GRAPHENE: A LIGHTWEIGHT TRANSFORMER WITH ISING MEAN-FIELD REFINEMENT FOR IMAGE SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

This paper proposes *Graphene*, a lightweight image super-resolution architecture with two complementary feature refinement mechanisms: Ising-model-inspired local feature refinement using an *Ising Refinement Block* (IRB) and stage-history-based feature refinement using a *History Attention* (HistAttn) module. The IRB reuses attention output features and pre-softmax attention logits to construct signed pairwise token couplings within each local window, applies unrolled mean-field updates, and refines features through a gated residual correction. These signed couplings allow alignment and opposition between latent spins to modulate local feature refinement. HistAttn attends over the current stage output and earlier representations at the same token position and adaptively blends the current stage output with a history-weighted feature aggregate without adding spatial pairwise attention. For *classical SR* and *lightweight SR* on Urban100 at $\times 4$, Graphene and Graphene-light outperform the state-of-the-art methods IET and IET-light by **0.24 dB** and **0.38 dB**, respectively, with **34.3%/13.7% fewer parameters**, **44.8%/44.4% fewer FLOPs**, **87.9%/84.5% lower latency**, and **78.3%/87.5% lower peak memory**. *Real-world SR* comparisons further show that Graphene recovers fine details and preserves textures across diverse degradations. `If the paper is accepted, the code will be released by the camera-ready deadline.`

1 INTRODUCTION

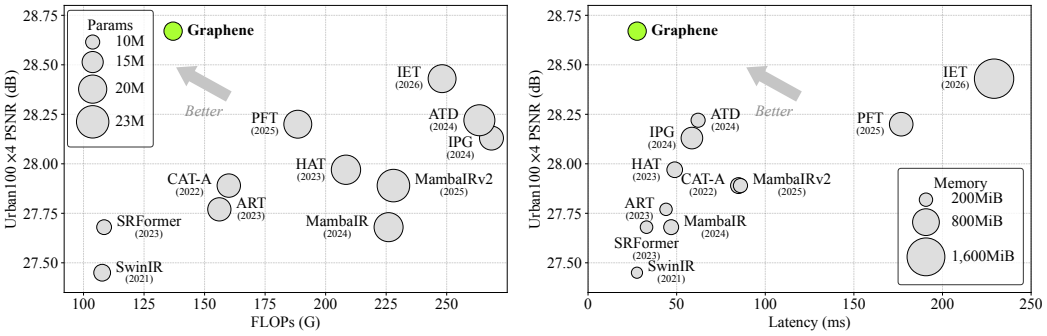

Figure 1: PSNR, FLOPs, latency, parameter count, and peak memory on Urban100 at $\times 4$. *Left:* PSNR versus FLOPs, with circle areas indicating parameter counts. *Right:* PSNR versus latency, with circle areas indicating peak memory. Profiling uses 64×64 LR inputs.

Image super-resolution (SR) reconstructs fine image structure from low-resolution (LR) observations. Windowed attention aggregates features within local regions, and recent architectures improve reconstruction by expanding or adapting the features available to each token (Liang et al., 2021; Chen et al., 2023; Zhou et al., 2023; Guo et al., 2024; 2025; Long et al., 2025; Meng et al., 2026). The challenge is to improve reconstruction while limiting parameter count, FLOPs, latency, and peak memory.

Reviewers: please read the Reviewer Guidelines (iclr.cc/Conferences/2027/ReviewerGuidelines) and the AI Policy for Reviewers (iclr.cc/Conferences/2027/AIPolicyForReviewers). If you used AI to expand, edit, or polish your review, please provide the input text to the LLM. Better still, consider skipping the LLM and submitting your original text: we, and the authors, are much more interested in your unedited thoughts than in what an LLM has to say. AI-assisted or not, you are putting your name and reputation behind your review: LLM-generated falsehoods, hallucinations or misrepresentations are subject to disciplinary action, which may include desk-rejecting all papers you have authored.

Graphene reuses two quantities already computed in a windowed Transformer: pre-softmax attention logits and intermediate stage outputs. The logits coordinate residual-correction gates within each window; the stored stage outputs supply representations from different depths at the same token position. Separate IRB and HistAttn modules implement these operations within the LR feature-processing trunk.

The *Ising Refinement Block* (IRB) converts logits from Permuted Self-Attention (PSA) (Zhou et al., 2023) into signed, symmetric interactions between latent spins. Three mean-field updates produce magnetizations that gate learned residual corrections to PSA output features. Positive and negative couplings favor aligned and opposing spins, respectively; the resulting gates coordinate correction magnitudes within the processing window.

The *History Attention* (HistAttn) module uses a learned probe to score the current stage output and earlier representations at each token position. Softmax weights form a feature aggregate, which is fused with the current output through a learned scalar. The weights depend on the input features; the aggregation adds no spatial pairwise attention.

Component ablations compare IRB and HistAttn at a fixed window size; coupling ablations vary signs and symmetry; and paired continuation compares intact IRB with a unary-only gate. Transfer experiments integrate IRB into HAT (Chen et al., 2023). Graphene has the highest reported PSNR in all 15 classical dataset–scale comparisons, and Graphene-light leads 12 of 15 lightweight comparisons, including all five at $\times 4$. Figure 1 compares classical SR PSNR against FLOPs and latency, with circle areas encoding parameter count and peak memory.

The contributions are:

- This work introduces *IRB*, an Ising-model-inspired refinement block that derives signed local interactions from attention logits and uses finite mean-field updates to gate residual feature corrections, *enabling coordinated refinement of fine local structures*.
- This work introduces *HistAttn*, a stage-history attention module that adaptively reuses current and earlier representations at each token position, allowing later stages to retain and *combine useful features from different depths without additional spatial pairwise attention*.
- This work achieves state-of-the-art reconstruction quality on *classical, lightweight, and real-world SR benchmarks* while maintaining *low parameter counts, FLOPs, latency, and memory usage*, as demonstrated for classical SR in Figure 1.

