# UARE: A Unified Vision-Language Model for Image Quality Assessment, Restoration, and Enhancement

## Supplementary Material

Our main paper presents the core ideas, architecture, and experimental results of UARE, a unified vision–language model for image quality assessment, restoration, and enhancement. In this ***supplementary material***, we provide additional information. Sec. A gives implementation details, including representative data examples in Sec. A.1 and training configurations in Sec. A.2. Sec. B reports more comparison results, with real-world super resolution in Sec. B.1, image restoration/enhancement in Sec. B.2, and a user study in Sec. B.3. Finally, Sec. C offers further discussion of UARE and a detailed analysis of its limitations.

## A. Implementation Details

### A.1. Data Examples

We have detailed the data construction process in our main paper. Here, we provide data examples for training UARE in Fig. A.1 and Fig. A.2 for IQA and restoration/enhancement, respectively. For IQA, our data include diverse instruction formats such as free-form quality description, scalar quality scoring, and reference-based pairwise comparison. For restoration and enhancement, we construct instructions for single, multiple, and high-order degradations, as well as interleaved text–image data where the model must first analyze the degradation and user intent and then plan the enhancement steps. These examples span a wide range of scenes (indoor/outdoor, day/night, natural and urban) and degradation types (blur, low light, noise, haze, and complex mixed artifacts), highlighting the richness and compositionality of our training corpus.

### A.2. Training Details

Table A.1 summarizes the full training recipe of UARE. For all stages, we use a constant learning rate of $2 \times 10^{-5}$, zero weight decay, gradient-norm clipping of 1.0, and AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.95$, $\epsilon = 1.0 \times 10^{-15}$) with EMA ratios of 0.990, 0.995, 0.995, and 0.995 for the single-degradation, multi-degradation, high-order degradation, and unified fine-tuning stages, respectively. The three stage-1 curricula are trained for 10K, 20K, and 1.5K steps with 250 warm-up steps, while the unified stage is trained for 10K steps with a 500-step warm-up. In the unified stage, we jointly optimize the IQA cross-entropy loss and the restoration MSE loss with a weight ratio of $0.25 : 1$. Restoration images are randomly resized such that the minimum short side and maximum long side fall in $(512, 1024)$, and IQA images are resized to $(378, 980)$. We apply a diffusion timestep shift of 4.0 for all diffusion-based restoration branches. The bottom part of Table A.1 lists the data sampling ratios: stage-1 curricula only sample their corresponding degradation type (single, multi, or high-order), whereas the unified stage mixes IQA, single-, multi-, and high-order degradation data and interleaved IQA–restoration pairs, forming a balanced curriculum for learning unified quality assessment and restoration.

