NTIRE 2025 Challenge on Image Super-Resolution ($\times$4): Methods and Results

Zheng Chen[†]   Kai Liu[†]   Jue Gong[†]   Jingkai Wang[†]   Lei Sun[†]   Zongwei Wu[†]

Radu Timofte[†]   Yulun Zhang[†*]   Xiangyu Kong   Xiaoxuan Yu   Hyunhee Park

Suejin Han   Hakjae Jeon   Dafeng Zhang   Hyung-Ju Chun   Donghun Ryou

Inju Ha   Bohyung Han   Lu Zhao   Yuyi Zhang   Pengyu Yan   Jiawei Hu

Pengwei Liu   Fengjun Guo   Hongyuan Yu   Pufan Xu   Zhijuan Huang

Shuyuan Cui   Peng Guo   Jiahui Liu   Dongkai Zhang   Heng Zhang   Huiyuan Fu

Huadong Ma   Yanhui Guo   Sisi Tian   Xin Liu   Jinwen Liang   Jie Liu   Jie Tang

Gangshan Wu   Zeyu Xiao   Zhuoyuan Li   Yinxiang Zhang   Wenxuan Cai

Vijayalaxmi Ashok Aralikatti   Nikhil Akalwadi   G Gyaneshwar Rao   Chaitra Desai

Ramesh Ashok Tabib   Uma Mudenagudi   Marcos V. Conde   Alejandro Merino

Bruno Longarela   Javier Abad   Weijun Yuan   Zhan Li   Zhanglu Chen

Boyang Yao   Aagam Jain   Milan Kumar Singh   Ankit Kumar   Shubh Kawa

Divyavardhan Singh   Anjali Sarvaiya   Kishor Upla   Raghavendra Ramachandra

Chia-Ming Lee   Yu-Fan Lin   Chih-Chung Hsu   Risheek V Hiremath

Yashaswini Palani   Yuxuan Jiang   Qiang Zhu   Siyue Teng   Fan Zhang

Shuyuan Zhu   Bing Zeng   David Bull   Jingwei Liao   Yuqing Yang   Wenda Shao

Junyi Zhao   Qisheng Xu   Kele Xu   Sunder Ali Khowaja   Ik Hyun Lee

Snehal Singh Tomar   Rajarshi Ray   Klaus Mueller   Sachin Chaudhary

Surya Vashisth   Akshay Dudhane   Praful Hambarde   Satya Naryan Tazi

Prashant Patil   Santosh Kumar Vipparthi   Subrahmanyam Murala   Bilel Benjdira

Anas M. Ali   Wadii Boulila   Zahra Moammeri   Ahmad Mahmoudi-Aznaveh

Ali Karbasi   Hossein Motamednia   Liangyan Li   Guanhua Zhao   Kevin Le

Yimo Ning   Haoxuan Huang   Jun Chen

Abstract

This paper presents the NTIRE 2025 image super-resolution ($\times$4) challenge, one of the associated competitions of the 10th NTIRE Workshop at CVPR 2025. The challenge aims to recover high-resolution (HR) images from low-resolution (LR) counterparts generated through bicubic downsampling with a $\times$4 scaling factor. The objective is to develop effective network designs or solutions that achieve state-of-the-art SR performance. To reflect the dual objectives of image SR research, the challenge includes two sub-tracks:

*(1) **a restoration track**, emphasizes pixel-wise accuracy and ranks submissions based on PSNR; (2) **a perceptual track**, focuses on visual realism and ranks results by a perceptual score. A total of 286 participants registered for the competition, with 25 teams submitting valid entries. This report summarizes the challenge design, datasets, evaluation protocol, the main results, and methods of each team. The challenge serves as a benchmark to advance the state of the art and foster progress in image SR.*

1. Introduction

Single image super-resolution (SR) refers to the process of reconstructing a high-resolution (HR) image from its low-resolution (LR) counterpart, which has undergone an information-losing degradation process. As a fundamental problem in computer vision, SR underpins numerous practical applications, including video surveillance, digital pathology, and satellite imagery. Among the various SR

formulations, the classical image SR task [50, 69] is one of the most commonly adopted benchmarks.

In this setting, the LR image is typically generated through a predefined downsampling operation, most often bicubic interpolation. Such degradation removes significant high-frequency content, rendering the task of detail recovery an inherently ill-posed inverse problem. The goal is to recover the missing high-frequency components using learned priors from external data. Owing to its well-defined setting, the classical SR task is a standard benchmark, enabling fair comparison across methods. Moreover, models trained under this framework often generalize well to more complex degradations through transfer learning.

Early image super-resolution (SR) methods rely on interpolation or reconstruction-based algorithms [47, 64, 70]. However, these approaches often produce over-smoothed results that lack fine textures and realistic details. With the advent of deep learning, convolutional neural networks (CNNs) have become the dominant paradigm in SR [11, 24, 66]. Beginning with the pioneering work of SRCNN [11], later models incorporate deeper architectures [56, 66], residual learning [67], and attention mechanisms [10] to improve reconstruction accuracy. Meanwhile, Transformer-based architectures further advance SR by modeling long-range dependencies through self-attention [4, 40, 53]. Besides, recent state-space models (*i.e.*, Mamba) offer improved scalability and efficient sequence modeling [16, 18, 39]. These approaches are typically trained using pixel-wise losses such as mean squared error, and primarily focus on enhancing restoration quality to maximize fidelity.

In recent years, the field has increasingly shifted toward models that emphasize perceptual realism. Generative approaches, such as generative adversarial networks (GANs) [15, 27, 63], are introduced to produce more perceptually appealing outputs. More recently, diffusion models have emerged as a powerful alternative for generating realistic textures [21, 29, 46, 57, 58]. These models progressively transform random noise into high-resolution images through a denoising process conditioned on the LR input. The diffusion-based methods show strong potential in improving perception quality, which better aligns with human visual preferences. Overall, the development of SR methods reflects a dual objective: achieving high restoration quality for faithful reconstruction, and high perception quality for producing visually pleasing, realistic results, both of which are essential for practical application.

