-aware readout around high-gradient regions. Appendix B reports Gaussian color denoising and JPEG CAR results. These tasks use different degradations from SR, so they serve as representative transfer checks for the modeling idea rather than the primary source of the paper's empirical claim. GeoMamba improves over MambaIRv2 on the reported denoising setting and remains competitive on JPEG CAR at quality factor $q = 40$. The margins are modest on saturated metrics, but they are consistent with the underlying motivation: geometry-aware propagation controls non-local memory, while continuous modulation improves detail-sensitive reconstruction.

4.5 Component-wise Ablation

Table 3 isolates the contributions of GAD and CDM on lightweight $2\times$ SR. The baseline without either module obtains 38.23 dB on Set5, 33.97 dB on Set14, 32.34 dB on BSDS100, 33.07 dB on Urban100, and 39.35 dB on Manga109. Adding GAD alone improves all five datasets, with gains of 0.03 dB, 0.01 dB, 0.01 dB, 0.08 dB, and 0.06 dB, respectively. The larger gains on Urban100 and Manga109 are consistent with the role of GAD: these datasets contain repeated structures where semantic reordering can place visually similar but spatially distant tokens next to each other, making geometry-aware state attenuation more important.

Adding CDM alone also improves every dataset, but with a different gain pattern. Compared with the baseline, CDM improves Set5, Set14, BSDS100, Urban100, and Manga109 by 0.04 dB, 0.08 dB, 0.02 dB, 0.08 dB, and 0.02 dB, respectively. This behavior matches the purpose of CDM: it strengthens the readout around local contrast and high-frequency regions rather than primarily controlling long-range state propagation. We provide a more detailed CDM design ablation in Appendix A, where we compare discrete prompting, gradient-free modulation, and the full gradient-aware continuous modulation.

Combining GAD and CDM gives the best overall results, improving the baseline by 0.07 dB on Set5, 0.16 dB on Set14, 0.03 dB on BSDS100, 0.23 dB on Urban100, and 0.09 dB on Manga109. The full model also outperforms either single-module variant, especially on Set14 and Urban100. This indicates that the two modules are complementary: GAD decides how much history should be propagated across reordered tokens, while CDM decides how the resulting state should be read out to preserve local detail.

Table 4: Ablation study on the distance metric within GAD for GeoMamba. GAD denotes the geometry-adaptive discretization.

Distance	Set5		Set14		BSDS100		Urban100		Manga109	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Euclidean	38.26	0.9616	33.98	0.9218	32.35	0.9019	33.15	0.9371	39.41	0.9787
Manhattan	38.21	0.9613	33.96	0.9214	32.35	0.9019	33.08	0.9369	39.27	0.9779
Rank Jump	38.26	0.9615	33.97	0.9216	32.34	0.9018	33.13	0.9369	39.38	0.9786

4.6 Distance Metric in Geometry-Adaptive Discretization

Table 4 studies how the geometry signal in GAD should be measured. We compare Euclidean distance, Manhattan distance, and a rank-jump distance under the same lightweight $2\times$ SR setting. Euclidean distance gives the best overall results, reaching 38.26 dB on Set5, 33.98 dB on Set14, 32.35 dB on BSDS100, 33.15 dB on Urban100, and 39.41 dB on Manga109. Manhattan distance performs worse on most datasets, with the largest PSNR drop on Manga109, while rank-jump distance is close to Euclidean on Set5 but lower on Set14, BSDS100, Urban100, and Manga109. These results clarify the type of geometry that matters for the proposed discretization. GAD does not need a complicated proxy for semantic disorder; it benefits from a direct physical measure of how far the recurrent state moves on the image plane after reordering. Euclidean distance is isotropic, aligns with the continuous-space interpretation of the SSM time-step, and naturally treats diagonal and horizontal/vertical transitions by their actual spatial length. We therefore use Euclidean distance as the default GAD metric in all main experiments.

5 Conclusion

This paper presents GeoMamba, a geometry-aware and continuously modulated state-space architecture for image restoration. Our analysis shows that semantic token reordering, while useful for aggregating similar textures, can break the geometric meaning of SSM transitions and can also expose the limits of discrete prompt modulation. GeoMamba addresses these issues with two complementary components: Geometry-Adaptive Discretization (GAD), which conditions the time-step on spatial transition distance to suppress unrelated history across discontinuous jumps, and the Continuous Dynamic Matrix (CDM), which uses contrast-aware pixel-wise modulation to reduce prompt quantization and preserve high-frequency details. Experiments on lightweight and classical image super-resolution demonstrate consistent gains over strong Transformer- and Mamba-based baselines, while additional results on denoising and JPEG artifact reduction show better or competitive transfer performance. The ablation results further show that GAD and CDM each contribute measurable gains, and their combination gives the best restoration accuracy. These results suggest that future state-space restoration models should treat token order not only as an efficiency choice, but also as a geometric constraint that affects how recurrent memory should be propagated.

References

[1] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *ECCV*, pages 252–268, 2018.

[2] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. In *CVPR*, pages 12299–12310, 2021.

[3] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *ECCV*, pages 17–33. Springer, 2022.

[4] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *CVPR*, pages 22367–22377, 2023.

[5] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xin Yuan, et al. Cross aggregation transformer for image restoration. *NeurIPS*, 35:25478–25490, 2022.

[6] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *ICCV*, pages 12312–12321, 2023.

[7] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, pages 11065–11074, 2019.

[8] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, pages 184–199. Springer, 2014.

[9] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. In *First conference on language modeling*, 2024.

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

[11] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. Mambair: A simple baseline for image restoration with state-space model. In *European conference on computer vision*, pages 222–241. Springer, 2024.

[12] Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28124–28133, 2025.

[13] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, pages 1646–1654, 2016.

[14] Wenbo Li, Xin Lu, Shengju Qian, Jiangbo Lu, Xiangyu Zhang, and Jiaya Jia. On efficient transformer-based image pre-training for low-level vision. *arXiv preprint arXiv:2112.10175*, 2021.

[15] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *ICCVW*, pages 1833–1844, 2021.

[16] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, pages 136–144, 2017.

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

[18] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin Transformer: Hierarchical vision transformer using shifted windows. In *ICCV*, pages 10012–10022, 2021.

[19] Xiaotong Luo, Yuan Xie, Yulun Zhang, Yanyun Qu, Cuihua Li, and Yun Fu. Latticenet: Towards lightweight image super-resolution with lattice block. In *ECCV*, pages 272–289. Springer, 2020.

