Time-Aware One Step Diffusion Network for Real-World Image Super-Resolution

Supplementary Material

In this supplementary material, we provide the following content:

- Detailed derivation about the Variational Score Distillation loss in Section 7
- Visual comparisons and quantitative metrics of TADSR across different timesteps in Section 8.1
- Ablation study on the blurred MSE loss in Section 8.2
- Ablation study on the hyperparameters of TAVSD loss in Section 8.3
- Efficiency comparison between TADSR and other diffusion-based Real-ISR methods in Section 9
- Comparisons with GAN-based Real-ISR methods in Section 10
- Extended visual comparisons with SD-based Real-ISR approaches in Section 11

7. Detailed Derivation

According to the original diffusion process in SD [21], at step t, the current state z_t satisfies:

$$\boldsymbol{z}_t = \alpha_t \boldsymbol{z}_0 + \beta_t \boldsymbol{\epsilon}, t = 1, 2, \ldots, T, \tag{11}$$

where α_t and β_t are the scale parameters in diffusion, $\boldsymbol{\epsilon} \sim \mathcal{N}(\mathbf{0}, \boldsymbol{I}^2)$ and $\boldsymbol{z}_0$ is HR latent in Real-ISR task. Therefore, we can express z_0 in terms of z_t and ϵ as $\boldsymbol{z}_0 = \frac{\boldsymbol{z}_t - \beta_t \boldsymbol{\epsilon}}{\alpha_t}$. Then, we can rewrite Eq. (2) in the main paper as follows:

$$\begin{aligned}
\nabla_\theta \mathcal{L}_{VSD} &= \mathbb{E}_{t,\epsilon}\left[\omega(t)\left(\epsilon_\psi(\hat{z}_t; t, c) - \epsilon_\phi(\hat{z}_t; t, c)\right)\frac{\partial \hat{z}}{\partial \theta}\right], \\
&= \mathbb{E}_{t,\epsilon}\left[\omega(t)\frac{\alpha_t}{\beta_t}\left(\hat{z}_\phi(\hat{z}_t; t, c) - \hat{z}_\psi(\hat{z}_t; t, c)\right)\frac{\partial \hat{z}}{\partial \theta}\right], \\
&= \mathbb{E}_{t,\epsilon}\left[\omega'(t)\left(\hat{z}_\phi(\hat{z}_t; t, c) - \hat{z}_\psi(\hat{z}_t; t, c)\right)\frac{\partial \hat{z}}{\partial \theta}\right],
\end{aligned} \tag{12}$$

where ϵ_ψ is the pre-trained diffusion model (teacher model), ϵ_ϕ represents its replica with trainable LoRA [13] (LoRA model), $\hat{z}_\psi$ and $\hat{z}_\phi$ represent the latent images predicted by the teacher model and the LoRA model respectively, c is a text embedding of a caption describing the input image, and ω_t is a time-varying weighting function. Therefore, we can represent the VSD [30] loss using the residual between the latent images predicted by the teacher model and the LoRA model, which is then decoded into pixel space to analyze the timestep-dependent guidance.

Table 3. Quantitative comparison of ablation study on blurred MSE loss, evaluated on DrealSR [31] dataset.

Methods	PSNR ↑	SSIM ↑	MUSIQ ↑	TOPIQ ↑	QALIGN ↑
$t_s = 100$	29.576	0.8021	64.022	0.6201	3.4666
$t_s = 200$	**29.610**	**0.8059**	63.735	0.6108	3.4750
$t_s = 300$	29.167	0.7935	65.367	0.6463	3.6069
$t_s = 400$	29.245	0.7992	64.681	0.6149	3.5635
$t_s = 500$	28.387	0.7758	**67.016**	**0.6758**	**3.7491**
$t_s = 600$	28.473	0.7834	66.620	0.6597	3.7372

Table 4. Quantitative comparison of ablation study on blurred MSE loss, evaluated on DrealSR [31] dataset.

Methods	PSNR ↑	SSIM ↑	MUSIQ ↑	QALIGN ↑
w/o blurred MSE	29.074	0.7841	64.732	3.5299
TADSR ($t_s = 300$)	**29.167**	**0.794**	65.367	3.6069
TADSR	28.387	0.7758	**67.016**	**3.7491**

8. More Ablation Study

8.1. Different Timesteps

Figure 8 presents TADSR's results at different timesteps t_s, demonstrating a gradual transition from fidelity to realism reconstruction as the t_s increases. Specifically: (1) In the first row, TADSR progressively generates richer eyelash textures and sharper contours; (2) The second row shows how patterned shadows gradually transform into stain-like artifacts; (3) For the third row, TADSR reconstructs plausible architectural stripes not present in the low-quality input; and (4) The fourth row reveals emerging yellow pistils in flower centers. These progressive changes evidence TADSR's enhanced utilization of the pre-trained generative priors in SD at larger t_s, effectively balancing the fidelity-realism trade-off condition on t_s. In addition, we also present the performance of our method at different timesteps. As shown in Table 3, with increasing timesteps, reference metrics (PSNR, SSIM [28]) tend to decrease while no-reference metrics (MUSIQ [14], TOPIQ [5], QALIGN [32]) tend to increase. This is consistent with the visual comparison results in Figure 8, demonstrating that our method can achieve one-step realism–fidelity controllable generation simply by adjusting the timestep.