2 Related Work

Windowed Transformer SR progresses from SwinIR (Liang et al., 2021), which adopts alternating window-based and shifted-window self-attention from the Swin Transformer (Liu et al., 2021), to HAT (Chen et al., 2023), which adds channel-attention convolution and overlapping cross-window attention to strengthen local and neighboring-window interactions. SRFormer (Zhou et al., 2023) introduces Permuted Self-Attention (PSA), which compresses and spatially rearranges keys and values while preserving full-resolution queries to reduce attention cost, and a convolutional feed-forward network (ConvFFN) to recover high-frequency details. Beyond these windowed Transformers, MambaIR (Guo et al., 2024) adapts Mamba (Gu & Dao, 2024) to SR using four-direction 2D selective scans with local convolution and channel attention; MambaIRv2 (Guo et al., 2025) combines state-space modeling, semantic prompts, pixel reordering, local attention, and the SRFormer-style ConvFFN (Zhou et al., 2023) for non-causal restoration; PFT (Long et al., 2025) progressively filters irrelevant tokens through cross-layer Progressive Focused Attention; and IET (Meng et al., 2026) uses Individualized Exploratory Attention to select content-adaptive candidates for each token.

3 Methods

3.1 Architecture and Design Rationale

Graphene processes LR features with a convolutional stem, a sequence of residual Graphene stages, a trunk convolution, and a PixelShuffle reconstruction head (Shi et al., 2016). Each Graphene block contains LayerNorm, PSA (Zhou et al., 2023), an IRB when enabled, and a residual LayerNorm–

ConvFFN branch. HistAttn is inserted between successive stages. Figure 2 summarizes the full configuration.

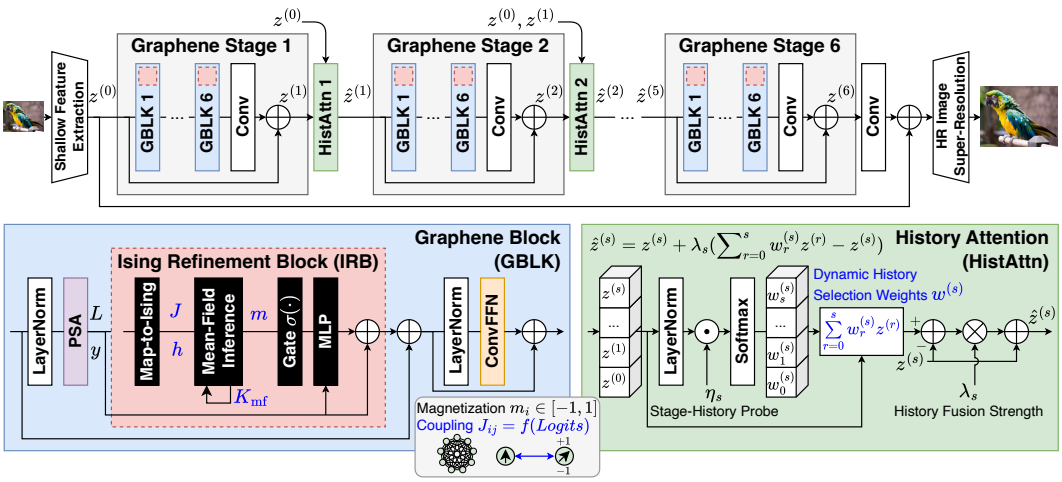

Figure 2: Overall architecture of Graphene. *Top:* Given an LR input, Graphene applies shallow feature extraction with a convolutional stem, deep feature extraction through six Graphene stages, each containing six Graphene Blocks (GBLKs), followed by a trunk convolution, and HR reconstruction with a PixelShuffle head (Shi et al., 2016) and a long residual connection. *Bottom-left:* Each GBLK applies LayerNorm, PSA (Zhou et al., 2023), an Ising Refinement Block (IRB), and a residual LayerNorm–ConvFFN branch. The IRB takes the PSA output features y and pre-softmax logits L as inputs: each token feature y_i is projected to a scalar field h_i and an initial magnetization $m_i^{(0)}$, while the logits L are mapped to a signed, symmetric window-level coupling matrix J with a zero diagonal. Thus, the resulting couplings define a window-local, undirected spin interaction graph. The IRB then performs K_{mf} mean-field updates to obtain magnetizations m, which gate a lightweight residual MLP correction. The final magnetizations modulate token-wise residual corrections through their gates. *Bottom-right:* HistAttn connects successive Graphene stages by matching a learnable stage-history probe η_s with LayerNorm-normalized stored features, producing dynamic history selection weights $w_r^{(s)}$ over the stage history $\{z^{(0)}, \ldots, z^{(s)}\}$, where r indexes the stored feature. The selected history is then fused with the current stage output $z^{(s)}$ using a learnable history fusion strength λ_s: $\hat{z}^{(s)} = z^{(s)} + \lambda_s(\sum_{r=0}^{s} w_r^{(s)} z^{(r)} - z^{(s)})$. This update refines the current stage output by adaptively blending it with a history-weighted feature aggregate.

Two refinement axes. Within each head, PSA aggregates value features with nonnegative attention weights. The IRB reuses the underlying logits to coordinate the strengths of token-wise residual corrections through signed latent-spin interactions, without computing another query–key affinity matrix. Earlier stages provide alternative representations at the same spatial location. HistAttn makes their weighting depend on the token features, while residual fusion preserves a direct path from the current stage output. Together, these operations augment local refinement and depth-wise reuse without global spatial pairwise attention.

Full and lightweight configurations. Graphene uses six stages, six blocks per stage, 180 channels, and window side length 32. Graphene-light uses four stages, six blocks per stage, 60 channels, and window side length 16. Both use six attention heads, PSA reduction factor $\alpha_{\mathrm{psa}} = 2$, and $K_{\mathrm{mf}} = 3$ mean-field updates per enabled IRB. IRB is enabled in every full-model block and only in the last block of each lightweight stage. Complete dimensions, initialization, and reconstruction settings are given in Appendix B of the supplementary material.