[20] Annan Yu, Dongwei Lyu, Soon Hoe Lim, Michael W. Mahoney, and N. Benjamin Erichson. Tuning frequency bias of state space models. In *The Thirteenth International Conference on Learning Representations*, 2025. URL `https://openreview.net/forum?id=wkHcXDv7cv`.

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

[22] Jiale Zhang, Yulun Zhang, Jinjin Gu, Jiahua Dong, Linghe Kong, and Xiaokang Yang. Xformer: Hybrid x-shaped transformer for image denoising. *arXiv preprint arXiv:2303.06440*, 2023.

[23] Kai Zhang, Yawei Li, Wangmeng Zuo, Lei Zhang, Luc Van Gool, and Radu Timofte. Plug-and-play image restoration with deep denoiser prior. *IEEE TPAMI*, 44(10):6360–6376, 2021.

[24] Xindong Zhang, Hui Zeng, Shi Guo, and Lei Zhang. Efficient long-range attention network for image super-resolution. In *ECCV*, pages 649–667. Springer, 2022.

[25] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pages 286–301, 2018.

[26] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR*, pages 2472–2481, 2018.

[27] Yulun Zhang, Kunpeng Li, Kai Li, Bineng Zhong, and Yun Fu. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082*, 2019.

[28] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image restoration. *IEEE TPAMI*, 43(7):2480–2495, 2020.

[29] Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. SRFormer: Permuted self-attention for single image super-resolution. *arXiv preprint arXiv:2303.09735*, 2023.

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

This appendix provides additional evidence for the main experimental claims. We first isolate the design choices inside CDM, then evaluate GeoMamba on denoising and JPEG compression artifact reduction with the U-shaped restoration architecture, and finally analyze the diagnostic visualization of geometry-aware state propagation.

A Detailed CDM Ablation

Table 5 provides a focused ablation on the design of CDM. To isolate the readout mechanism, all variants are evaluated under lightweight $2\times$ SR with GAD disabled. The baseline removes both GAD and CDM, Discrete Prompt uses a finite prompt-style readout, CDM w/o Gradient keeps continuous modulation but removes the gradient cue, and Full CDM uses the proposed gradient-aware continuous modulation. The comparison separates discrete prompting, continuous modulation, and the additional gradient-aware cue. Discrete Prompt gives limited gains over the baseline on Set5, Set14, and Urban100, but slightly decreases Manga109, suggesting that a finite prompt-style readout provides only coarse adaptation for high-frequency line structures. CDM w/o Gradient performs better than Discrete Prompt on all five datasets, improving the baseline by 0.03 dB on Set5, 0.05 dB on Set14, 0.01 dB on BSDS100, and 0.05 dB on Urban100. This indicates that token-wise continuous readout is useful because it adapts the observation process to local image content.

Full CDM further improves over CDM w/o Gradient on every dataset, reaching 38.27 dB on Set5, 34.05 dB on Set14, 32.36 dB on BSDS100, 33.15 dB on Urban100, and 39.37 dB on Manga109. Compared with the baseline, the full design brings gains of 0.04 dB, 0.08 dB, 0.02 dB, 0.08 dB, and 0.02 dB, respectively. The additional improvements from the gradient cue are most visible on Set14 and Urban100, supporting the motivation that CDM should condition the readout on local contrast and high-frequency structure rather than only on semantic features.

Table 5: Ablation study on the CDM design under lightweight $2\times$ SR. All variants disable GAD to isolate the readout modulation mechanism.

Variant	Set5		Set14		BSDS100		Urban100		Manga109	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Baseline	38.23	0.9614	33.97	0.9214	32.34	0.9017	33.07	0.9365	39.35	0.9785
Discrete Prompt	38.25	0.9614	34.00	0.9216	32.34	0.9017	33.10	0.9368	39.34	0.9783
CDM w/o Gradient	38.26	0.9615	34.02	0.9218	32.35	0.9018	33.12	0.9370	39.35	0.9784
Full CDM	38.27	0.9616	34.05	0.9220	32.36	0.9019	33.15	0.9373	39.37	0.9785

B Additional Restoration Tasks

We further evaluate GeoMamba on Gaussian color image denoising and JPEG compression artifact reduction. For these two restoration tasks, we use a U-shaped encoder-decoder architecture rather than the straight-through SR backbone. This design is appropriate because noise, blocking artifacts, and ringing appear at multiple spatial scales; the encoder-decoder structure provides multi-scale features for artifact removal, while the reduced intermediate resolutions lower the cost of state-space processing. To verify that the gain is not only caused by the architectural change, we also report a GeoMamba-Linear variant that keeps a straight-through backbone similar to MambaIRv2.

Table 6: Quantitative PSNR comparison on *Gaussian color image denoising* at noise level $\sigma = 15$.

Method	CBSD68	Kodak24	McMaster	Urban100
DRUNet [23]	34.30	35.31	35.40	34.81
SwinIR [15]	34.42	35.34	35.61	35.13
Restormer [21]	34.40	35.35	35.61	35.13
Xformer [22]	34.43	35.39	35.68	35.29
MambaIR [11]	34.48	35.42	35.70	35.37
MambaIRv2 [12]	34.48	35.43	35.73	35.42
GeoMamba-Linear	34.49	35.44	35.74	35.47
GeoMamba-UShape	34.49	35.44	35.74	35.48

Table 7: Quantitative PSNR/SSIM comparison on *JPEG compression artifact reduction* under quality factor $q = 40$.

Method	Classic5		LIVE1	
RNAN [27]	34.27	0.9061	34.47	0.9299
RDN [28]	34.27	0.9061	34.51	0.9302
DRUNet [23]	34.41	0.9075	34.58	0.9312
SwinIR [15]	34.52	0.9082	34.67	0.9317
MambaIR [11]	34.53	0.9084	34.70	0.9320
MambaIRv2 [12]	34.64	0.9093	34.73	0.9323
GeoMamba-Linear	34.64	0.9094	34.73	0.9323
GeoMamba-UShape	34.64	0.9095	34.73	0.9323

Table 6 reports Gaussian color image denoising results at $\sigma = 15$. GeoMamba-UShape achieves the best PSNR on all four datasets, improving MambaIRv2 by 0.01 dB on CBSD68, Kodak24,

and McMaster, and by 0.06 dB on Urban100. GeoMamba-Linear remains essentially tied with GeoMamba-UShape on CBSD68, Kodak24, and McMaster, and is only 0.01 dB lower on Urban100. This result indicates that the proposed GAD and CDM modules remain effective under additive noise, while the U-shaped architecture provides an additional benefit by aggregating multi-scale degradation cues. The larger improvement on Urban100 is consistent with the SR results because structured scenes with repeated edges and dense patterns benefit more from geometry-aware propagation and detail-aware modulation.