8.2. Blurred MSE Loss

To avoid gradient inconsistency arising from the ill-posed problem of the Real-ISR task while fully leveraging genera-

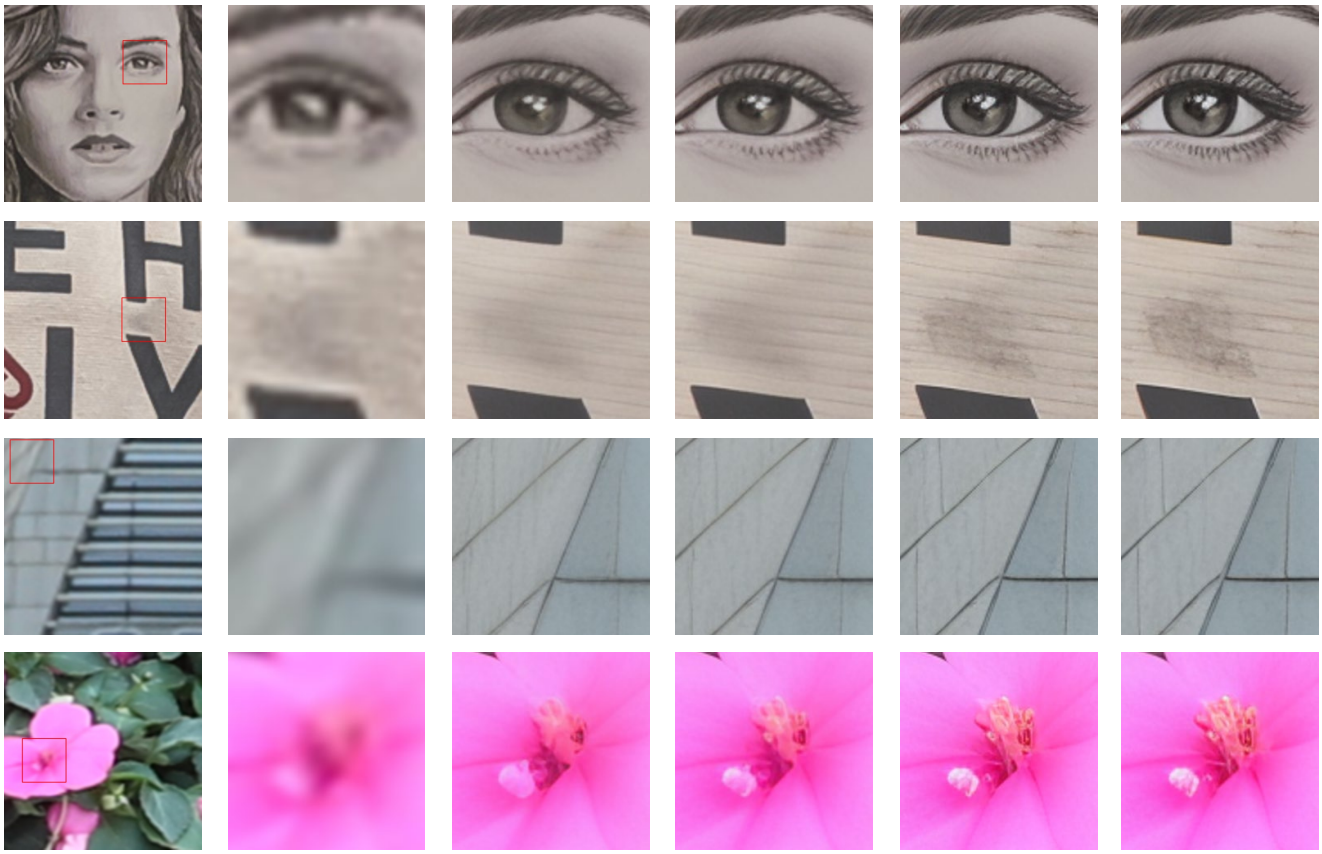

Figure 8. Vision comparisons of TADSR at different timesteps t_s. Zoom in for a better view.

tive prior of SD, we introduce a blurred MSE loss to replace the original MSE loss. Specifically, we first apply a Gaussian blur to both the reconstructed image $G_\theta(x_L)$ and the HQ image x_H before computing the MSE loss. The blurred MSE loss can be formed as:

$$\mathcal{L}_{MSE}^{blur} = \mathcal{L}_{MSE}\left(G_\theta(x_L) * G_{t_s}, x_H * G_{t_s}\right). \tag{13}$$

Where $*$ denotes the convolution operation, G_{t_s} is the Gaussian convolution kernel whose size is determined by t_s. Let k_{t_s} as the kernel size of G_{t_s}, it satisfies:

$$k_{t_s} = 5 + 4 * \lfloor \frac{t_s}{200} \rfloor. \tag{14}$$

To validate the effectiveness of the proposed blurred MSE loss, we performed an ablation study by removing it. As shown in Tab. 4, when the blurred MSE loss is removed, the no-reference metrics degrade (MUSIQ [14], QALIGN [32]) significantly while the reference metrics (PSNR, SSIM [28]) improve, demonstrating a trade-off effect where fidelity is enhanced at the expense of realism. To better align with the reference metrics, we selected TADSR's output at $t_s = 300$. With the blurred MSE loss incorporated, TADSR achieves improvements across all metrics, indicating that this loss function enables a more optimal balance between fidelity and realism.