3.2 Ising Refinement Block

Attention logits as local couplings. Consider a window with $n = W^2$ query tokens and H attention heads. PSA produces logits between query i and reduced key token u:

$$L_{iu}^{(a)} = \frac{\langle q_i^{(a)}, k_u^{(a)} \rangle}{\sqrt{d_h}} + b_{iu}^{(a)} + \mathcal{M}_{iu}, \qquad d_h = C_{\mathrm{feat}}/H, \tag{1}$$

where a indexes heads, C_{feat} is the feature width, b is the relative-position bias, and $\mathcal{M}$ is the optional shifted-window mask. Queries and keys are projections of LayerNorm-normalized input features. PSA compresses and permutes the key/value branches, leaving n/α_{psa}^2 key/value tokens per window.

Let $\pi(j)$ map fine token j to its reduced key cell. Lifting the logits and averaging heads gives

$$\bar{L}_{ij}^{(a)} = \frac{L_{i,\pi(j)}^{(a)}}{\alpha_{\text{psa}}^2}, \qquad \tilde{J}_{ij} = \begin{cases} H^{-1} \sum_{a=1}^{H} \bar{L}_{ij}^{(a)}, & \text{valid pair,} \\ 0, & \text{otherwise.} \end{cases} \tag{2}$$

Each reduced key cell is replicated α_{psa}^2 times; the reciprocal factor preserves its summed contribution before subsequent transformations. With $\alpha_{\text{psa}} = 2$, each reduced logit gives four copies scaled by $1/4$. Head averaging supplies one interaction graph. Invalid shifted-window pairs are zeroed at this step to remove the attention-mask sentinel values.

The couplings are then centered, symmetrized, and made diagonal-free:

$$\hat{J}_{ij} = \tilde{J}_{ij} - \frac{1}{n} \sum_{k=1}^{n} \tilde{J}_{ik}, \qquad J_{ij} = \frac{\gamma}{2} (\hat{J}_{ij} + \hat{J}_{ji})\, \mathbf{1}[i \neq j], \tag{3}$$

where $\gamma > 0$ controls coupling strength. Centering expresses affinities relative to each query's mean and permits both signs; symmetrization supplies reciprocal interactions for the undirected Ising energy. The final matrix has no self-couplings, but its row sums need not remain zero. Centering and symmetrization can produce nonzero couplings at previously masked positions. The graph remains confined to the processing window.

Latent spins and finite mean-field inference. For PSA output feature $y_i \in \mathbb{R}^{C_{\text{feat}}}$, learned affine projections define a unary field and an initial magnetization:

$$h_i = W_h^\top y_i + b_h, \qquad m_i^{(0)} = \tanh(W_m^\top y_i + b_m). \tag{4}$$

The update is motivated by binary spins $\xi_i \in \{-1, +1\}$ with energy

$$E(\boldsymbol{\xi}) = -\sum_i h_i \xi_i - \frac{1}{2} \sum_{i,j} J_{ij} \xi_i \xi_j, \qquad p(\boldsymbol{\xi}) = Z^{-1} \exp[-\beta E(\boldsymbol{\xi})], \tag{5}$$

where $\beta > 0$ is the inverse temperature and Z is the partition function. Under a factorized distribution q, the magnetization $m_i = \mathbb{E}_q[\xi_i]$ is the mean latent spin. Mean-field stationarity gives the self-consistency equation used by IRB as a differentiable recurrence:

$$m_i^{(t+1)} = \tanh\left(\beta \left(h_i + \sum_{j=1}^{n} J_{ij} m_j^{(t)} \right) \right), \qquad t = 0, \ldots, K_{\text{mf}} - 1. \tag{6}$$

Positive and negative couplings give reinforcing and opposing influences between latent states, respectively. The field and coupling matrix remain fixed within each unroll. Three updates per IRB give 108 and 12 updates along the full and lightweight forward paths, respectively, with a separate local system in each block.

Spin-gated residual correction. The final magnetization controls a token-wise residual MLP:

$$g_i = \sigma(\kappa m_i^{(K_{\text{mf}})}), \qquad \tilde{y}_i = y_i + g_i R(y_i), \tag{7}$$

where σ is the logistic sigmoid, κ controls gate scale, and R applies LayerNorm followed by linear–GELU–linear layers. The identity path retains the PSA feature; the residual determines the correction direction, and signed spin interactions modulate its positive gate. For $\kappa = 1$, $m_i \in [-1, 1]$ gives $g_i \in [0.269, 0.731]$ approximately, with $g_i = 0.5$ at neutral magnetization. Small learned residuals can preserve homogeneous features, while larger residuals permit stronger refinement. Graphene-light additionally scales R by a learned per-IRB scalar initialized to zero.

The finite unroll is bounded by $\tanh$ and does not require convergence to a fixed point. For a fixed window, $\beta \|J\|_2 < 1$ is a sufficient contraction condition, but it is not enforced during training. Appendix A of the supplementary material gives the mean-field derivation, convergence-rate bound, and proofs verified in Lean (de Moura & Ullrich, 2021).

3.3 HISTORY ATTENTION

Let $z^{(0)} \in \mathbb{R}^{B\times N\times C_{\text{feat}}}$ denote shallow features, with batch size B and N spatial tokens. Let $\mathcal{G}_s$ be stage s, and define $\hat{z}^{(0)} = z^{(0)}$. The stage output is $z^{(s)} = \mathcal{G}_s(\hat{z}^{(s-1)})$ for $s = 1, \dots, S$. HistAttn refines this output before the next stage, for $s < S$.