Table 7 evaluates JPEG CAR under quality factor $q = 40$. GeoMamba-UShape obtains 34.64 dB/0.9095 on Classic5 and 34.73 dB/0.9323 on LIVE1. Compared with MambaIRv2, it matches PSNR on Classic5 while improving SSIM by 0.0002, and it matches both PSNR and SSIM on LIVE1. GeoMamba-Linear is also highly competitive, reaching 34.64 dB/0.9094 on Classic5 and 34.73 dB/0.9323 on LIVE1. These results show that the U-shaped design provides a small but consistent advantage on JPEG CAR, while the proposed formulation remains competitive with the strongest Mamba baseline. The small margins are expected because JPEG CAR at a fixed quality factor is already highly saturated among recent restoration models.

Figure 6 reports inference latency over varying input resolutions. GeoMamba remains comparable to or faster than MambaIRv2 across the measured resolutions, showing that the additional geometry and modulation cues do not remove the linear-complexity advantage of the state-space backbone. For denoising and JPEG CAR, the U-shaped architecture further improves practical efficiency by processing part of the state-space computation at reduced spatial resolutions.

C Visualization of Geometry-Aware State Propagation

Figure 5 visualizes how GAD changes state propagation after semantic reordering. In a reordered sequence, adjacent tokens can be visually similar but spatially far apart. A content-only time-step may therefore propagate memory across a physically discontinuous transition, which can introduce structure from an unrelated location. GeoMamba instead predicts a time-step that increases with the spatial distance between consecutive reordered tokens. This behavior supports the intended role of GAD: local transitions preserve normal state propagation, while large spatial jumps receive stronger memory attenuation.

The visualization also explains the dataset-level trend observed in the main experiments. GeoMamba gives larger gains on Urban100 and Manga109, where repeated windows, line patterns, and cartoon strokes create many semantically similar but spatially separated tokens. In such cases, semantic reordering is useful for gathering related evidence, but it also increases the risk of non-local memory leakage. GAD reduces this risk by making the recurrent transition aware of image geometry, which yields cleaner state representations around repeated structures.

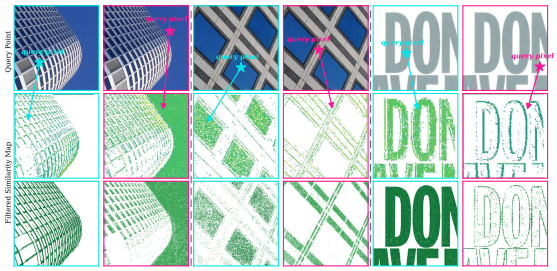

Figure 5: Visualization of the attentive state space of MambaIRv2 and the Geometry-Adaptive Discretization (GAD) in GeoMamba.

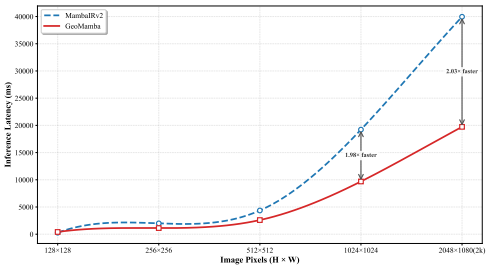

Figure 6: Inference latency across varying input resolutions.

D Limitations and Future Work

While GeoMamba demonstrates clear gains over state-of-the-art baselines, several limitations of the current work suggest directions for future investigation. GeoMamba is designed for single-image restoration and models 2D feature sequences using SGN reordering. Extending geometry-aware discretization to video restoration introduces an additional temporal dimension where frame-to-frame motion interacts with spatial token transitions. The current GAD formulation only considers

within-frame spatial distance; future work could incorporate optical flow or motion coherence to condition the time-step on both spatial and temporal discontinuities. This is particularly relevant for video super-resolution and frame interpolation, where maintaining temporal consistency across geometry-aware state propagation is essential.

E Additional Qualitative Results

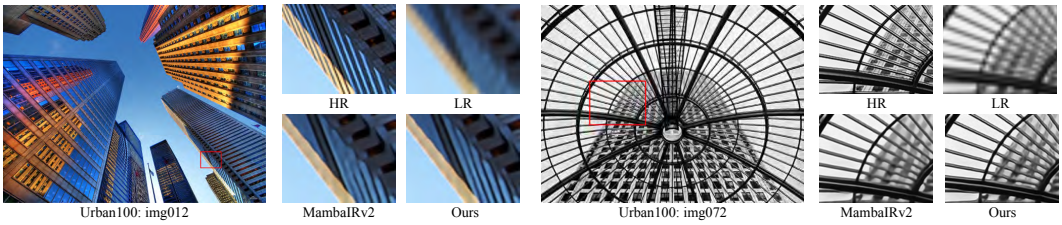

Figure 7: Additional qualitative comparison on lightweight image super-resolution.