## B. More Comparison Results

### B.1. Real-World Super Resolution

**More quantitative results.** We further report more quantitative results in Tab. B.1 on RealSR [2], DRealSR [20] and

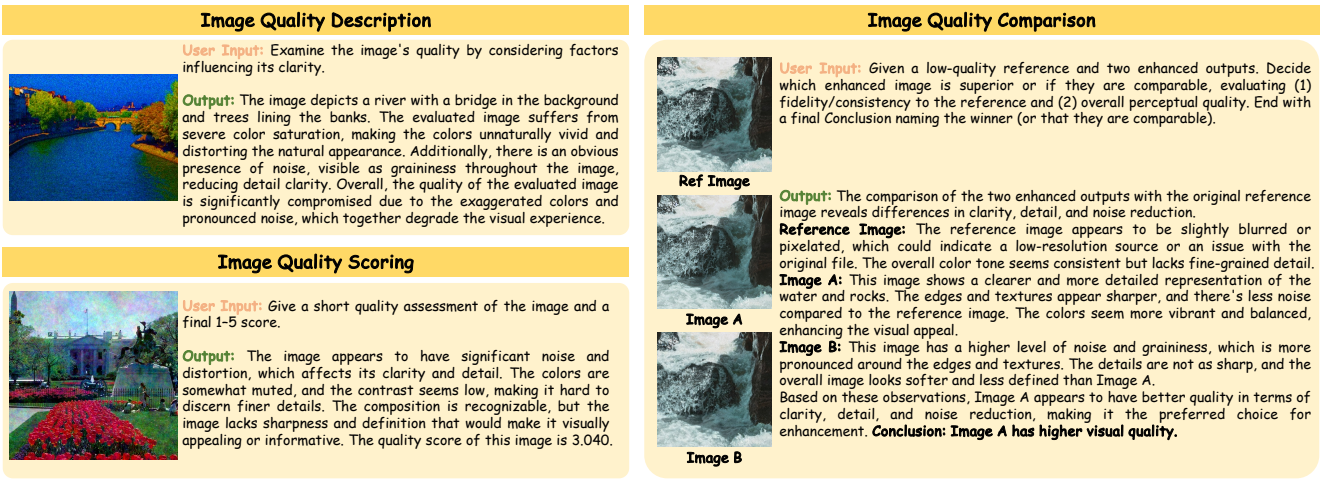


Figure A.1. Data examples for IQA training in UARE, including quality description, image quality scoring, and image comparison.

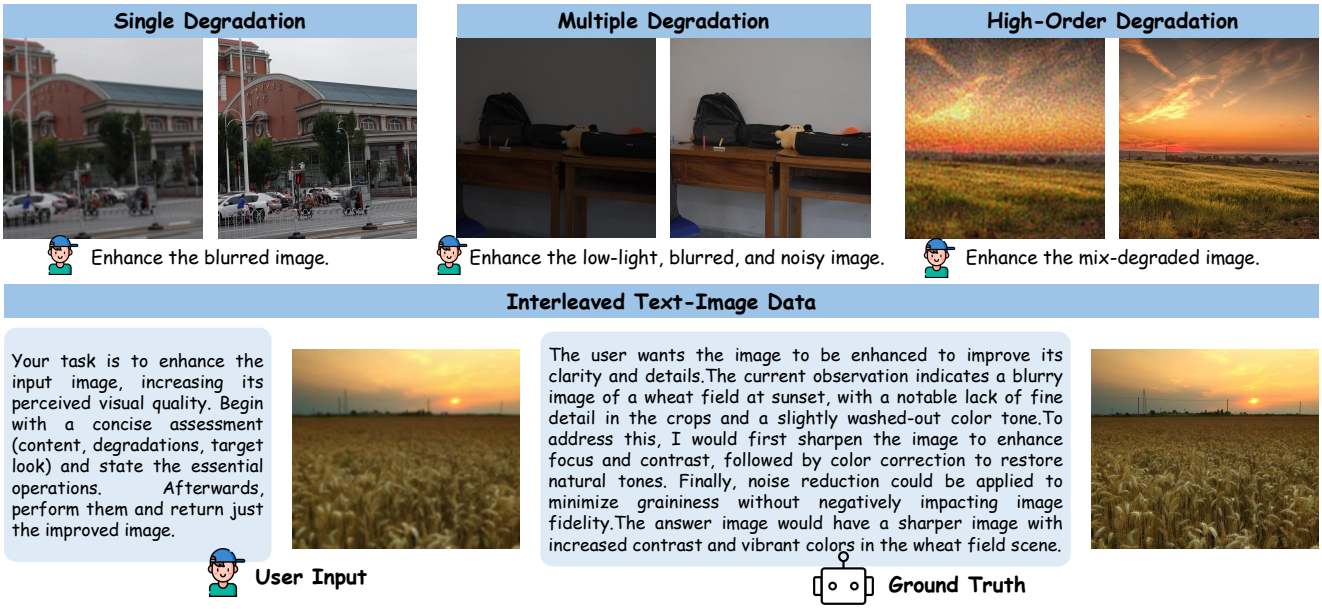


Figure A.2. Data examples for restoration and enhancement training in UARE, covering single, multiple, and high-order degradations as well as interleaved text–image pairs.