At position p, a learned, input-independent probe $\eta_s \in \mathbb{R}^{C_{\text{feat}}}$ scores the stored features:

$$e_{r,p}^{(s)} = \frac{\langle \eta_s, \mathrm{LN}(z_p^{(r)})\rangle}{\sqrt{C_{\text{feat}}}} + \rho\,\mathbf{1}[r = s], \qquad w_{r,p}^{(s)} = \frac{\exp(e_{r,p}^{(s)})}{\sum_{\ell=0}^{s}\exp(e_{\ell,p}^{(s)})}. \tag{8}$$

The current stage is included in the history, and ρ biases selection toward it. Although the probe is shared across inputs, its scores and selection weights depend on the features at each token position. The aggregate and fused feature are

$$c_p^{(s)} = \sum_{r=0}^{s} w_{r,p}^{(s)} z_p^{(r)}, \qquad \hat{z}_p^{(s)} = z_p^{(s)} + \lambda_s(c_p^{(s)} - z_p^{(s)}). \tag{9}$$

The weights make $c_p^{(s)}$ a convex history aggregate. The learned scalar $\lambda_s \in \mathbb{R}$ controls residual fusion: zero preserves the current feature and one selects the aggregate. Because λ_s is unconstrained, the complete fusion is affine and may extrapolate. No HistAttn follows the final stage, so the deep feature is $z^{(S)}$.

Computational cost. For one image, window size $n = W^2$, fixed head count, and fixed MLP ratios, PSA costs $O(NC_{\text{feat}}^2 + NnC_{\text{feat}}/\alpha_{\text{psa}}^2)$ per block. IRB adds $O(K_{\text{mf}}Nn + NC_{\text{feat}}^2)$ for the coupling updates and residual MLP. HistAttn processes $s+1$ representations at insertion s, giving $O(S^2NC_{\text{feat}})$ total history-fusion cost. Working coupling storage is $O(Nn)$ per IRB and stored history occupies $O(SNC_{\text{feat}})$ feature entries. For fixed windows and architecture, the operation counts and stored coupling and feature entries scale linearly with image area.

4 EXPERIMENTS

4.1 EXPERIMENTAL SETUP

Data and evaluation. The training recipe uses DF2K, comprising DIV2K (Timofte et al., 2017) and Flickr2K (Lim et al., 2017), with bicubic LR inputs. Classical and lightweight SR are evaluated on Set5 (Bevilacqua et al., 2012), Set14 (Zeyde et al., 2010), BSD100 (Martin et al., 2001), Urban100 (Huang et al., 2015), and Manga109 (Matsui et al., 2017). PSNR/SSIM (Wang et al., 2004) are computed on the Y channel after removing a border equal to the scale factor. Graphene predictions use EMA weights, with the image conversion and evaluation conventions specified in Appendix B of the supplementary material.

Training. The base schedule runs for $500{,}000$ iterations on 128×128 LR patches with ℓ_1 loss. Muon (Jordan et al., 2024) updates linear-layer matrices in the shared LR encoder; Adam (Kingma & Ba, 2015), with $(\beta_1, \beta_2) = (0.9, 0.999)$, updates the remaining parameters. Both start at 2×10^{-4} and halve at $250{,}000$, $400{,}000$, $450{,}000$, and $475{,}000$ iterations. Training uses random flips and right-angle rotations, BF16 autocast (Micikevicius et al., 2018) with FP32 parameters and optimizer states, and global batch size 32 across 16 32GB RTX 5090 GPUs. The ablations below use their separately stated budgets.

Efficiency. Profiling uses 64×64 LR inputs and counts two FLOPs per multiply–accumulate. Latency and peak allocated memory are measured on one NVIDIA GH200 GPU at batch size one. Graphene and Graphene-light use BF16 autocast, matching training, with `torch.compile`. Latency is the median of three run means, each using 50 warm-up and 200 timed passes. Appendix B of the supplementary material provides the protocol and baseline settings.

4.2 ABLATION OF IRB

The ablations use DIV2K, $10{,}000$ training iterations, and Urban100 $\times 2$ evaluation, with the full backbone and three mean-field updates. The controls in Tables 1 and 2 train from scratch under this protocol, matching shared-parameter initialization, minibatches, and optimization settings.

Table 1: Component baselines and IRB interactions on Urban100 at $\times 2$. Panel (b) retains HistAttn and the residual MLP at window side length 32; Δ is relative to full Graphene.

(a) Component baselines

Window	IRB	HistAttn	Params (M)	FLOPs (G)	PSNR (dB)	Δ (dB)
16	–	–	10.3	91.0	30.14	−0.79
32	–	–	10.5	111.4	30.33	−0.60
32	–	Enabled	10.5	111.6	30.53	−0.40
32	Enabled	–	12.9	131.8	30.88	−0.05
32	Enabled	Enabled	12.9	132.0	**30.93**	0.00

(b) Gate and coupling controls

IRB configuration	Params (M)	FLOPs (G)	PSNR (dB)	Δ (dB)
Constant gate ($g_i = 0.5$)	12.8	130.9	30.85	−0.08
Unary-only gate ($J = 0$)	12.9	131.0	30.89	−0.04
Nonnegative, symmetric	12.9	132.0	30.90	−0.03
Signed, asymmetric	12.9	132.0	30.86	−0.07
Signed, symmetric (full IRB)	12.9	132.0	**30.93**	0.00

Module contributions. At window side length 32 (Table 1a), IRB improves PSNR from 30.33 to 30.88 dB, while HistAttn alone reaches 30.53 dB. Their combination reaches 30.93 dB, adding 0.60 dB to the same-window baseline. IRB adds 2.4M parameters and 20.4G FLOPs; HistAttn adds approximately 0.2G FLOPs. Increasing the baseline window from 16 to 32 separately adds 0.19 dB.

Interactions beyond residual capacity. The controls in Table 1b retain the residual MLP and HistAttn. The constant gate fixes $g_i = 0.5$; the unary-only gate sets $J = 0$ while retaining the learned field. Unary gating improves PSNR from 30.85 to 30.89 dB, and full IRB adds 0.04 dB to reach 30.93 dB. These differences support feature-dependent gating and pairwise coordination beyond the added residual MLP.

Signed, symmetric couplings. The controls in Table 1b give 30.90 dB with nonnegative couplings and 30.86 dB without symmetry, respectively 0.03 and 0.07 dB below full IRB. These results favor signed, symmetric interactions. Three mean-field updates also give the highest PSNR among the tested unroll lengths (Section C.1 of the supplementary material).