Collaborating with the 10th edition of the NTIRE workshop at CVPR 2025, we organize the NTIRE 2025 Challenge on example-based single image super-resolution (×4). This challenge focuses on recovering high-resolution details from a single low-resolution input, following the classical bicubic degradation setting. The goal is to advance the development of effective SR models and provide a uni-fied benchmark for comparison.

To comprehensively evaluate performance, the challenge includes two tracks: one targeting restoration quality, which focuses on pixel-wise accuracy, and the other targeting perception quality, which emphasizes the perceptual realism of the output. This dual-track design reflects the evolving goals of image SR research and encourages solutions that balance fidelity with perceptual quality.

This challenge is one of the NTIRE 2025 Workshop associated challenges on: ambient lighting normalization [52], reflection removal in the wild [60], shadow removal [51], event-based image deblurring [48], image denoising [49], XGC quality assessment [37], UGC video enhancement [45], night photography rendering [13], image super-resolution (x4) [6], real-world face restoration [7], efficient super-resolution [44], HR depth estimation [61], efficient burst HDR and restoration [28], cross-domain few-shot object detection [14], short-form UGC video quality assessment and enhancement [31, 32], text to image generation model quality assessment [20], day and night raindrop removal for dual-focused images [30], video quality assessment for video conferencing [23], low light image enhancement [38], light field super-resolution [55], restore any image model (RAIM) in the wild [35], raw restoration and super-resolution [8] and raw reconstruction from RGB on smartphones [9].

2. NTIRE 2025 Image Super-Resolution (×4)

The NTIRE 2025 Image Super-Resolution (×4) Challenge, which is one of the associated challenges of NTIRE 2025, has two primary objectives. Firstly, it intends to offer a thorough overview of the latest advancements and emerging tendencies within the field of image super-resolution (SR). Secondly, it functions as a platform that enables both academic researchers and industrial professionals to converge and investigate possible collaborative opportunities. The following part delves into the specific aspects.

2.1. Dataset

The challenge provides three official datasets, including DIV2K [50], Flickr2K [36], and LSDIR [33]. Additionally, supplementary data are allowed to be used. For the challenge, the LR-HR image pairs are constructed with the HQ images via bicubic interpolation with a ×4 downsampling.

DIV2K The DIV2K dataset contains 1,000 2K images, which are divided into three parts: 800/100/100 for training, validation, and testing, respectively. To ensure fairness, participants do not have access to the high-resolution (HR) images from the DIV2K validation set except during the testing phase. Furthermore, the HR test images remain hidden throughout the entire challenge.

Flickr2K The Flickr2K dataset has 2,650 2K images, covering a wide range of quality levels and contents. For this challenge, all images in the Flickr2K dataset are available.
LSDIR The LSDIR dataset comprises 86,991 images from the Flickr platform. It is divided into three parts: 84,991/1000/1000 for training, validation, and testing.

2.2. Track and Competition

This year, the competition features two tracks: the Restoration track and the Perceptual track.

Restoration Track. Consistent with last year's challenge [5], all teams are ranked based on the PSNR of their enhanced HR images compared to the GT HR images from the DIV2K testing dataset.

Perceptual Track. This year, a perceptual track has been introduced. For this track, seven widely used IQA metrics are collected to evaluate the restored results thoroughly. These metrics include LPIPS, DISTS, CLIP-IQA, MANIQA, MUSIQ, and NIQE. The teams are ultimately ranked based on the overall perceptual score:

$$\text{Score} = (1 - \text{LPIPS}) + (1 - \text{DISTS}) + \text{CLIP-IQA}$$
$$+\text{MANIQA} + \frac{\text{MUSIQ}}{100} + \max\left(0, \frac{10 - \text{NIQE}}{10}\right). \quad (1)$$

Challenge Phases. *(1) Development and Validation Phase:* During this phase, participants have access to two datasets: (a) 800 pairs of LR/HR training images and 100 LR validation images from the DIV2K dataset, and (b) 84,991 LR/HR training image pairs and 1,000 LR validation images from the LSDIR dataset. Participants are also permitted to use additional data for training purposes. The restored HR images generated by their models can be submitted to the Codalab server for evaluation based on eight performance metrics, with immediate feedback provided.

(2) Testing Phase: In the final phase, participants receive 100 LR test images, but the corresponding HR ground truth images are not provided. They are required to upload their SR outputs to the Codalab server and submit their code and a detailed report to the organizers via email. After the challenge ends, the organizers will validate the code and send the results to the participants.

Evaluation Protocol. The evaluation process involves eight standard metrics: PSNR, SSIM, LPIPS, DISTS, NIQE, ManIQA, MUSIQ, and CLIP-IQA. For the evaluation, a 4-pixel border around each image is excluded, and the calculations are performed on the Y channel of the YCbCr color space. The results are primarily based on the submissions made to the Codalab server. Code submitted by participants is used for reproduction and verification, with small discrepancies in precision being considered acceptable. A script for calculating these metrics can be found at https://github.com/zhengchen1999/NTIRE2025_ImageSR_x4, and the repository also includes the source code and pre-trained models.

3. Challenge Results

The challenge includes two sub-tracks: Track 1 (restoration quality) and Track 2 (perception quality). The results and rankings of all participating teams are provided in Tab. 1.

Track 1 (Restoration Quality). The SamsungAICamera team achieves the top performance (33.46 dB). Notably, two teams surpass last year's best PSNR score (31.94 dB), and ten teams obtain results above 31.00 dB, highlighting a clear improvement in reconstruction accuracy.

Track 2 (Perception Quality). The SNUCV team ranks first with the highest perceptual score (4.3472). Two teams achieve a score above 4.0, and seven teams exceed 3.6, indicating advancements in the perceptual quality.