Table 5. Quantitative comparison of ablation study on the hyperparameters of TAVSD loss, evaluated on RealSR [2] dataset.

Methods		PSNR ↑	SSIM ↑	MUSIQ ↑	MANIQA ↑
λ	γ				
0.25	0	**25.757**	**0.7218**	69.408	0.6541
0.50	**0**	25.166	0.7150	71.182	0.6715
0.75	0	24.398	0.7038	**71.484**	**0.6774**
0.50	50	25.090	0.7159	70.742	0.6702
0.50	100	24.809	0.7115	70.918	0.6727

8.3. Hyperparameters in TAVSD

To verify the sensitivity of our method to the hyperparameters in TAVSD, we conducted ablation studies by varying their values. We set the timestep t_s to 500 and evaluated our method under different hyperparameters on the RealSR [2] dataset. As shown in Table 5, with the increase of λ and γ,

Table 6. A comprehensive evaluation against state-of-the-art GAN-based methods across synthetic and real-world datasets. The top-performing results under each metric are marked in **red**.

Datasets	Methods	PSNR ↑	SSIM ↑	LPIPS ↑	CLIPIQA ↑	MUSIQ ↑	MAINIQA ↑	TOPIQ ↑	QALIGN ↑
DIV2k-Val	BSRGAN	**24.583**	0.6269	0.3351	0.5246	61.196	0.5041	0.5460	3.1708
	RealESRGAN	24.293	**0.6372**	0.3112	0.5277	61.058	0.5485	0.5297	3.2768
	LDL	23.828	0.6344	0.3256	0.5179	60.038	0.5328	0.5144	3.1797
	TADSR	23.815	0.6028	**0.3078**	**0.7353**	**69.649**	**0.6443**	**0.7044**	**4.0783**
DrealSR	BSRGAN	**28.701**	0.8028	0.2858	0.5092	57.165	0.4845	0.5060	2.9580
	RealESRGAN	28.615	0.8051	0.2819	0.4525	54.268	0.4903	0.4623	2.8645
	LDL	28.197	**0.8124**	**0.2792**	0.4475	53.949	0.4894	0.4518	2.8564
	TADSR	28.387	0.7758	0.3235	**0.7398**	**67.016**	**0.6309**	**0.6758**	**3.7491**
RealSR	BSRGAN	**26.379**	**0.7651**	**0.2656**	0.5116	63.287	0.5420	0.5505	3.1843
	RealESRGAN	25.686	0.7614	0.2710	0.4494	60.370	0.5505	0.5148	3.1073
	LDL	25.281	0.7565	0.2750	0.4555	60.928	0.5495	0.5125	3.0888
	TADSR	25.166	0.7150	0.3168	**0.7283**	**71.182**	**0.6715**	**0.7082**	**3.9477**

Table 7. The inference time and the number of parameters of diffusion-based Real-ISR methods. The top-performing results under each metric are marked in **red**.

	StableSR	DiffBIR	SeeSR	SinSR	OSEDiff	S3Diff	AdcSR	TSDSR	PisaSR	TADSR
Inference Step	200	50	50	1	1	1	1	1	2	1
Inference time(s)	10.40	9.83	5.64	0.1785	0.1463	0.4704	**0.0825**	0.0947	0.1675	0.1465
#Params(MB)	1563	1682	2514	**119**	1775	1327	456	2207	1302	1777

our method exhibits a decrease in reference metrics while no-reference metrics improve, reflecting a trade-off of fidelity for enhanced realism.

This phenomenon is consistent with the functionality of TAVSD. As discussed in Section 1, the pre-trained SD model exhibits different generative priors at different timesteps: smaller timesteps tend to favor fidelity, while larger timesteps tend to favor generation. This also means that the teacher model in VSD will provide generation guidance based on semantic priors when the time step is large. Therefor, when λ and γ increase, the same t_s is mapped to a larger t_v, causing the teacher model to provide guidance more biased toward generation. In terms of metrics, this is manifested as an increase in no-reference metrics and a decrease in reference-based metrics. Considering the balance between realism and fidelity, we ultimately choose $\lambda = 0.5$ and $\gamma = 0$ as the default setting for our model.