Complementary evidence. The supplementary material reports six-seed variation (30.92 ± 0.08 dB, mean $\pm$ sample SD; Section C.2), transfer gains in HAT (Section C.3), and a 0.0163 dB paired-continuation advantage over $J = 0$ (Section C.4). Both absolute scores decrease during continuation, and boundary-specific gains remain unresolved. These diagnostics do not quantify uncertainty in the gate and history comparisons.

4.3 Ablation of HistAttn

The controls in Table 2 retain full IRB. A learned shallow skip sets $c_p^{(s)} = z_p^{(0)}$ in Equation (9). Static selection replaces the feature-dependent scores with learned stage-specific logits shared across images and tokens. It accesses the same history as HistAttn, including the current stage. All enabled variants learn λ_s at the same stage insertions; Section C.1 of the supplementary material gives the control equations.

Table 2: Adaptive history selection on Urban100 at $\times 2$ with full IRB and window side length 32. Δ is relative to HistAttn.

History configuration	Params (M)	FLOPs (G)	PSNR (dB)	Δ (dB)
No history fusion	12.9	131.8	30.88	−0.05
Learned shallow skip	12.9	131.8	30.89	−0.04
Static history weights	12.9	131.8	30.90	−0.03
HistAttn	12.9	132.0	**30.93**	0.00

Selection beyond a skip connection. HistAttn improves PSNR from 30.88 to 30.93 dB (Table 2). A learned shallow skip reaches 30.89 dB, while static history weights give 30.90 dB. HistAttn's additional 0.04 and 0.03 dB support selecting representations according to token features beyond a shallow skip or fixed depth mixture, without spatial pairwise attention.

4.4 Activation and History Analysis

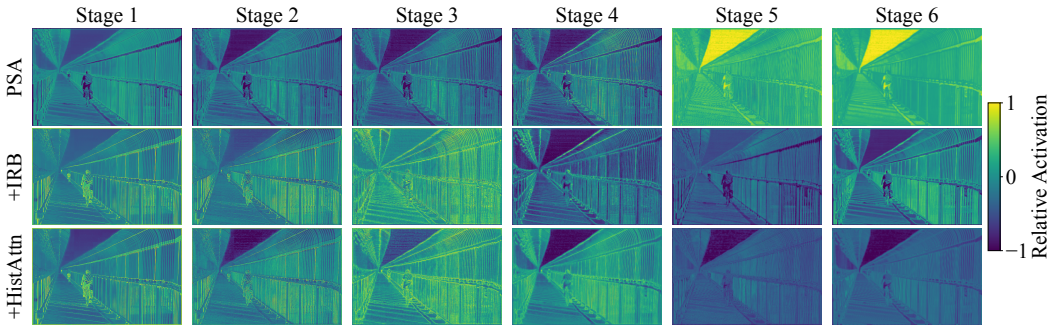

Figure 3: Stage-wise channel-mean activation maps for Urban100 img_024 at $\times 2$. The PSA row uses SRFormer; the other rows retain IRB or HistAttn while disabling the other module in Graphene at inference. Each panel is normalized independently, showing relative spatial organization across stages.

Activation patterns. In Figure 3, the IRB row shows localized contrasts around facade and window-frame structures, while the HistAttn row changes the distribution of responses across the facade with depth. These are observations of relative spatial patterns: each panel is normalized independently, and the Graphene rows disable one module at inference. They complement the retrained component comparisons in Table 1.

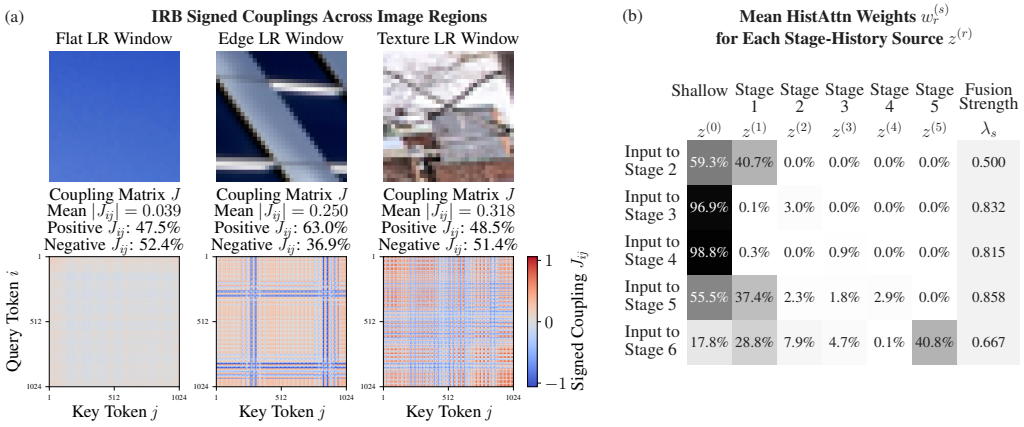

Figure 4: IRB and HistAttn analysis on Urban100 at $\times 2$. *Left:* Representative flat, edge, and texture windows and their signed coupling matrices. *Right:* Mean history selection weights before residual fusion, together with the learned fusion strengths. Individual-window values are distinct from category averages.

Structure-dependent interactions and history selection. Figure 4 summarizes the measured couplings and history selection. Across 18,221 LR windows, the mean absolute coupling in the selected edge and texture categories is $6.3\times$ and $7.8\times$ that in the flat category, using unrounded category means. HistAttn modules feeding Stages 3–4 assign 96.9–98.8% of their selection weight to the shallow feature. The insertion feeding Stage 6 assigns 40.8% to the current stage output and 59.2% to earlier features, before residual fusion. Appendix D of the supplementary material provides the category definitions, measurement protocols, and cross-scale comparisons.

4.5 PARAMETERS, FLOPS, LATENCY, PEAK MEMORY, AND PSNR

Table 3: Classical and lightweight $\times 4$ SR. Best and second-best PSNR values within each group are bold and underlined, with ties treated equally. Graphene's parameter counts, FLOPs, latency, and peak memory are also bolded. Profiling uses 64×64 LR inputs. Full PSNR/SSIM comparisons at $\times 2$, $\times 3$, and $\times 4$ are in Appendix E of the supplementary material.