Further details on the evaluation protocol of the two tracks are provided in Sec. 2.2. The team order in Sec. 4 follows the same order as presented in Tab. 1, with the top 5 teams and their method details highlighted here. Due to space limitations, the remaining teams are listed in Sec. A of the supplementary materials. Team member information can also be found in Sec. B of the supplementary materials.

3.1. Architectures and main ideas

Throughout the challenge, various innovative techniques were introduced to boost the SR performance. Here, the team members summarize some of the principal concepts.

1. **Transformer-based architectures remain a mainstream approach.** Transformer-based methods, such as HAT [3], SwinIR [34], and DAT [4], continue to deliver strong reconstruction results by capturing long-range dependencies. Many teams utilized pre-trained Transformer models and fine-tuned them with hybrid attention and self-attention mechanisms. The BBox team, for example, used Transformer models to capture both local and global information, improving reconstruction quality through model ensembles.

2. **Integrating Mamba architectures for improved global context modeling.** Recognizing the importance of capturing extensive contextual information, some participants employed the Mamba architecture to better model long-range spatial dependencies and global context, leading to significant performance improvements in super-resolution tasks. The XiaomiMM team, for example, integrated Mamba into the HAT model, adapting branches based on image statistics and combining multiple Mamba-enhanced networks.

3. **Dynamic fusion of multiple network architectures boosts performance.** Several teams utilized dynamic model fusion strategies, combining outputs from Transformer-based models and CNNs. By leveraging the strengths of both architectures, these methods captured local textures and global structural information, significantly improving reconstruction quality. The winning

Team Name	Rank Track 1	Rank Track 2	PSNR Track 1	SSIM	LPIPS	DISTS	NIQE	ManIQA	MUSIQ	CLIP-IQA	Perceptual Score Track 2
SamsungAICamera	1	3	33.46	0.9124	0.1681	0.0929	4.4089	0.3566	62.1058	0.4934	3.7692
SNUCV	24	1	22.53	0.6326	0.2113	0.1082	2.9635	0.4939	71.4919	0.7543	4.3472
BBox	2	4	31.97	0.8793	0.2082	0.1140	5.0643	0.3656	61.2975	0.5206	3.6706
MicroSR	22	2	26.34	0.7594	0.2340	0.1261	3.7380	0.3552	64.2008	0.6317	3.8950
XiaomiMM	3	5	31.93	0.8775	0.2144	0.1186	5.2336	0.3684	60.8244	0.5152	3.6355
NJU_MCG	4	11	31.19	0.8661	0.2255	0.1233	5.3914	0.3649	59.9235	0.4984	3.5747
X-L	5	13	31.15	0.8653	0.2276	0.1227	5.3885	0.3597	59.5546	0.4990	3.5652
Endeavour	6	12	31.15	0.8653	0.2269	0.1229	5.3787	0.3610	59.6981	0.4994	3.5697
CidautAi	20	6	30.48	0.8564	0.2029	0.0894	4.7080	0.3384	59.3074	0.4472	3.6157
KLETech-CEVI	7	7	31.13	0.8649	0.2264	0.1217	5.3118	0.3661	60.0893	0.5087	3.5964
JNU620	8	10	31.12	0.8647	0.2271	0.1212	5.3608	0.3587	59.6759	0.5048	3.5758
ACVLAB	14	8	30.82	0.8635	0.2302	0.1210	5.2777	0.3642	59.9242	0.5071	3.5916
CV_SVNIT	9	9	31.11	0.8647	0.2259	0.1224	5.3500	0.3619	59.8108	0.5010	3.5778
HyperPix	10	14	31.03	0.8633	0.2286	0.1238	5.3826	0.3593	59.5252	0.4982	3.5621
BVIVSR	11	15	30.99	0.8630	0.2288	0.1234	5.4281	0.3602	59.6739	0.4969	3.5588
AdaDAT	12	17	30.91	0.8605	0.2329	0.1256	5.4721	0.3584	59.1563	0.4968	3.5410
Junyi	13	19	30.91	0.8605	0.2364	0.1278	5.5369	0.3542	58.7826	0.4926	3.5167
ML_SVNIT	15	20	30.82	0.8589	0.2357	0.1268	5.4299	0.3497	58.4427	0.4831	3.5117
SAK_DCU	16	16	30.80	0.8595	0.2328	0.1268	5.3981	0.3551	59.2546	0.4937	3.5419
VAI-GM	17	18	30.76	0.8579	0.2366	0.1279	5.4573	0.3496	58.8017	0.4919	3.5193
Quantum Res	18	22	30.52	0.8523	0.2482	0.1330	5.6358	0.3386	56.9043	0.4754	3.4381
PSU	19	21	30.50	0.8528	0.2476	0.1296	5.5450	0.3404	57.6043	0.4801	3.4649
IVPLAB-sbu	21	23	26.74	0.8490	0.4512	0.1992	5.4675	0.3384	56.2507	0.4890	3.1927
MCMIR	23	24	24.53	0.7220	0.2624	0.1460	6.1282	0.2773	42.8159	0.3456	3.0298
Aimanga	N/A	N/A	22.44	0.6553	0.2965	0.1396	2.8260	0.4752	68.8273	0.7333	4.1781
IPCV_Team	N/A	N/A	31.01	0.8643	0.2255	0.1228	5.3553	0.3664	60.2005	0.5033	3.5878

Table 1. **Results of NTIRE 2025 Image Super-Resolution Challenge.** PSNR and the other seven perceptual metric scores are measured on the DIV2K testing (100) dataset. For the restoration track, the team rankings are based primarily on PSNR. As for the perceptual track, the team ranking is based on the perceptual score, a weighted average of seven perceptual metrics. The ranking mainly depends on the higher ranking among the two tracks, and secondly on the average value of the rankings in the two tracks. Team order in Sec. 4 follows the same sequence as this table. Teams (*i.e.*, Aimanga and IPCV_Team) that submitted late or failed to submit their code are not ranked.

team, SamsungAICamera, for instance, combined a pre-trained HAT Transformer with the CNN-based NAFNet, achieving top performance in the restoration track.