Table A.1. **Training recipe of UARE.**

| | **single deg.** | **multi deg.** | **high-order deg.** | **Uni ft.** |
|---|---|---|---|---|
| **Hyperparameters** | | | | |
| Learning rate | $2 \times 10^{-5}$ | | | |
| LR scheduler | Constant | | | |
| Weight decay | 0.0 | | | |
| Gradient norm clip | 1.0 | | | |
| Optimizer | AdamW ($\beta_1 = 0.9$, $\beta_2 = 0.95$, $\epsilon = 1.0 \times 10^{-15}$) | | | |
| Loss weight (CE : MSE) | - | - | - | 0.25 : 1 |
| Warm-up steps | 250 | 250 | 250 | 500 |
| Training steps | 10K | 20K | 1.5k | 10K |
| EMA ratio | 0.990 | 0.995 | 0.995 | 0.995 |
| Training seen tokens | 9.6B | 19.2B | 1.3B | 4.6B |
| Res. resolution (min short side, max long side) | (512, 1024) | | | |
| IQA resolution (min short side, max long side) | (378, 980) | | | |
| Diffusion timestep shift | 4.0 | | | |
| **Data sampling ratio** | | | | |
| IQA | 0.0 | 0.0 | 0.0 | 0.25 |
| Single degradation | 1.0 | 0.0 | 0.0 | 0.05 |
| Multiple degradation | 0.0 | 1.0 | 0.0 | 0.1 |
| high-order degradation | 0.0 | 0.0 | 1.0 | 0.2 |
| Interleaved IQA and restoration | 0.0 | 0.0 | 0.0 | 0.4 |

DIV2K [1]. We compare UARE with **twelve** SR methods: Real-ESRGAN [16], FeMASR [3], SwinIR [9], InvSR [28], StableSR [15], DiffBIR [11], SeeSR [22], PASD [25], ResShift [27], SinSR [17], OSEDiff [21], S3Diff [30], and PURE [19]. The evaluation metrics follow the main paper: for reference-based fidelity, we report PSNR and SSIM [18] on the Y channel in YCbCr space; for reference-based perceptual quality, we use LPIPS and DISTS; for no-reference quality, we adopt NIQE [33], LIQE [34], MUSIQ [7], MANIQA [24], and TOPIQ [4]. As shown in Tab. B.1,

Table B.1. Quantitative comparison of different methods on RealSR, DRealSR, and DIV2K. Throughout this paper, best, second-best, and third-best results are highlighted in **bold red**, underlined blue, *italic green*. ↑ / ↓ indicate higher/lower is better.