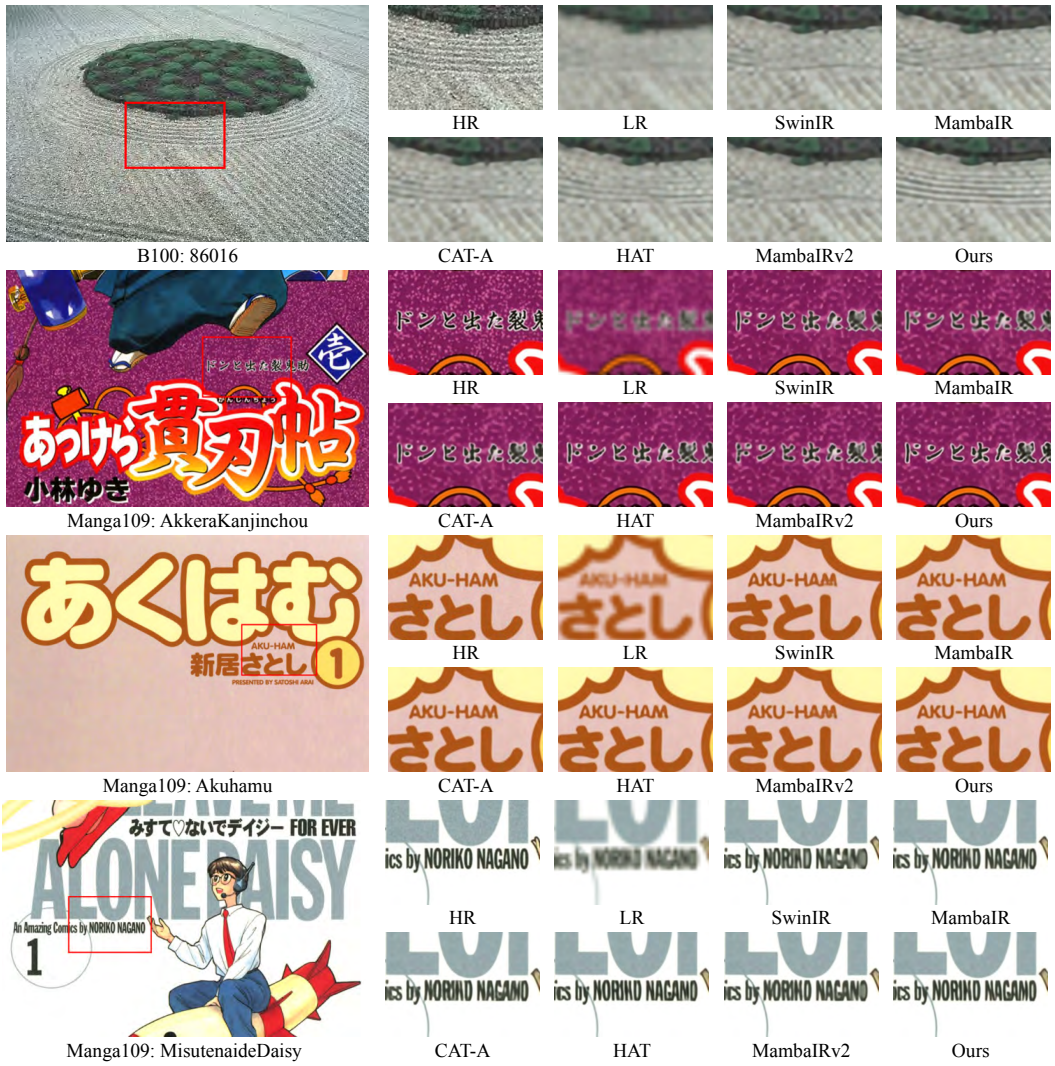

Figure 8: Additional qualitative comparison on classic image super-resolution.

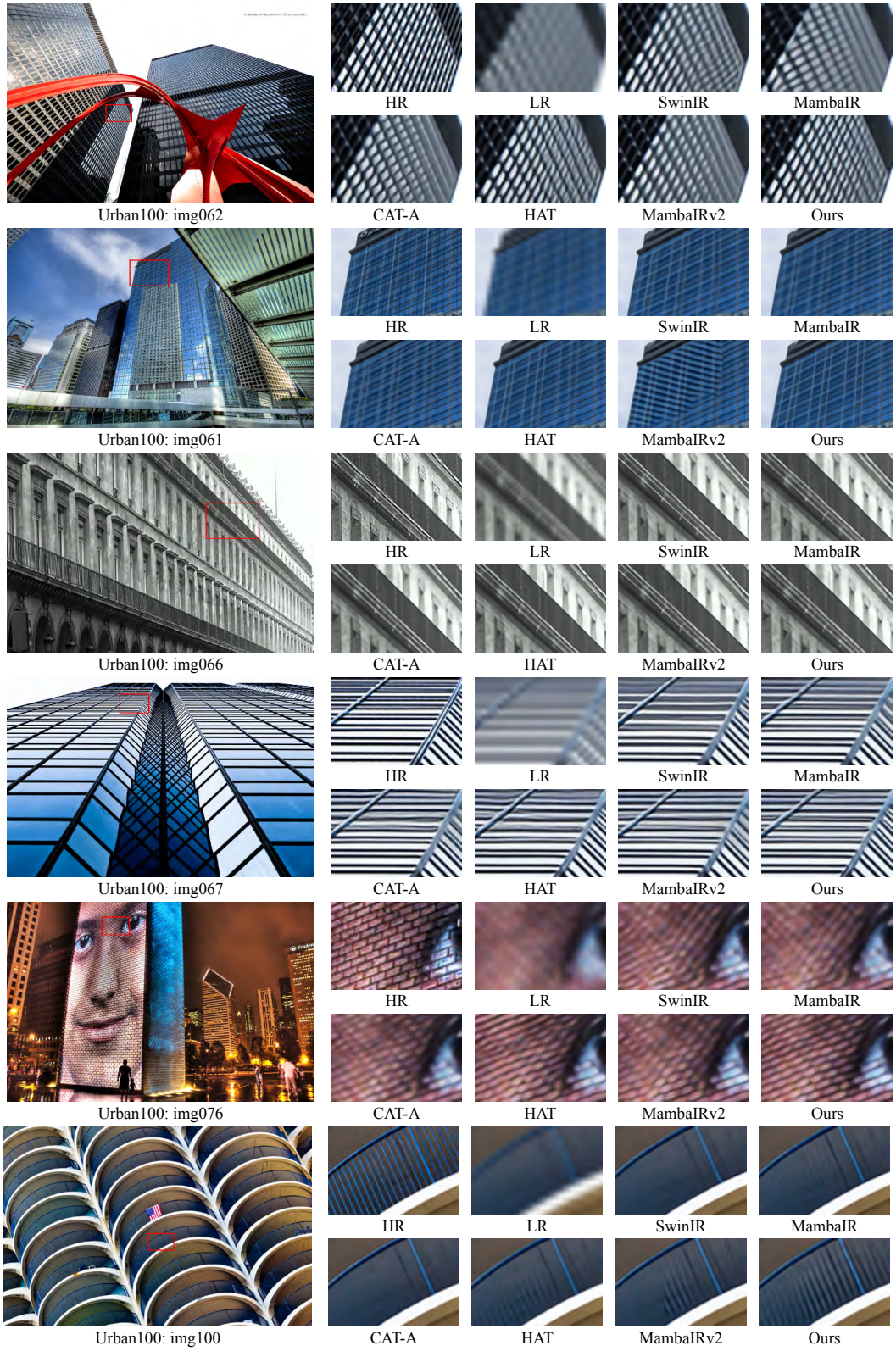

Figure 9: Additional qualitative comparison on classic image super-resolution.