4. **Incorporating frequency-domain loss functions for enhanced detail reconstruction.** To preserve fine details and textures, participants employed frequency-domain loss functions, such as Stationary Wavelet Transform (SWT) loss, to improve high-frequency detail recovery and reduce artifacts. The KLETech-CEVI team, for example, used wavelet-based loss in HAT framework, applying Symlet filters to maintain texture sharpness and prevent chroma distortion.

5. **Advanced multi-stage training strategies to progressively refine model performance.** Participants adopted multi-stage training methods, gradually increasing patch sizes during training, which improved reconstruction quality and model stability. For example, the BBox team used a three-stage training strategy, adjusting patch sizes and loss functions across stages, followed by CLIP-

based semantic filtering and fine-tuning, ultimately enhancing both fidelity and perceptual quality.

3.2. Participants

This year, the image SR challenge attracted 286 registered participants, with 25 teams submitting valid entries. In comparison to last year's challenge [5], there was an increase in the number of valid submissions (from 20 to 24) and a significant improvement in the results. These submissions have set a new benchmark for the state-of-the-art in image super-resolution ($\times 4$).

3.3. Fairness

A set of rules has been established to ensure the fairness of the competition. **(1)** The use of DIV2K test HR images for training is strictly prohibited, although the DIV2K test LR images are available. **(2)** Participants are allowed to train using additional datasets, such as the Flickr2K and LSDIR datasets. **(3)** The use of data augmentation techniques during both training and testing is considered a fair practice.
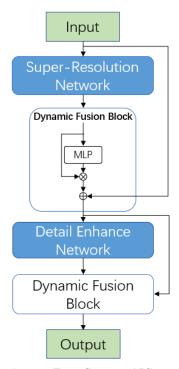
Figure 1. **Team SamsungAICamera**

Figure 2. **Team SamsungAICamera**

3.4. Conclusions

The insights gained from analyzing the results of the NTIRE 2025 image super-resolution (SR) challenge are summarized as follows:

1. The integration of Transformer and CNN hybrid architectures has demonstrated exceptional performance in modeling global context and reconstructing local details, achieving a balanced approach.

2. Mamba-based state space models have been widely adopted, showcasing their scalability in computational efficiency and modeling capabilities, offering a robust alternative to traditional self-attention mechanisms.

3. Advanced training strategies, including multi-stage pipelines, progressive patch training, and CLIP-based semantic filtering, have boosted model generalization and robustness. Frequency-domain losses further enhanced high-frequency texture restoration.

4. The innovative use of generative priors, particularly pre-trained diffusion models combined with CLIP-based reference-free perceptual losses, has achieved superior perceptual quality with minimal training.

4. Challenge Methods and Teams

4.1. SamsungAICamera

Description. The proposed solution is shown in Fig. 1. Image super-resolution (SR) is a critical task in computer vision, aiming to recon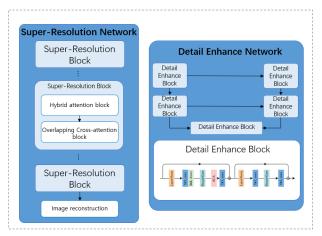struct high-quality images from their low-resolution versions. While significant advances have been made in recent years, existing methods often face challenges in effectively capturing both global context and fine local details simultaneously.

To address these limitations, SamsungAICamera's approach combines two powerful networks: the transformer-based network HAT [3] and the convolution-based network NAFnet [2]. HAT try to leverage global and local information more effective by combine channel attention and overlapping cross-attention, but still struggle for capturing fine-grained local features. This team design the Super-Resolution Network and Detail Enhance Network refer to these two method as shown in Fig. 2. They achieve superior performance in image super-resolution, by dynamically fuse the feature extracted by transformer-based network and convolution-based network This design not only addresses the limitations of each individual component but also opens new possibilities for hybrid architectures in image super-resolution tasks.

Implementation Details. *Datasets.*The team used three datasets in total: the DIV2K dataset, the LSDIR dataset, and a self-collected custom dataset consisting of 2 million images. The specific ways in which the team utilized these training sets across different training phases will be detailed in the training details section.

Trainning strategy. The model training consists of two stages. In the first stage, The team pre-train the entire network using the custom dataset of 2 million images and LS-DIR datasets, with an initial learning rate of $1e^{-4}$ and a training time of approximately 360 hours. In the second stage, they fine-tune the network module using the DIV2K datasets, with an initial learning rate of $1e^{-5}$ and a training duration of about 120 hours, which enhanced the model's ability to restore details.

They used data augmentation methods such as channel shuffle and mixing and perform progressive learning where the network is trained on different image patch sizes gradually enlarged from 320 to 448 and 768. As the patch size
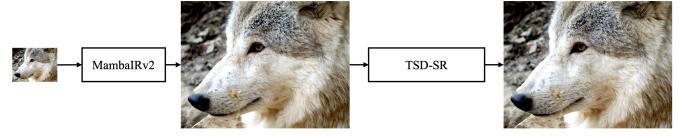
Figure 3. **Team SNUCV**

increases, the performance can gradually improve. As for training loss, they train the model by alternately iterating L1 loss, L2 loss, and Stationary Wavelet Transform(SWT) loss [26]. They found that adding SWT loss during training helps the model escape from local optima. Their model was trained on an A100 80G GPU.

4.2. SNUCV

Description. As illustrated in Fig. 3, the SNUCV team first employs a classic super-resolution (SR) model to perform $4\times$ upsampling of the input image. Subsequently, a one-step diffusion model is applied to further enhance the super-resolution process. For the $4\times$ upsampling model, the team adopts MambaIRv2 [17], while the diffusion model utilizes the TSD-SR [12] architecture. To fully leverage the generative prior of the diffusion model, they refrain from additional training and directly use its pretrained weights, focusing instead on fine-tuning the upsampler.