| Test Dataset | Method | PSNR↑ | SSIM↑ | LPIPS↓ | DISTS↓ | NIQE↓ | LIQE↑ | MUSIQ↑ | MANIQA↑ | TOPIQ↑ |
|---|---|---|---|---|---|---|---|---|---|---|
| RealSR | Real-ESRGAN [16] | 23.62 | *0.7185* | 0.2763 | 0.2063 | 5.7619 | 3.3163 | 59.87 | 0.3749 | 0.5097 |
| | FeMASR [3] | 23.26 | 0.7030 | 0.2850 | 0.2254 | 5.7053 | 3.1587 | 58.05 | 0.3435 | 0.4848 |
| | SwinIR [9] | 23.75 | **0.7250** | **0.2608** | *0.1981* | 5.6989 | 3.0798 | 58.95 | 0.3546 | 0.4816 |
| | InvSR [28] | 22.90 | 0.6844 | 0.2634 | 0.1980 | 5.9996 | 3.7639 | 67.20 | 0.4270 | 0.5546 |
| | StableSR [15] | 23.73 | 0.6979 | 0.2792 | 0.2023 | 5.5914 | 3.0532 | 61.65 | 0.3826 | 0.5201 |
| | DiffBIR [11] | 23.20 | 0.6346 | 0.3350 | 0.2162 | **4.5879** | 3.5529 | 65.25 | 0.4620 | 0.6033 |
| | SeeSR [22] | 24.34 | 0.7187 | 0.2754 | 0.2134 | 6.4146 | 3.3938 | 65.53 | *0.4856* | 0.6246 |
| | PASD [25] | **24.50** | 0.7115 | *0.2716* | **0.1954** | 6.0067 | 2.8541 | 58.52 | 0.3831 | 0.4969 |
| | ResShift [27] | *24.17* | 0.6528 | 0.4336 | 0.2812 | 8.6273 | 2.6610 | 53.38 | 0.3412 | 0.4210 |
| | SinSR [17] | 23.68 | 0.6649 | 0.3490 | 0.2445 | 6.5101 | 3.2255 | 61.03 | 0.4230 | 0.5383 |
| | OSEDiff [21] | 23.07 | 0.6850 | 0.2941 | 0.2109 | 5.5054 | **4.0681** | 68.95 | 0.4876 | 0.6441 |
| | S3Diff [30] | 23.16 | 0.6810 | 0.2748 | 0.1986 | *5.3003* | *4.0080* | *67.57* | 0.4677 | *0.6301* |
| | PURE [19] | 21.31 | 0.5738 | 0.3859 | 0.2468 | 5.6419 | 3.7881 | 66.57 | 0.4829 | *0.6301* |
| | **UARE (Ours)** | 21.38 | 0.6464 | 0.3095 | 0.2344 | 5.2981 | 4.0658 | **69.67** | **0.5260** | **0.6796** |
| DRealSR | Real-ESRGAN [16] | 27.26 | *0.7745* | 0.2841 | *0.2085* | 6.6994 | 2.8595 | 53.43 | 0.3438 | 0.4559 |
| | FeMASR [3] | 25.32 | 0.7221 | 0.3164 | 0.2241 | 5.8831 | 2.9538 | 53.32 | 0.3169 | 0.4722 |
| | SwinIR [9] | 27.01 | 0.7703 | 0.2793 | 0.2070 | 6.5370 | 2.8340 | 52.42 | 0.3310 | 0.4484 |
| | InvSR [28] | 25.55 | 0.7087 | 0.3188 | 0.2192 | 6.0231 | *3.7525* | *64.25* | 0.4301 | 0.5726 |
| | StableSR [15] | **28.28** | **0.7981** | **0.2687** | **0.2026** | 7.2816 | 2.5068 | 51.62 | 0.3226 | 0.4355 |
| | DiffBIR [11] | 26.08 | 0.6578 | 0.4144 | 0.2564 | **4.4856** | 3.3993 | 61.81 | 0.4612 | 0.6084 |
| | SeeSR [22] | *28.14* | 0.7798 | *0.2832* | 0.2241 | 7.4833 | 2.7943 | 55.89 | 0.3976 | 0.5436 |
| | PASD [25] | 28.18 | 0.7722 | 0.2970 | 0.2108 | 7.4421 | 2.6129 | 51.42 | 0.3595 | 0.4587 |
| | ResShift [27] | 27.39 | 0.6907 | 0.4996 | 0.3077 | 9.1788 | 1.7905 | 40.58 | 0.2457 | 0.3414 |
| | SinSR [17] | 26.72 | 0.6933 | 0.4031 | 0.2624 | 6.8825 | 2.7781 | 53.36 | 0.3677 | 0.4959 |
| | OSEDiff [21] | 25.60 | 0.7403 | 0.3088 | 0.2158 | 6.1544 | 3.9797 | 65.24 | 0.4879 | 0.6273 |
| | S3Diff [30] | 26.18 | 0.7197 | 0.3161 | 0.2099 | *5.9531* | 3.9255 | 63.34 | *0.4635* | *0.6181* |
| | PURE [19] | 23.04 | 0.5718 | 0.4461 | 0.2674 | 6.3939 | 3.7390 | 60.68 | 0.4362 | 0.5888 |
| | **UARE (Ours)** | 21.31 | 0.5736 | 0.4071 | 0.2613 | 6.4290 | **4.0445** | **67.71** | **0.5121** | **0.6652** |
| DIV2K | Real-ESRGAN [16] | 19.41 | 0.4901 | 0.4123 | 0.2586 | 4.5100 | 3.6731 | 61.63 | 0.3835 | 0.5449 |
| | FeMASR [3] | 18.46 | 0.4339 | 0.4139 | 0.2382 | *4.1173* | 3.4794 | 61.01 | 0.3068 | 0.5151 |
| | SwinIR [9] | 19.11 | *0.4772* | 0.4285 | 0.2647 | 4.7146 | 3.2109 | 58.22 | 0.3251 | 0.4844 |
| | InvSR [28] | 18.93 | 0.4597 | 0.4182 | 0.2685 | 5.8936 | 3.5459 | 61.91 | 0.4066 | 0.5573 |
| | StableSR [15] | **19.85** | **0.4940** | 0.4796 | 0.2887 | 5.7479 | 1.8466 | 43.25 | 0.2181 | 0.3276 |
| | DiffBIR [11] | 18.94 | 0.4332 | 0.4009 | 0.2238 | **3.6594** | 3.8573 | 67.20 | 0.4574 | 0.6467 |
| | SeeSR [22] | 19.11 | 0.4580 | *0.3769* | 0.2339 | 4.5817 | 3.7445 | 66.31 | *0.4686* | 0.6330 |
| | PASD [25] | 18.98 | 0.4562 | 0.4293 | 0.2373 | 4.7846 | 3.6022 | 63.46 | 0.4025 | 0.5653 |
| | ResShift [27] | *19.15* | 0.4311 | 0.4900 | 0.2808 | 7.4321 | 2.8862 | 56.02 | 0.3534 | 0.4662 |
| | SinSR [17] | 18.58 | 0.4059 | 0.4483 | 0.2455 | 6.0533 | 3.4629 | 64.12 | 0.4483 | 0.5997 |
| | OSEDiff [21] | 18.86 | 0.4563 | 0.3579 | *0.2209* | 4.1756 | 3.8877 | 67.83 | 0.4422 | 0.6269 |
| | S3Diff [30] | 18.76 | 0.4490 | **0.3299** | **0.1990** | 4.2026 | 4.2692 | *69.31* | 0.4675 | 0.6679 |
| | PURE [19] | 16.71 | 0.3661 | 0.4449 | 0.2293 | 4.9545 | **4.2701** | 70.06 | **0.5201** | *0.6621* |
| | **UARE (Ours)** | 16.59 | 0.3857 | 0.4074 | 0.2138 | 3.7931 | *4.2627* | **70.45** | 0.5028 | **0.6864** |