9. Comparison of Efficiency with Other One-step Real-ISR Methods

We compare the number of parameters and inference time of one-step diffusion-based Real-ISR models in Table 7. Inference time is measured on the $\times 4$ SR task with 128×128 LQ images using a single NVIDIA 3090 24G GPU. Compared with OSEDiff [33], our method achieves roughly the same inference time and parameter count, while showing significant improvements in no-reference metrics and visual quality. PisaSR requires two inferences to achieve controllable Real-ISR due to the presence of two LoRA weights. In contrast, our method can obtain controllable Real-ISR with a single inference simply by adjusting the time step, resulting in fewer inference steps and shorter inference time.

10. Comparisons with GAN-based Real-ISR Methods

We compare TADSR with three GAN-based Real-ISR methods: BSRGAN [41], RealESRGAN [26], and LDL [17]. Quantitative evaluations are conducted on the DIV2K [1], RealSR [2], and DRealSR [31] datasets, with results summarized in Tab. 6. The experimental results demonstrate that TADSR, leveraging the powerful generative priors of the pre-trained SD, achieves significantly superior no-reference metrics (e.g., CLIPIQA [24], MAINIQA [37]) compared to GAN-based methods.

Additionally, Fig. 9 presents a visual comparison between TADSR and other GAN-based methods. The results show that TADSR reconstructs more photorealistic and natural outcomes, including higher fidelity in text and architectural structures (from the first to the third group), and more realistic rope textures (in the fourth group).

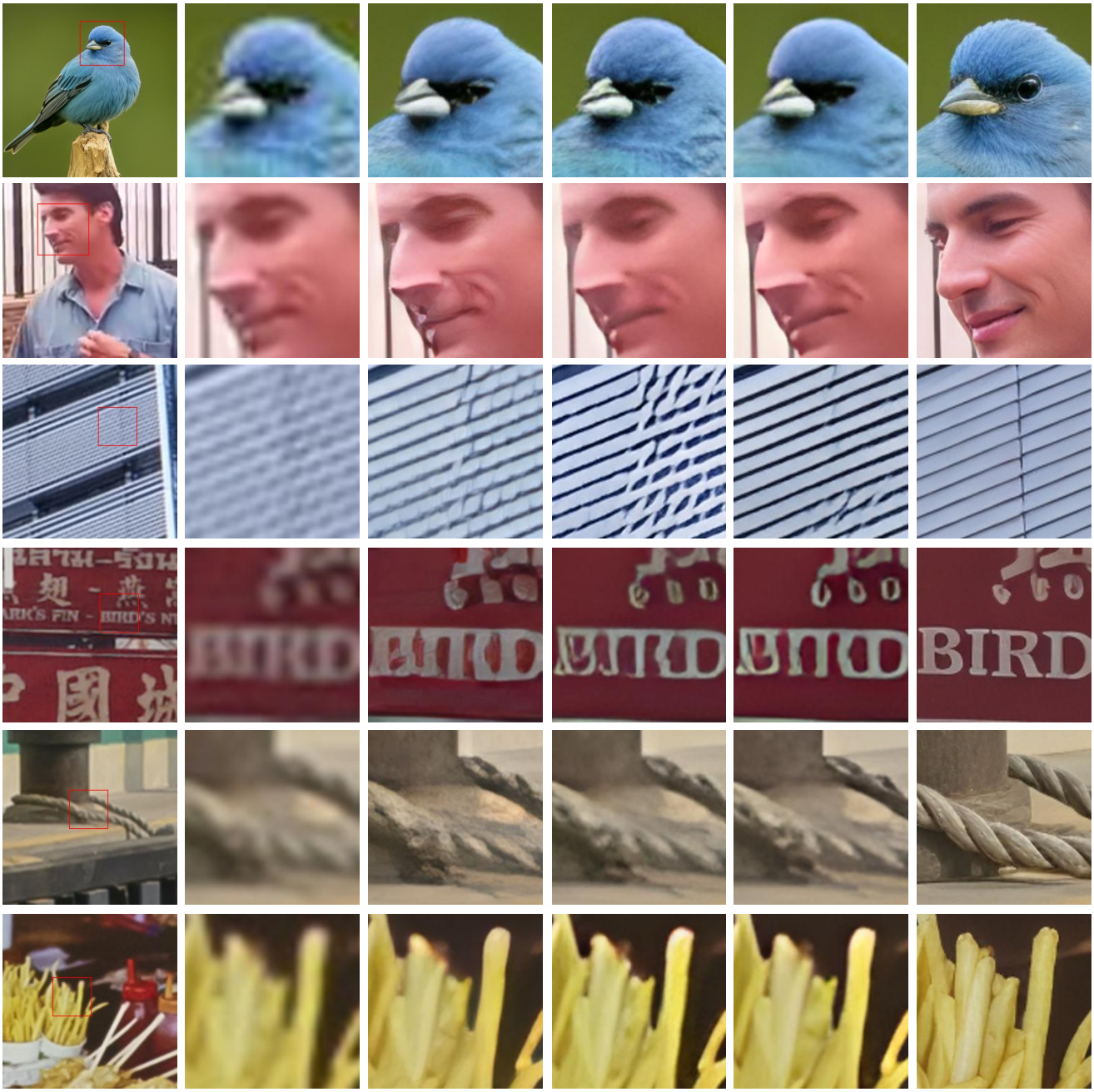

Figure 9. Vision comparisons between TADSR and GAN-based Real-ISR methods. Zoom in for a better view.