Let the diffusion model be M, the upsampler be U, the input image be I, and the ground-truth image be I_{gt}. To ensure that both the output $M(U(I))$ and the intermediate output $U(I)$ accurately resemble the ground-truth I_{gt}, the team defines a loss function composed of the LPIPS loss [65], denoted $\mathcal{L}_{\text{LPIPS}}$, and the L1 loss, denoted $\mathcal{L}_1$:

$$\mathcal{L}_{\text{accuracy}} = \mathcal{L}_1(U(I), I_{gt}) \\ + \alpha \cdot \mathcal{L}_{\text{LPIPS}}(M(U(I)), I_{gt}), \quad (2)$$

where $\mathcal{L}_{\text{LPIPS}}$ captures perceptual similarity between the final output and the ground-truth, and $\mathcal{L}_1$ enforces pixel-wise accuracy.

To further enhance perceptual quality, the team introduces a no-reference CLIP-based image quality assessment (IQA) loss [43, 54], denoted $\mathcal{L}_{\text{CLIP}}$. This loss utilizes the CLIP model's text encoder $\text{CLIP}_{\text{text}}$ and image encoder $\text{CLIP}_{\text{image}}$. Two text prompts are defined: "Good photo" and "Bad photo", and their embeddings are computed and normalized as:

$$T_{\text{pos}} = \text{norm}(\text{CLIP}_{\text{text}}(\text{"Good photo"})) \quad (3)$$
$$T_{\text{neg}} = \text{norm}(\text{CLIP}_{\text{text}}(\text{"Bad photo"})) \quad (4)$$

The generated image's feature representation is:
$$F_{\text{pred}} = \text{norm}(\text{CLIP}_{\text{image}}(M(U(I)))) \quad (5)$$

Then, the similarities between the image features and the text embeddings are computed:

$$S_{\text{pos}} = F_{\text{pred}} \cdot T_{\text{pos}}^\top \quad (6)$$
$$S_{\text{neg}} = F_{\text{pred}} \cdot T_{\text{neg}}^\top \quad (7)$$

The CLIP loss is defined to increase similarity with "Good photo" and reduce similarity with "Bad photo":

$$\mathcal{L}_{\text{CLIP}} = 1 - S_{\text{pos}} + S_{\text{neg}} \quad (8)$$

The total training loss is formulated as:
$$\mathcal{L} = \mathcal{L}_{\text{accuracy}} + \beta \cdot \mathcal{L}_{\text{CLIP}} \quad (9)$$
where they set $\alpha = 0.1$ and $\beta = 0.5$.

Implementation Details. When training the upsampler MambaIRv2, the team utilizes the AdamW [41] optimizer with hyperparameters $\beta_1 = 0.9$, $\beta_2 = 0.999$, and a learning rate of 1×10^{-5}. The ground-truth images are cropped into patches of size 256×256. Training is conducted over 100,000 iterations with a batch size of 4. The training dataset is DIV2K [50] and LSDIR [33].

4.3. BBox

Description. The solution proposed by the BBox is illustrated in Fig. 4. Transformer-based models have consistently demonstrated remarkable performance in the field of super-resolution, as exemplified by methods such as HAT [3], DAT [4], and SwinIR [34]. Recently, numerous studies have shown that Mamba architectures can also achieve impressive results in this domain, as evidenced by models such as MambaIR [18], MambaIRv2 [17], and S^3Mamba [59]. Notably, prior research indicates that Transformer-based models excel at modeling sequential relationships, while Mamba-based approaches are particularly adept at capturing long-range contextual information [5]. Both capabilities are crucial for pixel-dense super-resolution tasks, which require models to simultaneously capture spatial relationships between individual pixels and model their long-range contextual dependencies. Inspired by these complementary strengths, the team members adopt HAT-L and MambaIRv2-L as the foundation models, training them independently and employing a model ensemble strategy to effectively leverage their advantages, thereby achieving optimal performance.
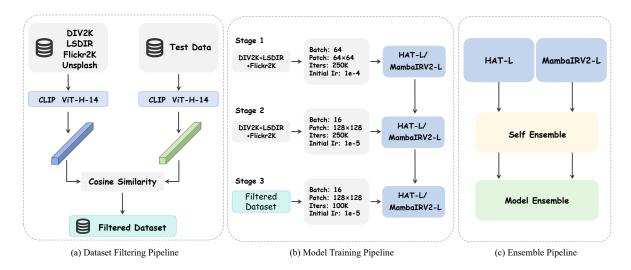
(a) Dataset Filtering Pipeline   (b) Model Training Pipeline   (c) Ensemble Pipeline
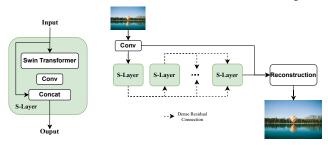
Figure 4. **Team BBox.**

Figure 5. **Team MicroSR**

Implementation Details. The training dataset comprises DIV2K, LSDIR, Flickr2K, and selected Unsplash [1] datasets. To augment the training data, the team members implement random flip and rotation strategies. For the HAT model, the team members initialize the network using pretrained weights from the HAT-L model, which was previously trained on the ImageNet dataset. In contrast, the team members train the MambaIRv2-L model from scratch.