UARE achieves higher SSIM and lower LPIPS/DISTS than PURE, and clearly outperforms all competing methods on MUSIQ, MANIQA, and TOPIQ, while ranking second-best on NIQE and LIQE. These results confirm that UARE delivers the best overall perceptual quality among all compared methods.

Additionally, we evaluate UARE on RealSet80 [27], which contains 80 low-resolution real-world images without ground-truth references. We compare UARE against BSRGAN, StableSR, DiffBIR, SeeSR, SinSR, OSEDiff,

Table B.2. Quantitative comparison of different methods on RealSet80 without ground truth.

| Method | NIQE↓ | LIQE↑ | MUSIQ↑ | MANIQA↑ | TOPIQ↑ |
|---|---|---|---|---|---|
| BSRGAN [32] | 5.1655 | 3.8884 | 64.85 | 0.3941 | 0.5821 |
| StableSR [15] | **4.0798** | 3.9074 | 67.67 | 0.4682 | 0.6440 |
| DiffBIR [11] | 6.1069 | 4.1113 | 68.10 | **0.5527** | 0.6736 |
| SeeSR [22] | 5.2244 | **4.3317** | 69.70 | *0.5362* | **0.6887** |
| SinSR [17] | 6.4250 | 3.6613 | 62.78 | 0.4483 | 0.5854 |
| OSEDiff [21] | *4.6362* | *4.2251* | 68.88 | 0.4995 | 0.6062 |
| PURE [19] | 5.3617 | 4.2528 | *69.55* | 0.5215 | *0.6647* |
| **UARE (Ours)** | 4.6044 | 4.1804 | **70.05** | 0.5363 | 0.6446 |

Table B.3. Quantitative comparison of different methods on four single-degradation subset of FoundIR. For each method, the first row lists PSNR/LPIPS and the second row lists NIQE/MANIQA.