Classical SR

Method	Params (M)	FLOPs (G)	Latency (ms)	Memory (MiB)	Set5	Set14	B100	U100	M109
SwinIR (Liang et al., 2021)	11.9	107.7	27.6	139.6	32.92	29.09	27.92	27.45	32.03
CAT-A (Chen et al., 2022)	16.6	160.0	84.9	282.1	33.08	29.18	27.99	27.89	32.39
ART (Zhang et al., 2023)	16.5	156.1	44.0	172.1	33.04	29.16	27.97	27.77	32.31
HAT (Chen et al., 2023)	20.8	208.4	48.9	274.2	33.04	29.23	28.00	27.97	32.48
SRFormer (Zhou et al., 2023)	10.5	108.5	33.0	176.2	32.93	29.08	27.94	27.68	32.21
IPG (Tian et al., 2024)	17.0	268.5	58.6	522.9	33.15	29.24	27.99	28.13	32.53
ATD (Zhang et al., 2024)	22.0	263.5	62.1	222.8	33.14	29.25	28.02	28.22	32.65
MambaIR (Guo et al., 2024)	20.6	226.0	46.9	249.8	33.03	29.20	27.98	27.68	32.32
MambaIRv2 (Guo et al., 2025)	23.1	228.0	86.0	224.6	33.14	29.23	28.00	27.89	32.57
PFT (Long et al., 2025)	19.8	188.5	176.7	621.8	33.15	29.29	28.02	28.20	32.63
IET (Meng et al., 2026)	19.8	248.1	229.1	1732.4	<u>33.22</u>	<u>29.35</u>	<u>28.06</u>	<u>28.43</u>	<u>32.81</u>
Graphene	**13.0**	**137.0**	**27.7**	**375.2**	**33.27**	**29.43**	**28.12**	**28.67**	**32.89**

Lightweight SR

Method	Params (K)	FLOPs (G)	Latency (ms)	Memory (MiB)	Set5	Set14	B100	U100	M109
SwinIR (Liang et al., 2021)	930.0	8.8	18.3	71.7	32.44	28.77	27.69	26.47	30.92
SRFormer (Zhou et al., 2023)	873.0	8.4	20.3	71.7	32.51	28.82	27.73	26.67	31.17
ATD (Zhang et al., 2024)	777.0	13.7	47.5	80.3	32.63	28.89	27.79	26.97	31.48
MambaIR (Guo et al., 2024)	924.0	12.0	27.2	68.0	32.51	28.85	27.75	26.75	31.26
MambaIRv2 (Guo et al., 2025)	790.0	10.1	48.5	87.3	32.51	28.84	27.75	26.82	31.24
PFT (Long et al., 2025)	792.0	10.3	108.4	326.5	32.63	28.92	27.79	27.20	31.51
UCAN (Tan et al., 2026)	705.0	11.3	71.6	195.4	32.65	28.95	27.79	26.89	31.50
IET (Meng et al., 2026)	801.0	12.6	72.4	578.3	<u>32.70</u>	<u>29.00</u>	<u>27.85</u>	<u>27.41</u>	<u>31.73</u>
Graphene-light	**691.6**	**7.0**	**11.2**	**72.1**	**32.80**	**29.19**	**27.94**	**27.79**	**31.93**

Classical SR. Table 3 compares $\times 4$ PSNR, parameter counts, FLOPs, latency, and peak memory. Graphene has the highest reported PSNR on all five datasets, extending to all 15 dataset–scale settings in Appendix E of the supplementary material. On Urban100 at $\times 4$, Graphene exceeds IET (Meng et al., 2026) by **0.24 dB**, using **34.3% fewer parameters**, **44.8% fewer FLOPs**, **87.9% lower latency**, and **78.3% lower peak memory**. Figure 1 visualizes these comparisons. Under the stated profiling settings, Graphene has lower latency and higher PSNR than SRFormer (Zhou et al., 2023), while SRFormer uses less peak memory.

Lightweight SR. The lower panel compares lightweight models, including UCAN (Tan et al., 2026). Graphene-light achieves the highest PSNR on the five $\times 4$ datasets and on 12 of the 15 settings across scales. Its Urban100 $\times 4$ gain over IET-light is **0.38 dB**, with **13.7% fewer parameters**, **44.4% fewer FLOPs**, **84.5% lower latency**, and **87.5% lower peak memory**. The four-stage Graphene-light model achieves these $\times 4$ PSNR values with 691.6K parameters.

Visual reconstruction and attribution. Figure 5 compares repeated structures in Urban100 img_045. Local Attribution Maps (LAM) (Gu & Dong, 2021) characterize the input locations associated with reconstruction of the selected output patch. Graphene obtains the highest displayed Diffusion Index (DI), indicating the broadest spatial attribution for this region. This measurement characterizes the selected prediction; it does not establish a dataset-wide explanation of the PSNR advantage. Section D.4 of the supplementary material specifies the attribution objective and protocol.

Real-world SR. Figure 6 compares Graphene with SwinIR, HAT, SRFormer, and MambaIR, the compared baselines providing real-world SR models. Graphene follows SwinIR's real-world training pipeline (Liang et al., 2021). The naturally degraded images have no aligned HR references, so the