Specifically, the training follows a three-stage progressive strategy with a multi-scale loss that computes losses at x2, x3, and x4 resolutions to enhance performance across different scales. (1) In the first stage, the team members train on the DIV2K, LSDIR, and Flickr2K datasets. The team members set the patch size to 64 and batch size to 64, conducting 250K iterations with an initial learning rate of 1×10^{-4}. During this stage, the team members minimize the L1 loss using the Adam optimizer [25]. (2) In the second stage, while maintaining the same training datasets, the team members increase the patch size to 128 and reduce the batch size to 16. The team members also switch from L1 loss to MSE loss. This stage comprises 250K iterations with a reduced initial learning rate of 1×10^{-5}. (3) In the final stage, to achieve superior model performance, the team members implement semantic selection using CLIP [43] features. the team members filter out 2,000 similar images

from DIV2K, LSDIR, Flickr2K, and Unsplash datasets for further fine-tuning. During this stage, the learning rate is set to 1×10^{-5}, and the model is trained for an additional 100K iterations. The multi-step learning rate decay method is applied across all three training phases. For both MambaIRv2-L and HAT-L models, the team members repeat the aforementioned three training stages five times to ensure optimal convergence.

In the inference stage, the team members employ a self-ensemble strategy to enhance the performance of both HAT-L and MambaIRv2-L models. For MambaIRv2-L specifically, the team members utilize multiple sliding windows of varying sizes during inference and integrate these results. Finally, the team members implement a weighted fusion method [68] to generate outputs from the adaptive combination of HAT-L and MambaIRv2-L models.

4.4. MicroSR

Description. The MicroSR team proposes a super-resolution method based on the architectural principles of DRCT [22] and SwinFIR [62], while also building on the foundation established by SwinIR [34]. A novel S-layer is introduced with residual connections to improve the capture of global information more efficiently. As illustrated in Fig. 5, the architecture comprises a shallow feature extraction layer, 12 S-layers constructed with dense residual connections, and a final reconstruction layer.

Implementation Details. The proposed solution adopts a multi-stage greedy training policy. In the first stage, the model is trained with MSE loss for 10 epochs. In the second stage, an adversarial loss is introduced to fine-tune the model for better perceptual quality. During training, the learning rate decays progressively by a factor of 0.1 after each stage. Within each stage, the learning rate remains fixed, and the best-performing checkpoint is selected and used to initialize the next stage.
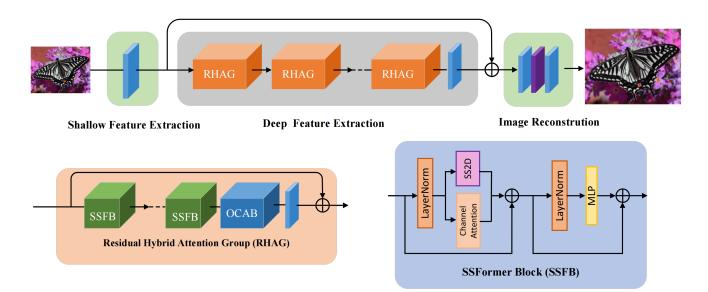
Figure 6. **Team XiaomiMM**

To further enhance performance, the team integrates a neural degradation algorithm [42] for data augmentation. The training data includes the DIV2K dataset [50], supplemented with real-world paired images [19], resulting in a more diverse dataset and improved robustness to unseen degradations. The team re-trains DRCT using the official training code on the augmented dataset. Training is initialized with the official pretrained model Real-DRCT-GAN_Finetuned. The final model contains 27.58M parameters, with 11.078G FLOPs. The average inference time per image is reported as 4.41 seconds.

4.5. XiaomiMM

Description. Based on the HAT model [3], the team members propose two variants to improve image super-resolution performance.

The first variant introduces a multi-branch structure, designed according to the statistical feature information (brightness, contrast, sharpness, noise, saturation) of the input image. Each of these five statistical features is divided into two categories, resulting in a ten-branch network structure. By utilizing fixed statistical information instead of allowing the network to learn these features autonomously, the team members introduce additional prior information and improve the model's generalization ability.

The second variant integrates the Mamba structure [39] into the HAT model, thereby fully leveraging the advantageous characteristics of Mamba. The detailed network structure is illustrated in Fig. 6.

Finally, the team members trained and fused the original HAT model and these two variants, achieving the best results in the experiments.

Implementation Details. The dataset utilized for training comprises DIV2K and LSDIR. During each training batch,

64 HR RGB patches measuring 256×256 are cropped and subjected to random flipping and rotation. The learning rate is initialized at 5×10^{-4} and is halved every 2×10^5 iterations. The network undergoes training for a total of 10^6 iterations, minimizing the L1 loss function using the Adam optimizer [25]. The team members repeated the aforementioned training settings four times after loading the trained weights. Subsequently, fine-tuning is executed using the L1 and L2 loss functions, with an initial learning rate of 1×10^{-5} for 5×10^5 iterations and an HR patch size of 512. The team members conducted fine-tuning on four models utilizing both L1 and L2 losses and employed batch sizes of 64 and 128. Finally, the team members integrated these five models to obtain the ultimate model and also used the K-means Based Fusion Strategy [5] to achieve better results.

5. Methods of the Remaining Teams

All the teams presented innovative ideas and thorough experiments for the competition. However, due to space limitations, a more in-depth discussion can be found in Sec. A of the supplementary materials, which contains detailed descriptions of the methods and implementation details for the remaining teams that participated in the challenge. While these teams have not been discussed in the main report, their approaches are still highlighted, offering insight into their unique strategies and technical implementations.