11. More Visual Comparisons with SD-based Real-ISR Methods

We provide more visual comparisons between TADSR and other SD-based SR methods in Fig. 10 and Fig. 11. Compared to other methods, TADSR consistently produces clearer, more realistic, and more natural results. Moreover, although our training is conducted at a resolution of 512×512, we provide visual comparisons of TADSR and other diffusion-based one-step Real-ISR methods on 2K-resolution images. As shown in Figure 12, TADSR is also capable of maintaining strong structural consistency and producing realistic, natural SR results on high-resolution images.

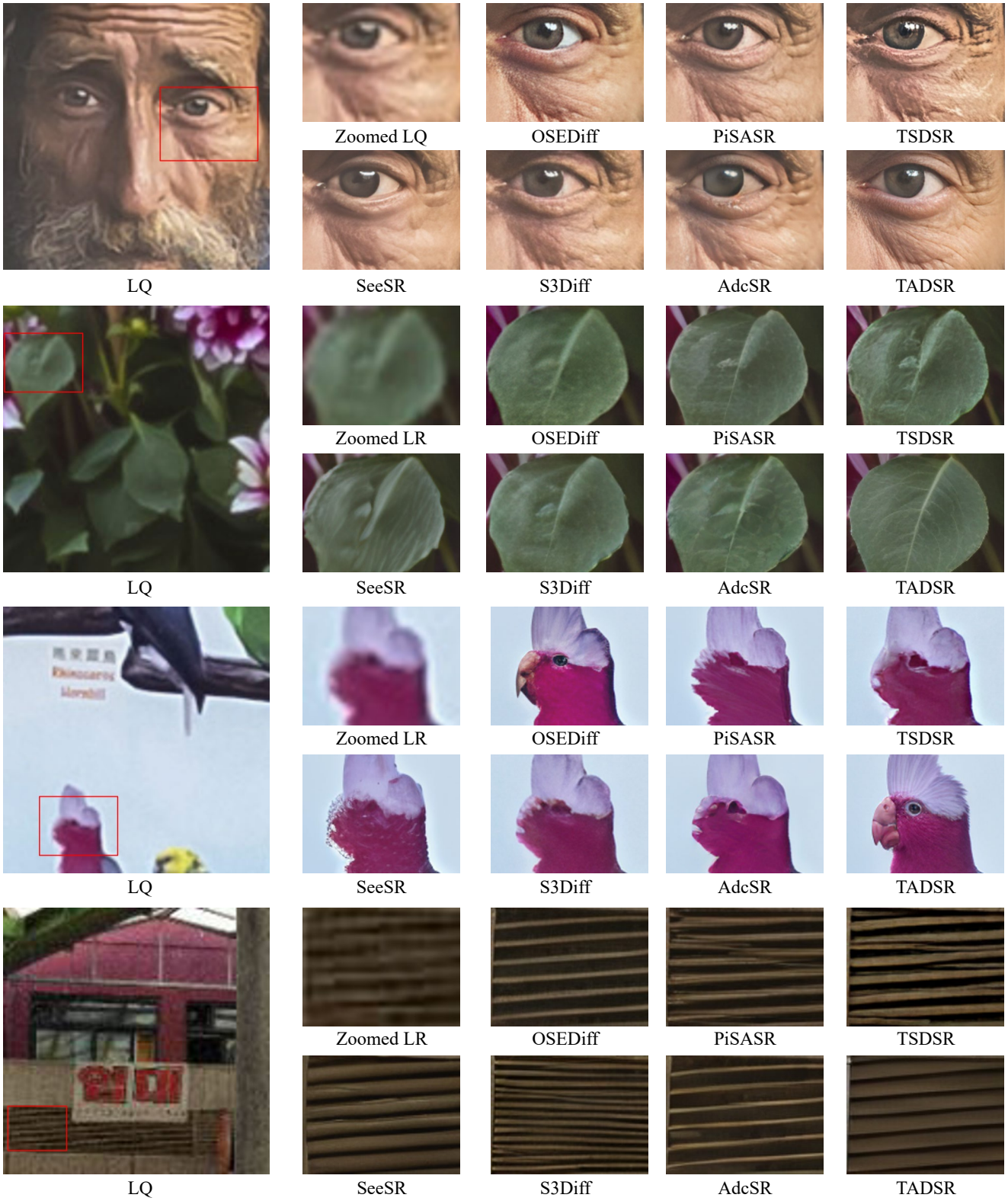

Figure 10. Vision comparisons between TADSR and SD-based Real-ISR methods (SeeSR [34], OSEDiff [33], S3Diff [40], PiSASR [22], AdcSR [4], TSDSR [9]). Zoom in for a better view.