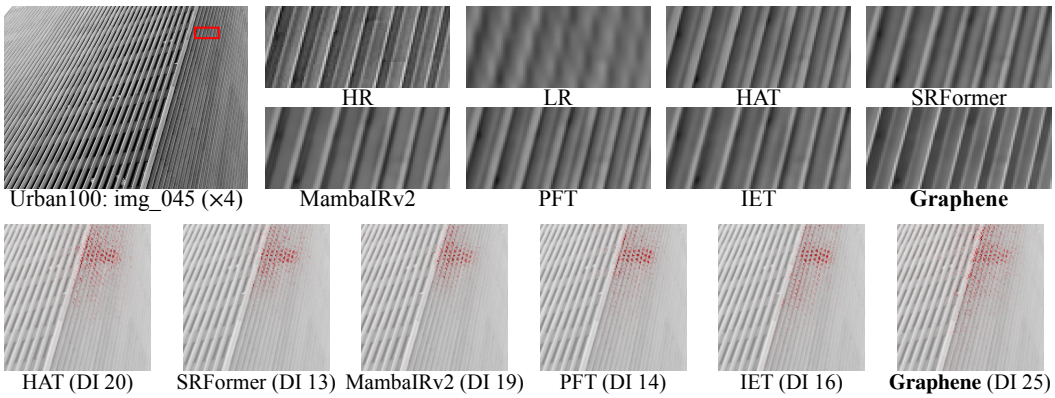

Figure 5: *Top:* $\times 4$ SR on Urban100 img_045. Graphene preserves the repeated line structure visible in the HR crop. *Bottom:* Local Attribution Maps for the outlined output region, with Diffusion Index (DI) values rounded to integers.

comparison is qualitative, as in SwinIR (Liang et al., 2021) and MambaIR (Guo et al., 2024). The displayed examples illustrate detail and texture recovery across different image content.

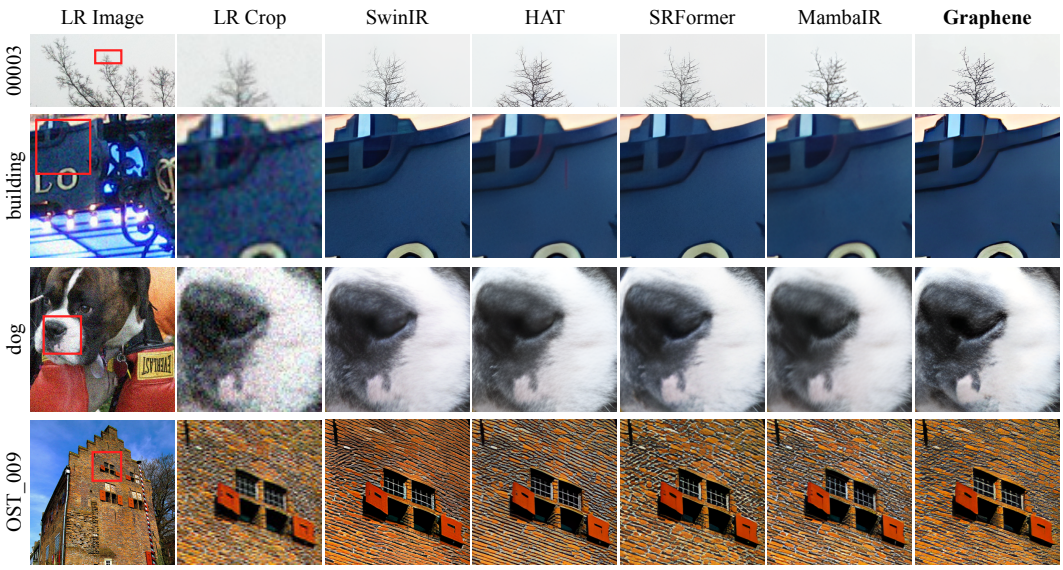

Figure 6: Qualitative $\times 4$ real-world SR on representative RealSRSet+5images examples. Graphene reconstructs fine details and textures in the displayed regions.

Limitations. The IRB requires materialized pre-softmax attention logits, limiting compatibility with FlashAttention-style kernels (Dao et al., 2022). Nevertheless, Graphene achieves lower inference latency than several state-of-the-art SR methods, demonstrating practical efficiency.

5 Conclusion

This paper introduces *Graphene*, a lightweight SR architecture combining Ising-model-inspired local refinement with stage-history reuse. The IRB uses signed, window-local couplings and mean-field updates to gate residual corrections, while HistAttn reuses earlier features at each token position without additional spatial pairwise attention. Graphene achieves state-of-the-art reconstruction quality with low parameter counts, FLOPs, latency, and memory usage.

AI USE STATEMENT

Generative AI tools were used to assist with writing and checking formatting compliance. The authors reviewed all AI-assisted changes and take responsibility for the final content of this work, including text, claims, and artifacts produced with the aid of generative AI.

ETHICS STATEMENT

This work uses public image super-resolution datasets and standard benchmark protocols and does not involve human subjects or private data. The authors have read and commit to the ICLR Code of Ethics.

REPRODUCIBILITY STATEMENT

Section 3 specifies the refinement equations and full/light architecture sizes, and Section 4 states the benchmark and ablation protocols. Appendix A of the supplementary material provides mathematical derivations and the scope of the companion Lean development; Appendix B gives architecture and evaluation details; Appendix C provides extended ablations and the paired-continuation protocol; Appendix D describes feature and attribution measurements; and Appendix E contains the full comparison tables. If the paper is accepted, the official code will be released by the camera-ready deadline.

REFERENCES

Marco Bevilacqua, Aline Roumy, Christine Guillemot, and Marie line Alberi Morel. Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. In *Proceedings of the British Machine Vision Conference*, pp. 135.1–135.10. BMVA Press, 2012. ISBN 1-901725-46-4. doi: 10.5244/C.26.135.

Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating More Pixels in Image Super-Resolution Transformer. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 22367–22377, 2023. doi: 10.1109/CVPR52729.2023.02142.

Zheng Chen, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Cross Aggregation Transformer for Image Restoration. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022.

Tri Dao, Daniel Y. Fu, Stefano Ermon, Atri Rudra, and Christopher Ré. FlashAttention: Fast and Memory-Efficient Exact Attention with IO-Awareness. In *Advances in Neural Information Processing Systems (NeurIPS)*, 2022. doi: 10.48550/arXiv.2205.14135.

Leonardo de Moura and Sebastian Ullrich. The Lean 4 Theorem Prover and Programming Language. In *Automated Deduction – CADE 28*, volume 12699 of *Lecture Notes in Computer Science*, pp. 625–635. Springer, 2021. doi: 10.1007/978-3-030-79876-5_37.

Albert Gu and Tri Dao. Mamba: Linear-Time Sequence Modeling with Selective State Spaces. In *First Conference on Language Modeling (COLM)*, 2024. doi: 10.48550/arXiv.2312.00752.