References

[1] Unsplash Full Dataset 1.2.2. https://unsplash.com/data. Accessed: 2024-03-21. 7

[2] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European conference on computer vision*, pages 17–33. Springer, 2022. 5

[3] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *CVPR*, 2023. 3, 5, 6, 8

[4] Zheng Chen, Yulun Zhang, Jinjin Gu, Linghe Kong, Xiaokang Yang, and Fisher Yu. Dual aggregation transformer for image super-resolution. In *ICCV*, 2023. 2, 3, 6

[5] Zheng Chen, Zongwei WU, Eduard Sebastian Zamfir, Kai Zhang, Yulun Zhang, Radu Timofte, Xiaokang Yang, et al. NTIRE 2024 challenge on image super-resolution (×4): Methods and results. In *CVPRW*, 2024. 3, 4, 6, 8

[6] Zheng Chen, Kai Liu, Jue Gong, Jingkai Wang, Lei Sun, Zongwei Wu, Radu Timofte, Yulun Zhang, et al. NTIRE 2025 challenge on image super-resolution (×4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[7] Zheng Chen, Jingkai Wang, Kai Liu, Jue Gong, Lei Sun, Zongwei Wu, Radu Timofte, Yulun Zhang, et al. NTIRE 2025 challenge on real-world face restoration: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[8] Marcos Conde, Radu Timofte, et al. NTIRE 2025 challenge on raw image restoration and super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[9] Marcos Conde, Radu Timofte, et al. Raw image reconstruction from RGB on smartphones. NTIRE 2025 challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[10] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In *CVPR*, 2019. 2

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *ECCV*, 2014. 2

[12] Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. *arXiv preprint arXiv:2411.18263*, 2024. 6

[13] Egor Ershov, Sergey Korchagin, Alexei Khalin, Artyom Panshin, Arseniy Terekhin, Ekaterina Zaychenkova, Georgiy Lobarev, Vsevolod Plokhotnyuk, Denis Abramov, Elisey Zhdanov, Sofia Dorogova, Yasin Mamedov, Nikola Banic, Georgii Perevozchikov, Radu Timofte, et al. NTIRE 2025 challenge on night photography rendering. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[14] Yuqian Fu, Xingyu Qiu, Bin Ren Yanwei Fu, Radu Timofte, Nicu Sebe, Ming-Hsuan Yang, Luc Van Gool, et al. NTIRE 2025 challenge on cross-domain few-shot object detection: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[15] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. In *NeurIPS*, 2014. 2

[16] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023. 2

[17] Hang Guo, Yong Guo, Yaohua Zha, Yulun Zhang, Wenbo Li, Tao Dai, Shu-Tao Xia, and Yawei Li. Mambairv2: Attentive state space restoration. *arXiv preprint arXiv:2411.15269*, 2024. 6

[18] 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 *ECCV*, pages 222–241. Springer, 2025. 2, 6

[19] Yanhui Guo, Xiaolin Wu, and Xiao Shu. Data acquisition and preparation for dual-reference deep learning of image super-resolution. *TIP*, 2022. 8

[20] Shuhao Han, Haotian Fan, Fangyuan Kong, Wenjie Liao, Chunle Guo, Chongyi Li, Radu Timofte, et al. NTIRE 2025 challenge on text to image generation model quality assessment. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[21] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. In *NeurIPS*, 2020. 2

[22] Chih-Chung Hsu, Chia-Ming Lee, and Yi-Shiuan Chou. Drct: Saving image super-resolution away from information bottleneck. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6133–6142, 2024. 7

[23] Varun Jain, Zongwei Wu, Quan Zou, Louis Florentin, Henrik Turbell, Sandeep Siddhartha, Radu Timofte, et al. NTIRE 2025 challenge on video quality enhancement for video conferencing: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[24] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *CVPR*, 2016. 2

[25] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 7, 8

[26] Cansu Korkmaz and A Murat Tekalp. Training transformer models by wavelet losses improves quantitative and visual performance in single image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 6661–6670, 2024. 6

[27] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, and Wenzhe

Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *CVPR*, 2017. 2

[28] Sangmin Lee, Eunpil Park, Angel Canelo, Hyunhee Park, Youngjo Kim, Hyungju Chun, Xin Jin, Chongyi Li, Chun-Le Guo, Radu Timofte, et al. NTIRE 2025 challenge on efficient burst hdr and restoration: Datasets, methods, and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[29] Jianze Li, Jiezhang Cao, Zichen Zou, Xiongfei Su, Xin Yuan, Yulun Zhang, Yong Guo, and Xiaokang Yang. Distillation-free one-step diffusion for real-world image super-resolution. *arXiv preprint arXiv:2410.04224*, 2024. 2

[30] Xin Li, Yeying Jin, Xin Jin, Zongwei Wu, Bingchen Li, Yufei Wang, Wenhan Yang, Yu Li, Zhibo Chen, Bihan Wen, Robby Tan, Radu Timofte, et al. NTIRE 2025 challenge on day and night raindrop removal for dual-focused images: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[31] Xin Li, Xijun Wang, Bingchen Li, Kun Yuan, Yizhen Shao, Suhang Yao, Ming Sun, Chao Zhou, Radu Timofte, and Zhibo Chen. NTIRE 2025 challenge on short-form ugc video quality assessment and enhancement: Kwaisr dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[32] Xin Li, Kun Yuan, Bingchen Li, Fengbin Guan, Yizhen Shao, Zihao Yu, Xijun Wang, Yiting Lu, Wei Luo, Suhang Yao, Ming Sun, Chao Zhou, Zhibo Chen, Radu Timofte, et al. NTIRE 2025 challenge on short-form ugc video quality assessment and enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[33] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, Rakesh Ranjan, Radu Timofte, and Luc Van Gool. Lsdir: A large scale dataset for image restoration. In *CVPRW*, 2023. 2, 6

[34] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCVW*, 2021. 3, 6, 7

[35] Jie Liang, Radu Timofte, Qiaosi Yi, Zhengqiang Zhang, Shuaizheng Liu, Lingchen Sun, Rongyuan Wu, Xindong Zhang, Hui Zeng, Lei Zhang, et al. NTIRE 2025 the 2nd restore any image model (RAIM) in the wild challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[36] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *CVPRW*, 2017. 2

[37] Xiaohong Liu, Xiongkuo Min, Qiang Hu, Xiaoyun Zhang, Jie Guo, et al. NTIRE 2025 XGC quality assessment challenge: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[38] Xiaoning Liu, Zongwei Wu, Florin-Alexandru Vasluianu, Hailong Yan, Bin Ren, Yulun Zhang, Shuhang Gu, Le Zhang, Ce Zhu, Radu Timofte, et al. NTIRE 2025 challenge on low light image enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[39] 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. 2, 8