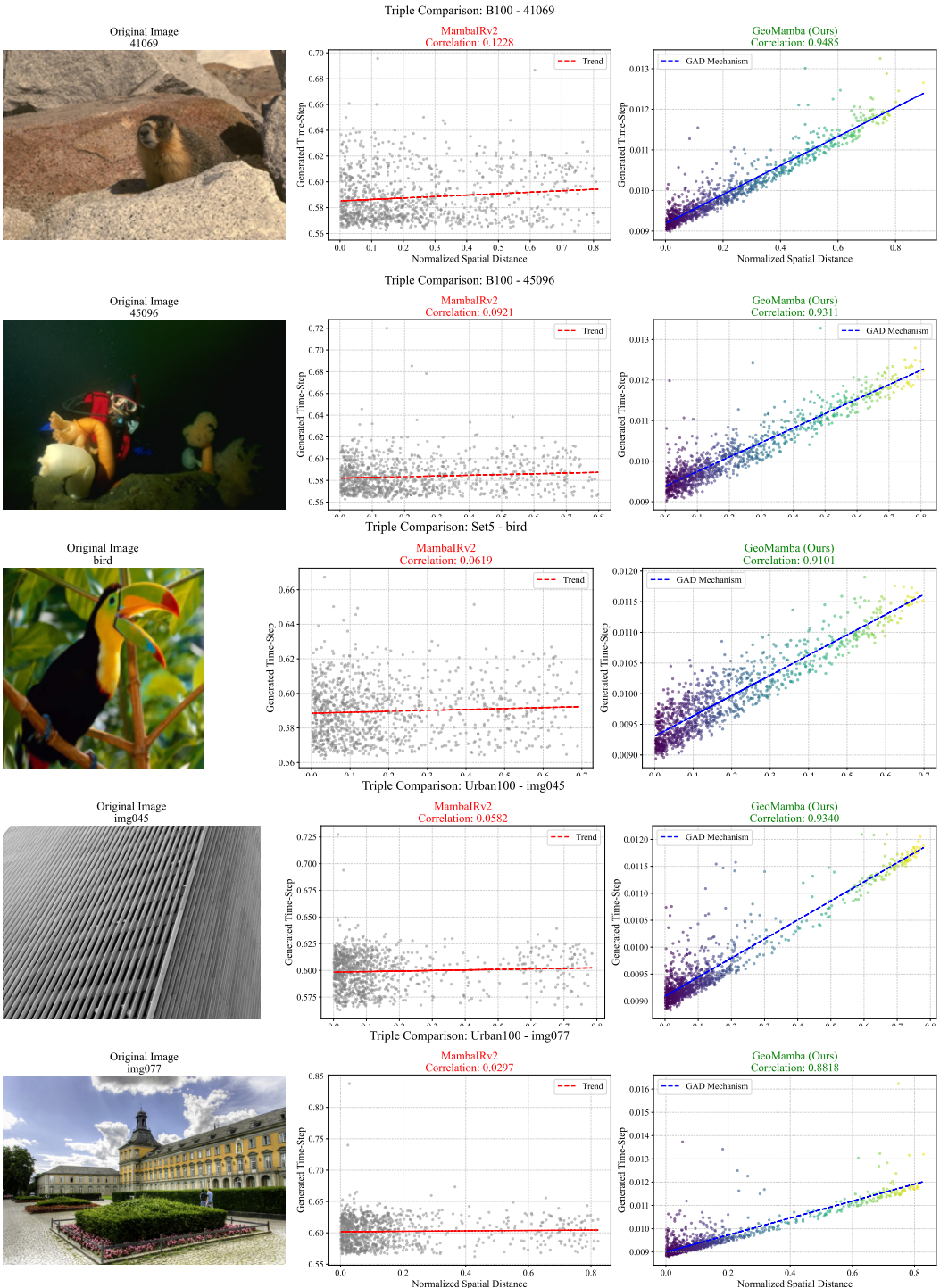

Figure 10: Qualitative examples for analyzing geometry-aware state propagation. GeoMamba better preserves boundaries and repeated structures when semantic reordering introduces long-range token transitions.