| Method | Blur | Haze | RainDrop | Lowlight |
|---|---|---|---|---|
| Restormer [29] | 21.53 / 0.3821 | 13.70 / 0.5687 | 24.63 / 0.3029 | 9.20 / 0.6037 |
| | 6.9553 / 0.1322 | 5.6395 / 0.2695 | 5.0340 / 0.2777 | 7.3095 / 0.2207 |
| PromptIR [14] | 21.64 / 0.4041 | 18.31 / 0.4762 | 26.67 / 0.2203 | 16.67 / 0.4804 |
| | 8.0897 / 0.1361 | 5.7593 / 0.2815 | 4.8115 / 0.2449 | 7.7780 / 0.2317 |
| DiffIR [11] | 21.61 / 0.3823 | 19.78 / 0.2345 | 26.19 / 0.2367 | 18.05 / 0.3146 |
| | 7.4800 / 0.1390 | 3.4547 / 0.3235 | **3.2762** / 0.2593 | **4.9243** / 0.2828 |
| DiffUIR [35] | **26.99** / *0.1912* | *20.50* / *0.2037* | 29.52 / *0.1292* | 14.88 / 0.2691 |
| | 6.1600 / *0.2734* | 3.7435 / 0.3540 | 4.0766 / 0.2697 | 6.0439 / 0.3615 |
| SUPIR [26] | 20.63 / 0.3180 | 13.66 / 0.3883 | 21.41 / 0.3552 | 7.43 / 0.6460 |
| | 4.8374 / 0.2702 | 3.8944 / 0.3482 | 4.7940 / *0.2933* | 7.6004 / 0.2680 |
| InstructIR [5] | 19.81 / 0.2335 | 17.27 / 0.2267 | 21.75 / 0.3456 | 20.78 / 0.2473 |
| | 6.1392 / 0.2394 | *3.4979* / 0.3663 | 4.7295 / 0.2935 | 4.9299 / 0.3184 |
| AutoDIR [6] | 19.98 / 0.3400 | 15.59 / 0.4130 | 21.46 / 0.3643 | **22.36** / 0.3487 |
| | 6.8772 / 0.1652 | 4.9741 / 0.2909 | 5.2210 / 0.2836 | 7.6178 / 0.2567 |
| FoundIR [8] | 26.10 / **0.1709** | **23.29** / 0.1896 | **30.86** / **0.0897** | *20.34* / *0.2499* |
| | *5.6797* / 0.2854 | 3.9543 / *0.3544* | *4.3657* / 0.2793 | 6.7511 / *0.3460* |
| **UARE (Ours)** | *22.38* / 0.1904 | 21.28 / **0.1635** | *28.26* / 0.0981 | 19.64 / **0.1841** |
| | **4.7313** / **0.3361** | **3.2674** / **0.4232** | 4.6515 / **0.3008** | *5.0119* / **0.4234** |

and PURE using five no-reference image quality metrics: NIQE, LIQE, MUSIQ, MANIQA, and TOPIQ. As reported in Tab. B.2, UARE ranks first on MUSIQ and second on NIQE and MANIQA, further demonstrating its effectiveness in challenging real-world scenarios.

**More qualitative comparisons.** Figs. B.1, B.2, B.3 and B.4 present visual comparisons across super-resolution images produced by these approaches. It can be seen that our method effectively restores fine image details, such as knots, text, and petals, while producing noticeably fewer artifacts. These results comprehensively confirm the effectiveness of UARE in image super-resolution.

## B.2. Image Restoration and Enhancement

**More quantitative results.** We have reported the restoration/enhancement results on the multi-degradation subsets of FoundIR [8] in the main paper. Here, we further compare UARE with Restormer [29], PromptIR [14], DiffIR [23], DiffUIR [35], SUPIR [26], InstructIR [5], AutoDIR [6], and FoundIR [8] on the single-degradation subsets of FoundIR, including blur, haze, raindrop, and low-light, as shown in Tab. B.3. UARE ranks first in MANIQA across all subsets. In addition, it achieves competitive PSNR, LPIPS, and NIQE results, indicating superior performance and a favorable trade-off between fidelity and perceptual quality.