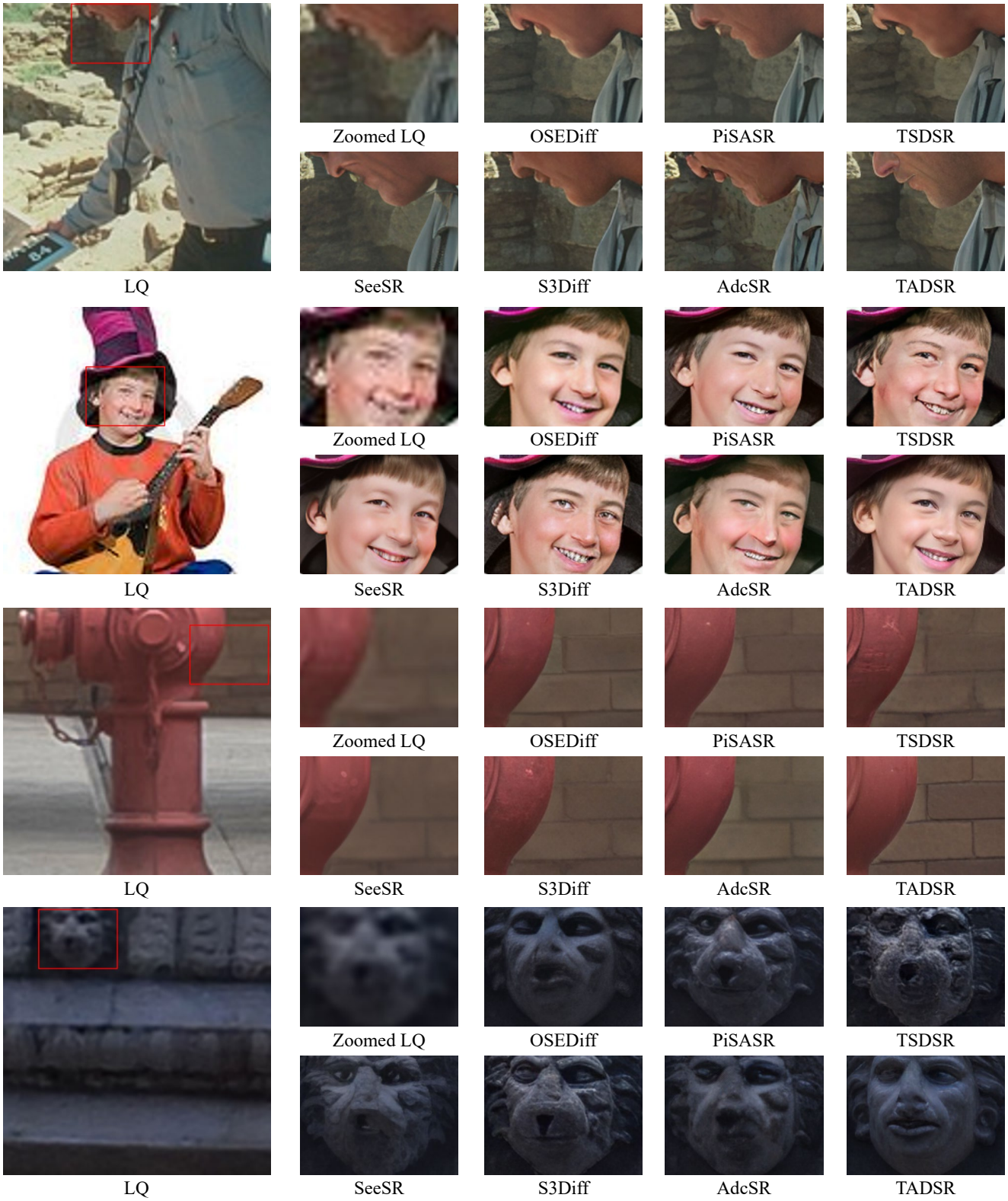

Figure 11. Vision comparisons between TADSR and SD-based Real-ISR methods (SeeSR [34], OSEDiff [33], S3Diff [40], PiSASR [22], AdcSR [4], TSDSR [9]). Zoom in for a better view.