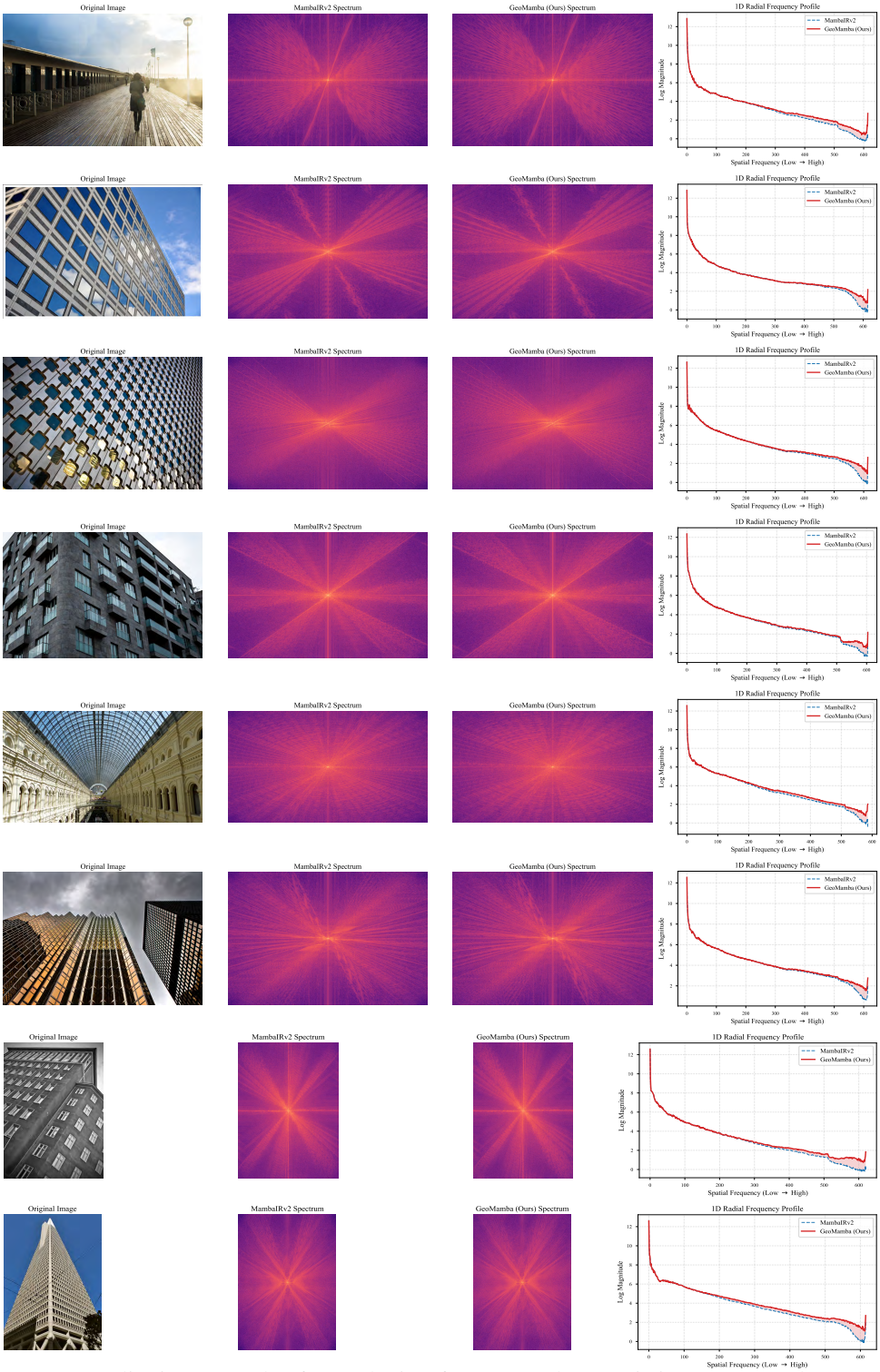

Figure 11: Qualitative examples for analyzing frequency characteristics. GeoMamba preserves more high-frequency detail around edges and textures, which is consistent with the gradient-aware continuous modulation in CDM.

NeurIPS Paper Checklist

The checklist is designed to encourage best practices for responsible machine learning research, addressing issues of reproducibility, transparency, research ethics, and societal impact. Do not remove the checklist: **The papers not including the checklist will be desk rejected.** The checklist should follow the references and follow the (optional) supplemental material. The checklist does NOT count towards the page limit.

Please read the checklist guidelines carefully for information on how to answer these questions. For each question in the checklist:

- You should answer [Yes], [No], or [N/A].
- [N/A] means either that the question is Not Applicable for that particular paper or the relevant information is Not Available.
- Please provide a short (1–2 sentence) justification right after your answer (even for [N/A]).

The checklist answers are an integral part of your paper submission. They are visible to the reviewers, area chairs, senior area chairs, and ethics reviewers. You will also be asked to include it (after eventual revisions) with the final version of your paper, and its final version will be published with the paper.

The reviewers of your paper will be asked to use the checklist as one of the factors in their evaluation. While [Yes] is generally preferable to [No], it is perfectly acceptable to answer [No] provided a proper justification is given (e.g., error bars are not reported because it would be too computationally expensive" or "we were unable to find the license for the dataset we used"). In general, answering [No] or [N/A] is not grounds for rejection. While the questions are phrased in a binary way, we acknowledge that the true answer is often more nuanced, so please just use your best judgment and write a justification to elaborate. All supporting evidence can appear either in the main paper or the supplemental material, provided in appendix. If you answer [Yes] to a question, in the justification please point to the section(s) where related material for the question can be found.

IMPORTANT, please:

- **Delete this instruction block, but keep the section heading "NeurIPS Paper Checklist"**,
- **Keep the checklist subsection headings, questions/answers and guidelines below.**
- **Do not modify the questions and only use the provided macros for your answers**.

1. **Claims**

   Question: Do the main claims made in the abstract and introduction accurately reflect the paper's contributions and scope?

   Answer: [Yes]

   Justification: Yes. The abstract and introduction accurately reflect our contributions. All claims are supported by extensive experimental results.

   Guidelines:

   - The answer [N/A] means that the abstract and introduction do not include the claims made in the paper.
   - The abstract and/or introduction should clearly state the claims made, including the contributions made in the paper and important assumptions and limitations. A [No] or [N/A] answer to this question will not be perceived well by the reviewers.
   - The claims made should match theoretical and experimental results, and reflect how much the results can be expected to generalize to other settings.
   - It is fine to include aspirational goals as motivation as long as it is clear that these goals are not attained by the paper.

2. **Limitations**

   Question: Does the paper discuss the limitations of the work performed by the authors?

   Answer: [Yes]

Justification: Yes. Section D discusses the limitations of our work.

Guidelines:

- The answer [N/A] means that the paper has no limitation while the answer [No] means that the paper has limitations, but those are not discussed in the paper.
- The authors are encouraged to create a separate "Limitations" section in their paper.
- The paper should point out any strong assumptions and how robust the results are to violations of these assumptions (e.g., independence assumptions, noiseless settings, model well-specification, asymptotic approximations only holding locally). The authors should reflect on how these assumptions might be violated in practice and what the implications would be.
- The authors should reflect on the scope of the claims made, e.g., if the approach was only tested on a few datasets or with a few runs. In general, empirical results often depend on implicit assumptions, which should be articulated.
- The authors should reflect on the factors that influence the performance of the approach. For example, a facial recognition algorithm may perform poorly when image resolution is low or images are taken in low lighting. Or a speech-to-text system might not be used reliably to provide closed captions for online lectures because it fails to handle technical jargon.
- The authors should discuss the computational efficiency of the proposed algorithms and how they scale with dataset size.
- If applicable, the authors should discuss possible limitations of their approach to address problems of privacy and fairness.
- While the authors might fear that complete honesty about limitations might be used by reviewers as grounds for rejection, a worse outcome might be that reviewers discover limitations that aren't acknowledged in the paper. The authors should use their best judgment and recognize that individual actions in favor of transparency play an important role in developing norms that preserve the integrity of the community. Reviewers will be specifically instructed to not penalize honesty concerning limitations.

3. **Theory assumptions and proofs**

   Question: For each theoretical result, does the paper provide the full set of assumptions and a complete (and correct) proof?