Jinjin Gu and Chao Dong. Interpreting Super-Resolution Networks with Local Attribution Maps. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 9195–9204, 2021. doi: 10.1109/CVPR46437.2021.00908.

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 (ECCV)*, pp. 222–241. Springer, 2024. doi: 10.1007/978-3-031-72649-1_13.

Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. MambaIRv2: Attentive State Space Restoration. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 28124–28133, 2025. doi: 10.1109/CVPR52734.2025.02619.

Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single Image Super-resolution from Transformed Self-Exemplars. In *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5197–5206, 2015. doi: 10.1109/CVPR.2015.7299156.

Keller Jordan, Yuchen Jin, Vlado Boza, Jiacheng You, Franz Cesista, Laker Newhouse, and Jeremy Bernstein. Muon: An Optimizer for Hidden Layers in Neural Networks, 2024.

Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. In *International Conference on Learning Representations (ICLR)*, 2015.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image Restoration Using Swin Transformer. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1833–1844, 2021. doi: 10.1109/ICCVW54120.2021.00210.

Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced Deep Residual Networks for Single Image Super-Resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1132–1140, 2017. doi: 10.1109/CVPRW.2017.151.

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 *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9992–10002, 2021. doi: 10.1109/ICCV48922.2021.00986.

Wei Long, Xingyu Zhou, Leheng Zhang, and Shuhang Gu. Progressive Focused Transformer for Single Image Super-Resolution. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2279–2288, 2025.

D. Martin, C. Fowlkes, D. Tal, and J. Malik. A Database of Human Segmented Natural Images and its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision (ICCV)*, volume 2, pp. 416–423 vol.2, 2001. doi: 10.1109/ICCV.2001.937655.

Yusuke Matsui, Kota Ito, Yuji Aramaki, Azuma Fujimoto, Toru Ogawa, Toshihiko Yamasaki, and Kiyoharu Aizawa. Sketch-based Manga Retrieval using Manga109 Dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017.

Chunyu Meng, Wei Long, and Shuhang Gu. From Local Windows to Adaptive Candidates via Individualized Exploratory: Rethinking Attention for Image Super-Resolution. In *European Conference on Computer Vision (ECCV)*, 2026.

Paulius Micikevicius, Sharan Narang, Jonah Alben, Gregory Diamos, Erich Elsen, David Garcia, Boris Ginsburg, Michael Houston, Oleksii Kuchaiev, Ganesh Venkatesh, and Hao Wu. Mixed Precision Training. In *International Conference on Learning Representations (ICLR)*, 2018.

Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P. Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. doi: 10.1109/CVPR.2016.207.

Cao Thien Tan, Phan Thi Thu Trang, Do Nghiem Duc, Ho Ngoc Anh, Hanyang Zhuang, and Nguyen Duc Dung. UCAN: Unified Convolutional Attention Network for Expansive Receptive Fields in Lightweight Super-Resolution. In *2026 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23409–23418, 2026.

Yuchuan Tian, Hanting Chen, Chao Xu, and Yunhe Wang. Image Processing GNN: Breaking Rigidity in Super-Resolution. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 24108–24117, 2024.

Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee, Xintao Wang, Yapeng Tian, Ke Yu, Yulun Zhang, Shixiang Wu, Chao Dong, Liang Lin, Yu Qiao, Chen Change Loy, Woong Bae, Jaejun Yoo, Yoseob Han, Jong Chul Ye, Jae-Seok Choi, Munchurl Kim, Yuchen Fan, Jiahui Yu, Wei Han,

Ding Liu, Haichao Yu, Zhangyang Wang, Honghui Shi, Xinchao Wang, Thomas S. Huang, Yunjin Chen, Kai Zhang, Wangmeng Zuo, Zhimin Tang, Linkai Luo, Shaohui Li, Min Fu, Lei Cao, Wen Heng, Giang Bui, Truc Le, Ye Duan, Dacheng Tao, Ruxin Wang, Xu Lin, Jianxin Pang, Jinchang Xu, Yu Zhao, Xiangyu Xu, Jinshan Pan, Deqing Sun, Yujin Zhang, Xibin Song, Yuchao Dai, Xueying Qin, Xuan-Phung Huynh, Tiantong Guo, Hojjat Seyed Mousavi, Tiep Huu Vu, Vishal Monga, Cristovao Cruz, Karen Egiazarian, Vladimir Katkovnik, Rakesh Mehta, Arnav Kumar Jain, Abhinav Agarwalla, Ch V. Sai Praveen, Ruofan Zhou, Hongdiao Wen, Che Zhu, Zhiqiang Xia, Zhengtao Wang, and Qi Guo. NTIRE 2017 Challenge on Single Image Super-Resolution: Methods and Results. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1110–1121, 2017. doi: 10.1109/CVPRW.2017.149.

Zhou Wang, A.C. Bovik, H.R. Sheikh, and E.P. Simoncelli. Image Quality Assessment: From Error Visibility to Structural Similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004. doi: 10.1109/TIP.2003.819861.

Roman Zeyde, Michael Elad, and Matan Protter. On Single Image Scale-Up Using Sparse-Representations. In *International Conference on Curves and Surfaces*, pp. 711–730. Springer, 2010.

Jiale Zhang, Yulun Zhang, Jinjin Gu, Yongbing Zhang, Linghe Kong, and Xin Yuan. Accurate Image Restoration with Attention Retractable Transformer. In *International Conference on Learning Representations (ICLR)*, 2023.

Leheng Zhang, Yawei Li, Xingyu Zhou, Xiaorui Zhao, and Shuhang Gu. Transcending the Limit of Local Window: Advanced Super-Resolution Transformer with Adaptive Token Dictionary. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2856–2865, 2024.

Yupeng Zhou, Zhen Li, Chun-Le Guo, Song Bai, Ming-Ming Cheng, and Qibin Hou. SRFormer: Permuted Self-Attention for Single Image Super-Resolution. In *2023 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 12734–12745, 2023. doi: 10.1109/ICCV51070.2023.01174.