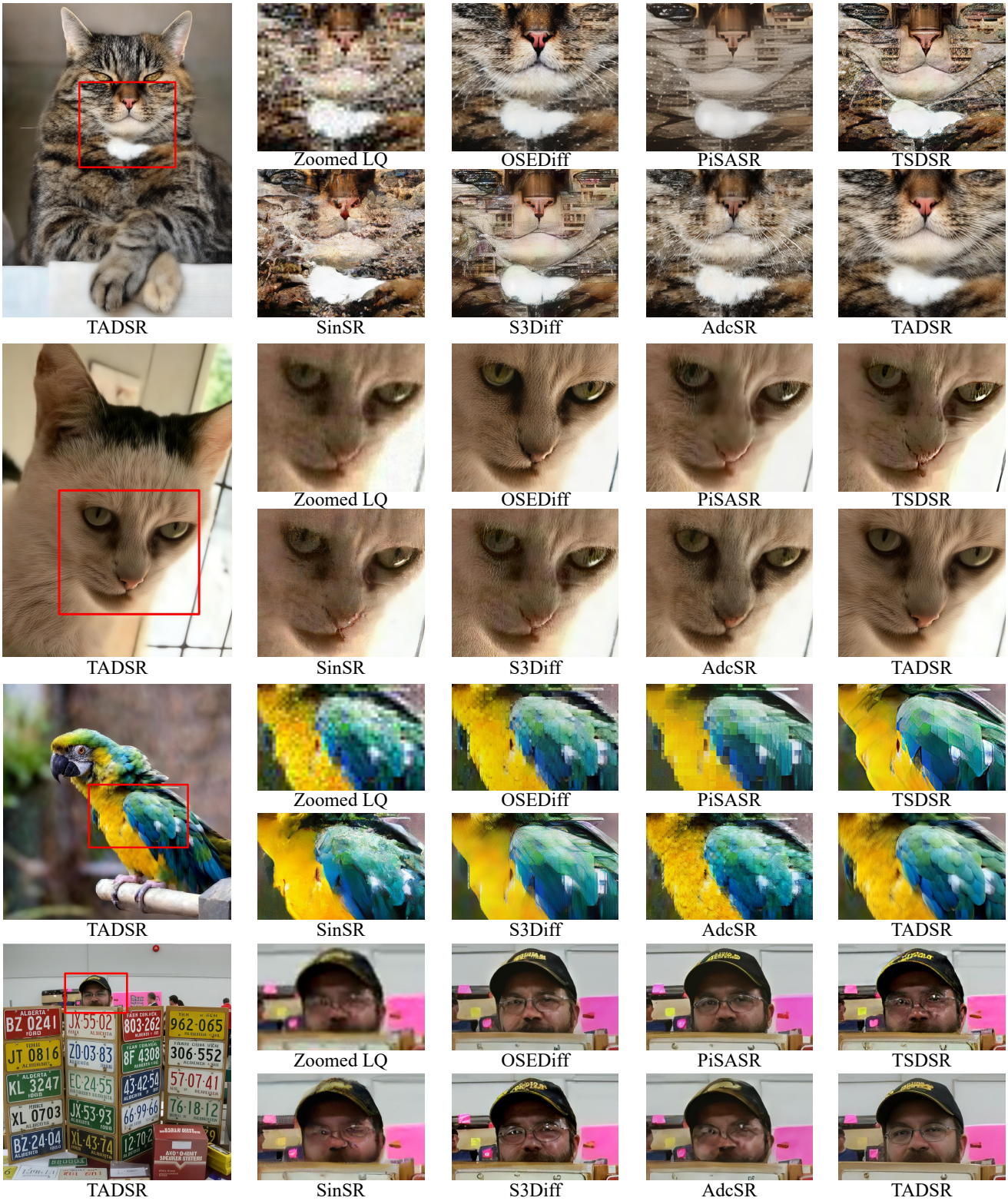

Figure 12. Vision comparisons between TADSR and Diffusion-based Real-ISR methods (SinSR [27], OSEDiff [33], S3Diff [40], PiSASR [22], AdcSR [4], TSDSR [9]). Zoom in for a better view.

References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *CVPRW*, pages 126–135, 2017. 7, 12

[2] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *ICCV*, pages 3086–3095, 2019. 7, 11, 12

[3] Xiaofeng Cao, Mingwei Xu, Xin Yu, Jiangchao Yao, Wei Ye, Shengjun Huang, Minling Zhang, Ivor Tsang, Yew-Soon Ong, James T Kwok, et al. Analytical survey of learning with low-resource data: From analysis to investigation. *ACM Computing Surveys*, 58(6):1–47, 2025. 2

[4] Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression for real-world image super-resolution. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 28208–28220, 2025. 2, 7, 14, 15, 16

[5] Chaofeng Chen, Jiadi Mo, Jingwen Hou, Haoning Wu, Liang Liao, Wenxiu Sun, Qiong Yan, and Weisi Lin. Topiq: A top-down approach from semantics to distortions for image quality assessment. *IEEE TIP*, 33:2404–2418, 2024. 7, 10

[6] 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. 2

[7] 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.

[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. 2

[9] 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. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 23174–23184, 2025. 2, 7, 14, 15, 16

[10] Zheng-Peng Duan, Jiawei Zhang, Xin Jin, Ziheng Zhang, Zheng Xiong, Dongqing Zou, Jimmy Ren, Chun-Le Guo, and Chongyi Li. Dit4sr: Taming diffusion transformer for real-world image super-resolution. *arXiv preprint arXiv:2503.23580*, 2025. 1

[11] Jonathan Ho and Tim Salimans. Classifier-free diffusion guidance. *arXiv preprint arXiv:2207.12598*, 2022. 5

[12] Jonathan Ho, Ajay Jain, and Pieter Abbeel. Denoising diffusion probabilistic models. *NeurIPS*, 33:6840–6851, 2020. 1

[13] Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*, 2021. 10

[14] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *ICCV*, pages 5148–5157, 2021. 7, 10, 11