   Answer: [Yes]

   Justification: Yes, please see Section 3.

   Guidelines:

   - The answer [N/A] means that the paper does not include theoretical results.
   - All the theorems, formulas, and proofs in the paper should be numbered and cross-referenced.
   - All assumptions should be clearly stated or referenced in the statement of any theorems.
   - The proofs can either appear in the main paper or the supplemental material, but if they appear in the supplemental material, the authors are encouraged to provide a short proof sketch to provide intuition.
   - Inversely, any informal proof provided in the core of the paper should be complemented by formal proofs provided in appendix or supplemental material.
   - Theorems and Lemmas that the proof relies upon should be properly referenced.

4. **Experimental result reproducibility**

   Question: Does the paper fully disclose all the information needed to reproduce the main experimental results of the paper to the extent that it affects the main claims and/or conclusions of the paper (regardless of whether the code and data are provided or not)?

   Answer: [Yes]

   Justification: Please see Section 4.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.

- If the paper includes experiments, a [No] answer to this question will not be perceived well by the reviewers: Making the paper reproducible is important, regardless of whether the code and data are provided or not.
- If the contribution is a dataset and/or model, the authors should describe the steps taken to make their results reproducible or verifiable.
- Depending on the contribution, reproducibility can be accomplished in various ways. For example, if the contribution is a novel architecture, describing the architecture fully might suffice, or if the contribution is a specific model and empirical evaluation, it may be necessary to either make it possible for others to replicate the model with the same dataset, or provide access to the model. In general. releasing code and data is often one good way to accomplish this, but reproducibility can also be provided via detailed instructions for how to replicate the results, access to a hosted model (e.g., in the case of a large language model), releasing of a model checkpoint, or other means that are appropriate to the research performed.
- While NeurIPS does not require releasing code, the conference does require all submissions to provide some reasonable avenue for reproducibility, which may depend on the nature of the contribution. For example
  (a) If the contribution is primarily a new algorithm, the paper should make it clear how to reproduce that algorithm.
  (b) If the contribution is primarily a new model architecture, the paper should describe the architecture clearly and fully.
  (c) If the contribution is a new model (e.g., a large language model), then there should either be a way to access this model for reproducing the results or a way to reproduce the model (e.g., with an open-source dataset or instructions for how to construct the dataset).
  (d) We recognize that reproducibility may be tricky in some cases, in which case authors are welcome to describe the particular way they provide for reproducibility. In the case of closed-source models, it may be that access to the model is limited in some way (e.g., to registered users), but it should be possible for other researchers to have some path to reproducing or verifying the results.

5. **Open access to data and code**

   Question: Does the paper provide open access to the data and code, with sufficient instructions to faithfully reproduce the main experimental results, as described in supplemental material?

   Answer: [No]

   Justification: Not at the current submission stage. We do not provide a public release of the codebase and complete reproduction package at this time, but we plan to release them in the final version.

   Guidelines:

   - The answer [N/A] means that paper does not include experiments requiring code.
   - Please see the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - While we encourage the release of code and data, we understand that this might not be possible, so [No] is an acceptable answer. Papers cannot be rejected simply for not including code, unless this is central to the contribution (e.g., for a new open-source benchmark).
   - The instructions should contain the exact command and environment needed to run to reproduce the results. See the NeurIPS code and data submission guidelines (`https://neurips.cc/public/guides/CodeSubmissionPolicy`) for more details.
   - The authors should provide instructions on data access and preparation, including how to access the raw data, preprocessed data, intermediate data, and generated data, etc.
   - The authors should provide scripts to reproduce all experimental results for the new proposed method and baselines. If only a subset of experiments are reproducible, they should state which ones are omitted from the script and why.
   - At submission time, to preserve anonymity, the authors should release anonymized versions (if applicable).

- Providing as much information as possible in supplemental material (appended to the paper) is recommended, but including URLs to data and code is permitted.

6. **Experimental setting/details**

   Question: Does the paper specify all the training and test details (e.g., data splits, hyperparameters, how they were chosen, type of optimizer) necessary to understand the results?

   Answer: [Yes]

   Justification: Please see Section 4.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The experimental setting should be presented in the core of the paper to a level of detail that is necessary to appreciate the results and make sense of them.
   - The full details can be provided either with the code, in appendix, or as supplemental material.

7. **Experiment statistical significance**

   Question: Does the paper report error bars suitably and correctly defined or other appropriate information about the statistical significance of the experiments?

   Answer: [No]

   Justification: However, we provided extensive experimental results, including quantitative metrics and visual quality, in Section 4 to support our claims.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The authors should answer [Yes] if the results are accompanied by error bars, confidence intervals, or statistical significance tests, at least for the experiments that support the main claims of the paper.
   - The factors of variability that the error bars are capturing should be clearly stated (for example, train/test split, initialization, random drawing of some parameter, or overall run with given experimental conditions).
   - The method for calculating the error bars should be explained (closed form formula, call to a library function, bootstrap, etc.)
   - The assumptions made should be given (e.g., Normally distributed errors).
   - It should be clear whether the error bar is the standard deviation or the standard error of the mean.
   - It is OK to report 1-sigma error bars, but one should state it. The authors should preferably report a 2-sigma error bar than state that they have a 96% CI, if the hypothesis of Normality of errors is not verified.
   - For asymmetric distributions, the authors should be careful not to show in tables or figures symmetric error bars that would yield results that are out of range (e.g., negative error rates).
   - If error bars are reported in tables or plots, the authors should explain in the text how they were calculated and reference the corresponding figures or tables in the text.

8. **Experiments compute resources**

   Question: For each experiment, does the paper provide sufficient information on the computer resources (type of compute workers, memory, time of execution) needed to reproduce the experiments?

   Answer: [Yes]

   Justification: Please see Table 1 and Figure 6.

   Guidelines:

   - The answer [N/A] means that the paper does not include experiments.
   - The paper should indicate the type of compute workers CPU or GPU, internal cluster, or cloud provider, including relevant memory and storage.

- The paper should provide the amount of compute required for each of the individual experimental runs as well as estimate the total compute.
- The paper should disclose whether the full research project required more compute than the experiments reported in the paper (e.g., preliminary or failed experiments that didn't make it into the paper).

9. **Code of ethics**

   Question: Does the research conducted in the paper conform, in every respect, with the NeurIPS Code of Ethics `https://neurips.cc/public/EthicsGuidelines`?