[40] 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*, 2021. 2

[41] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017. 6

[42] Fangzhou Luo, Xiaolin Wu, and Yanhui Guo. And: Adversarial neural degradation for learning blind image super-resolution. *NeurIPS*, 2024. 8

[43] Alec Radford, Jong Wook Kim, Chris Hallacy, Aditya Ramesh, Gabriel Goh, Sandhini Agarwal, Girish Sastry, Amanda Askell, Pamela Mishkin, Jack Clark, et al. Learning transferable visual models from natural language supervision. In *ICML*, 2021. 6, 7

[44] Bin Ren, Hang Guo, Lei Sun, Zongwei Wu, Radu Timofte, Yawei Li, et al. The tenth NTIRE 2025 efficient super-resolution challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[45] Nickolay Safonov, Alexey Bryntsev, Andrey Moskalenko, Dmitry Kulikov, Dmitriy Vatolin, Radu Timofte, et al. NTIRE 2025 challenge on UGC video enhancement: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[46] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *TPAMI*, 2022. 2

[47] Jian Sun, Jiejie Zhu, and Marshall F Tappen. Context-constrained hallucination for image super-resolution. In *CVPR*, 2010. 2

[48] Lei Sun, Andrea Alfarano, Peiqi Duan, Shaolin Su, Kaiwei Wang, Boxin Shi, Radu Timofte, Danda Pani Paudel, Luc Van Gool, et al. NTIRE 2025 challenge on event-based image deblurring: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[49] Lei Sun, Hang Guo, Bin Ren, Luc Van Gool, Radu Timofte, Yawei Li, et al. The tenth ntire 2025 image denoising challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[50] Radu Timofte, Eirikur Agustsson, Luc Van Gool, Ming-Hsuan Yang, Lei Zhang, Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, Kyoung Mu Lee, et al. Ntire 2017 challenge on single image super-resolution: Methods and results. In *CVPRW*, 2017. 2, 6, 8

[51] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Cailian Chen, Zongwei Wu, Radu Timofte, et al. NTIRE

2025 image shadow removal challenge report. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[52] Florin-Alexandru Vasluianu, Tim Seizinger, Zhuyun Zhou, Zongwei Wu, Radu Timofte, et al. NTIRE 2025 ambient lighting normalization challenge. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[53] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *NeurIPS*, 2017. 2

[54] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, 2023. 6

[55] Yingqian Wang, Zhengyu Liang, Fengyuan Zhang, Lvli Tian, Longguang Wang, Juncheng Li, Jungang Yang, Radu Timofte, Yulan Guo, et al. NTIRE 2025 challenge on light field image super-resolution: Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[56] Zhaowen Wang, Ding Liu, Jianchao Yang, Wei Han, and Thomas Huang. Deep networks for image super-resolution with sparse prior. In *ICCV*, 2015. 2

[57] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *CVPR*, 2024. 2

[58] Bin Xia, Yulun Zhang, Shiyin Wang, Yitong Wang, Xinglong Wu, Yapeng Tian, Wenming Yang, and Luc Van Gool. Diffir: Efficient diffusion model for image restoration. *arXiv preprint arXiv:2303.09472*, 2023. 2

[59] Peizhe Xia, Long Peng, Xin Di, Renjing Pei, Yang Wang, Yang Cao, and Zheng-Jun Zha. S^3mamba: Arbitrary-scale super-resolution via scaleable state space model. *arXiv preprint arXiv:2411.11906*, 2024. 6

[60] Kangning Yang, Jie Cai, Ling Ouyang, Florin-Alexandru Vasluianu, Radu Timofte, Jiaming Ding, Huiming Sun, Lan Fu, Jinlong Li, Chiu Man Ho, Zibo Meng, et al. NTIRE 2025 challenge on single image reflection removal in the wild: Datasets, methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[61] Pierluigi Zama Ramirez, Fabio Tosi, Luigi Di Stefano, Radu Timofte, Alex Costanzino, Matteo Poggi, Samuele Salti, Stefano Mattoccia, et al. NTIRE 2025 challenge on hr depth from images of specular and transparent surfaces. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, 2025. 2

[62] Dafeng Zhang, Feiyu Huang, Shizhuo Liu, Xiaobing Wang, and Zhezhu Jin. Swinfir: Revisiting the swinir with fast fourier convolution and improved training for image super-resolution. *arXiv preprint arXiv:2208.11247*, 2022. 7

[63] He Zhang, Vishwanath Sindagi, and Vishal M Patel. Image de-raining using a conditional generative adversarial network. *arXiv preprint arXiv:1701.05957*, 2017. 2

[64] Kaibing Zhang, Xinbo Gao, Dacheng Tao, and Xuelong Li. Single image super-resolution with non-local means and steering kernel regression. *TIP*, 2012. 2

[65] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, 2018. 6

[66] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *ECCV*, 2018. 2

[67] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *CVPR*, 2018. 2

[68] Yulun Zhang, Kai Zhang, Zheng Chen, Yawei Li, Radu Timofte, Junpei Zhang, Kexin Zhang, Rui Peng, Yanbiao Ma, Licheng Jia, et al. Ntire 2023 challenge on image super-resolution (x4): Methods and results. In *CVPRW*, 2023. 7

[69] Yulun Zhang, Kai Zhang, Zheng Chen, Yawei Li, Radu Timofte, et al. NTIRE 2023 challenge on image super-resolution (x4): Methods and results. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, 2023. 2

[70] Marcelo Victor Wüst Zibetti and Joceli Mayer. A robust and computationally efficient simultaneous super-resolution scheme for image sequences. *TCSVT*, 2007. 2