[15] Yawei Li, Kai Zhang, Jingyun Liang, Jiezhang Cao, Ce Liu, Rui Gong, Yulun Zhang, Hao Tang, Yun Liu, Denis Demandolx, et al. Lsdir: A large scale dataset for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1775–1787, 2023. 6

[16] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *ICCV*, pages 1833–1844, 2021. 2

[17] Jie Liang, Hui Zeng, and Lei Zhang. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In *CVPR*, pages 5657–5666, 2022. 1, 5, 12

[18] 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. 2

[19] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Ben Fei, Bo Dai, Wanli Ouyang, Yu Qiao, and Chao Dong. Diffbir: Towards blind image restoration with generative diffusion prior. *arXiv preprint arXiv:2308.15070*, 2023. 1, 3, 7

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

[21] Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *CVPR*, pages 10684–10695, 2022. 1, 10

[22] Lingchen Sun, Rongyuan Wu, Zhiyuan Ma, Shuaizheng Liu, Qiaosi Yi, and Lei Zhang. Pixel-level and semantic-level adjustable super-resolution: A dual-lora approach. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 2333–2343, 2025. 1, 2, 3, 7, 14, 15, 16

[23] Yang Tao, Wu Rongyuan, Ren Peiran, Xie Xuansong, and Zhang Lei. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *ECCV*, 2023. 1, 3

[24] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *AAAI*, pages 2555–2563, 2023. 7, 12

[25] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *arXiv preprint arXiv:2305.07015*, 2023. 1, 3, 7

[26] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *ICCV*, pages 1905–1914, 2021. 1, 2, 6, 12

[27] Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. Sinsr: Diffusion-based image super-resolution in a single step. In *CVPR*, 2024. 7, 16

[28] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 10, 11

[29] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE TIP*, 13(4):600–612, 2004. 7

[30] Zhengyi Wang, Cheng Lu, Yikai Wang, Fan Bao, Chongxuan Li, Hang Su, and Jun Zhu. Prolificdreamer: High-fidelity and

diverse text-to-3d generation with variational score distillation. *arXiv preprint arXiv:2305.16213*, 2023. 2, 10

[31] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In *ECCV*, pages 101–117. Springer, 2020. 7, 10, 12

[32] Haoning Wu, Zicheng Zhang, Weixia Zhang, Chaofeng Chen, Liang Liao, Chunyi Li, Yixuan Gao, Annan Wang, Erli Zhang, Wenxiu Sun, Qiong Yan, Xiongkuo Min, Guangtao Zhai, and Weisi Lin. Q-align: Teaching LMMs for visual scoring via discrete text-defined levels. In *ICML*, pages 54015–54029. PMLR, 2024. 7, 10, 11

[33] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *arXiv preprint arXiv:2406.08177*, 2024. 2, 3, 4, 5, 7, 12, 14, 15, 16

[34] 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. 1, 3, 6, 7, 14, 15

[35] Zhiqiang Wu, Zhaomang Sun, Tong Zhou, Bingtao Fu, Ji Cong, Yitong Dong, Huaqi Zhang, Xuan Tang, Mingsong Chen, and Xian Wei. Omgsr: You only need one mid-timestep guidance for real-world image super-resolution. *arXiv preprint arXiv:2508.08227*, 2025. 1

[36] Liangbin Xie, Xintao Wang, Xiangyu Chen, Gen Li, Ying Shan, Jiantao Zhou, and Chao Dong. Desra: detect and delete the artifacts of gan-based real-world super-resolution models. *arXiv preprint arXiv:2307.02457*, 2023. 1

[37] Sidi Yang, Tianhe Wu, Shuwei Shi, Shanshan Lao, Yuan Gong, Mingdeng Cao, Jiahao Wang, and Yujiu Yang. Maniqa: Multi-dimension attention network for no-reference image quality assessment. In *CVPR*, pages 1191–1200, 2022. 7, 12

[38] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *arXiv preprint arXiv:2307.12348*, 2023. 1

[39] Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 23153–23163, 2025. 2

[40] Aiping Zhang, Zongsheng Yue, Renjing Pei, Wenqi Ren, and Xiaochun Cao. Degradation-guided one-step image super-resolution with diffusion priors. *arXiv preprint arXiv:2409.17058*, 2024. 2, 7, 14, 15, 16

[41] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *ICCV*, pages 4791–4800, 2021. 1, 2, 12

[42] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *CVPR*, pages 586–595, 2018. 7

[43] 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*, pages 286–301, 2018. 2

[44] Qianqian Zhao, Chunle Guo, Tianyi Zhang, Junpei Zhang, Peiyang Jia, Tan Su, Wenjie Jiang, and Chongyi Li. A systematic investigation on deep learning-based omnidirectional image and video super-resolution. *arXiv preprint arXiv:2506.06710*, 2025. 2

[45] Yuanzhi Zhu, Ruiqing Wang, Shilin Lu, Junnan Li, Hanshu Yan, and Kai Zhang. Oftsr: One-step flow for image super-resolution with tunable fidelity-realism trade-offs. *arXiv preprint arXiv:2412.09465*, 2024. 1