   Answer: [Yes]

   Justification: We followed the NeurIPS Code of Ethics.

   Guidelines:

   - The answer [N/A] means that the authors have not reviewed the NeurIPS Code of Ethics.
   - If the authors answer [No], they should explain the special circumstances that require a deviation from the Code of Ethics.
   - The authors should make sure to preserve anonymity (e.g., if there is a special consideration due to laws or regulations in their jurisdiction).

10. **Broader impacts**

    Question: Does the paper discuss both potential positive societal impacts and negative societal impacts of the work performed?

    Answer: [N/A]

    Justification: There is no societal impact of the work performed.

    Guidelines:

    - The answer [N/A] means that there is no societal impact of the work performed.
    - If the authors answer [N/A] or [No], they should explain why their work has no societal impact or why the paper does not address societal impact.
    - Examples of negative societal impacts include potential malicious or unintended uses (e.g., disinformation, generating fake profiles, surveillance), fairness considerations (e.g., deployment of technologies that could make decisions that unfairly impact specific groups), privacy considerations, and security considerations.
    - The conference expects that many papers will be foundational research and not tied to particular applications, let alone deployments. However, if there is a direct path to any negative applications, the authors should point it out. For example, it is legitimate to point out that an improvement in the quality of generative models could be used to generate Deepfakes for disinformation. On the other hand, it is not needed to point out that a generic algorithm for optimizing neural networks could enable people to train models that generate Deepfakes faster.
    - The authors should consider possible harms that could arise when the technology is being used as intended and functioning correctly, harms that could arise when the technology is being used as intended but gives incorrect results, and harms following from (intentional or unintentional) misuse of the technology.
    - If there are negative societal impacts, the authors could also discuss possible mitigation strategies (e.g., gated release of models, providing defenses in addition to attacks, mechanisms for monitoring misuse, mechanisms to monitor how a system learns from feedback over time, improving the efficiency and accessibility of ML).

11. **Safeguards**

    Question: Does the paper describe safeguards that have been put in place for responsible release of data or models that have a high risk for misuse (e.g., pre-trained language models, image generators, or scraped datasets)?

    Answer: [N/A]

    Justification: This question is unrelated to the research topic of this article.

    Guidelines:

- The answer [N/A] means that the paper poses no such risks.
- Released models that have a high risk for misuse or dual-use should be released with necessary safeguards to allow for controlled use of the model, for example by requiring that users adhere to usage guidelines or restrictions to access the model or implementing safety filters.
- Datasets that have been scraped from the Internet could pose safety risks. The authors should describe how they avoided releasing unsafe images.
- We recognize that providing effective safeguards is challenging, and many papers do not require this, but we encourage authors to take this into account and make a best faith effort.

12. **Licenses for existing assets**

    Question: Are the creators or original owners of assets (e.g., code, data, models), used in the paper, properly credited and are the license and terms of use explicitly mentioned and properly respected?

    Answer: [Yes]

    Justification: Yes, we credited them in appropriate ways.

    Guidelines:

    - The answer [N/A] means that the paper does not use existing assets.
    - The authors should cite the original paper that produced the code package or dataset.
    - The authors should state which version of the asset is used and, if possible, include a URL.
    - The name of the license (e.g., CC-BY 4.0) should be included for each asset.
    - For scraped data from a particular source (e.g., website), the copyright and terms of service of that source should be provided.
    - If assets are released, the license, copyright information, and terms of use in the package should be provided. For popular datasets, `paperswithcode.com/datasets` has curated licenses for some datasets. Their licensing guide can help determine the license of a dataset.
    - For existing datasets that are re-packaged, both the original license and the license of the derived asset (if it has changed) should be provided.
    - If this information is not available online, the authors are encouraged to reach out to the asset's creators.

13. **New assets**

    Question: Are new assets introduced in the paper well documented and is the documentation provided alongside the assets?

    Answer: [N/A]

    Justification: his paper does not release new assets.

    Guidelines:

    - The answer [N/A] means that the paper does not release new assets.
    - Researchers should communicate the details of the dataset/code/model as part of their submissions via structured templates. This includes details about training, license, limitations, etc.
    - The paper should discuss whether and how consent was obtained from people whose asset is used.
    - At submission time, remember to anonymize your assets (if applicable). You can either create an anonymized URL or include an anonymized zip file.

14. **Crowdsourcing and research with human subjects**

    Question: For crowdsourcing experiments and research with human subjects, does the paper include the full text of instructions given to participants and screenshots, if applicable, as well as details about compensation (if any)?

    Answer: [N/A]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Including this information in the supplemental material is fine, but if the main contribution of the paper involves human subjects, then as much detail as possible should be included in the main paper.
- According to the NeurIPS Code of Ethics, workers involved in data collection, curation, or other labor should be paid at least the minimum wage in the country of the data collector.

15. **Institutional review board (IRB) approvals or equivalent for research with human subjects**

Question: Does the paper describe potential risks incurred by study participants, whether such risks were disclosed to the subjects, and whether Institutional Review Board (IRB) approvals (or an equivalent approval/review based on the requirements of your country or institution) were obtained?

Answer: [N/A]

Justification: This paper does not involve crowdsourcing nor research with human subjects.

Guidelines:

- The answer [N/A] means that the paper does not involve crowdsourcing nor research with human subjects.
- Depending on the country in which research is conducted, IRB approval (or equivalent) may be required for any human subjects research. If you obtained IRB approval, you should clearly state this in the paper.
- We recognize that the procedures for this may vary significantly between institutions and locations, and we expect authors to adhere to the NeurIPS Code of Ethics and the guidelines for their institution.
- For initial submissions, do not include any information that would break anonymity (if applicable), such as the institution conducting the review.

16. **Declaration of LLM usage**

Question: Does the paper describe the usage of LLMs if it is an important, original, or non-standard component of the core methods in this research? Note that if the LLM is used only for writing, editing, or formatting purposes and does *not* impact the core methodology, scientific rigor, or originality of the research, declaration is not required.

Answer: [N/A]

Justification: The core method development in this research does not involve LLMs as any important, original, or non-standard components.

Guidelines:

- The answer [N/A] means that the core method development in this research does not involve LLMs as any important, original, or non-standard components.
- Please refer to our LLM policy in the NeurIPS handbook for what should or should not be described.