Table B.4. User study results of different SR methods.

| Method | OSEDiff | S3Diff | PURE | **UARE (Ours)** |
|---|---|---|---|---|
| Total Votes | 25 | 48 | *31* | **186** |
| Voting Rate (%) | 8.62 | 16.55 | *10.69* | **64.14** |

Table B.5. User study results of different restoration methods.

| Method | DiffIR | DiffUIR | FoundIR | **UARE (Ours)** |
|---|---|---|---|---|
| Total Votes | 5 | *15* | 20 | **250** |
| Voting Rate (%) | 1.72 | *5.17* | 6.90 | **86.21** |

**More qualitative results.** Figs. B.5, B.6, B.7, B.8 and B.9 present visual comparison across restored/enhanced images produced by these approaches. It can be seen that UARE faithfully reconstructs challenging details, such as text in blurred or low-light regions, hair, and the fine structures of flowers and vegetation. These results comprehensively confirm the effectiveness of UARE across multiple image restoration and enhancement tasks.

## B.3. User Study

To further evaluate the effectiveness of our UARE, we conduct a user study comparing four SR and restoration methods, respectively. We employ ten LR images from the RealSR, DRealSR and DIV2K datasets, and ten LQ images from the FoundIR test set. Compared SR methods include OSEDiff [21], S3Diff [30] and PURE [19], while restoration methods include DiffIR [23], DiffUIR [35] and FoundIR [8]. Twenty-nine expert researchers are invited to choose the best super-resolution/restored image for each test sample based on two equally weighted criteria: (1) perceptual quality, focusing on clarity, detail, and realism, and (2) content consistency with the LR/LQ input, including alignment in image structure and texture.

As reported in Tab. B.4, UARE achieves a high voting rate of 64.14% in comparison with SR methods, which is significant better preference than other methods. Besides, as shown in Tab. B.5, UARE achieves a voting rate of 86.21% in comparison of different all-in-one restoration methods. These results show that users overwhelmingly prefer UARE over both SR and restoration baselines. In particular, UARE receives nearly four times as many votes as the best competing SR method and more than an order of magnitude more votes than the strongest all-in-one restoration baseline. These consistent user preferences demonstrate that UARE achieves a better balance between perceptual quality and content fidelity, as well as our unified IQA-and-restoration scheme.


Figure B.1. Visual comparison on the image named "Canon_043" from the RealSR dataset.

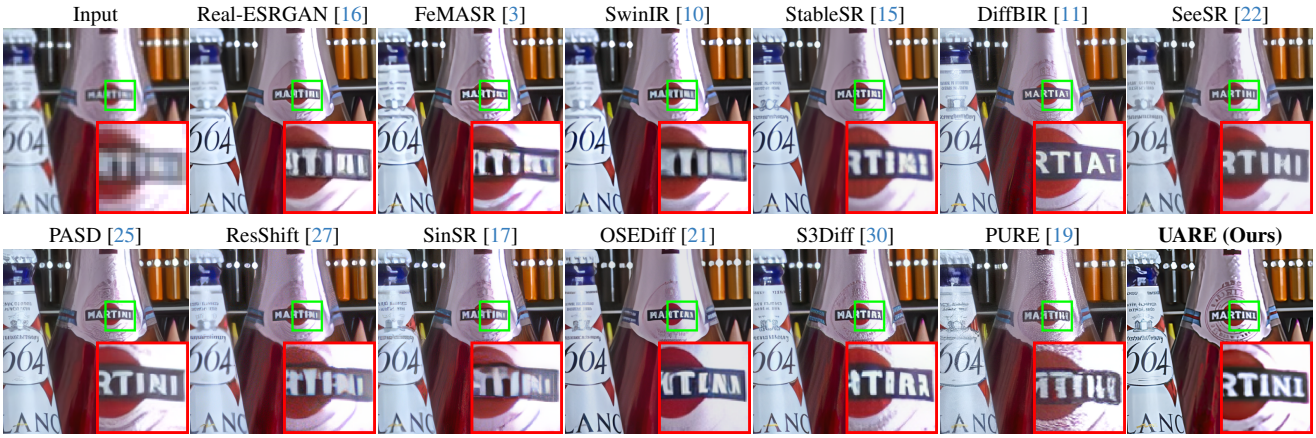


Figure B.2. Visual comparison on the image named "Canon_050" from the RealSR dataset.

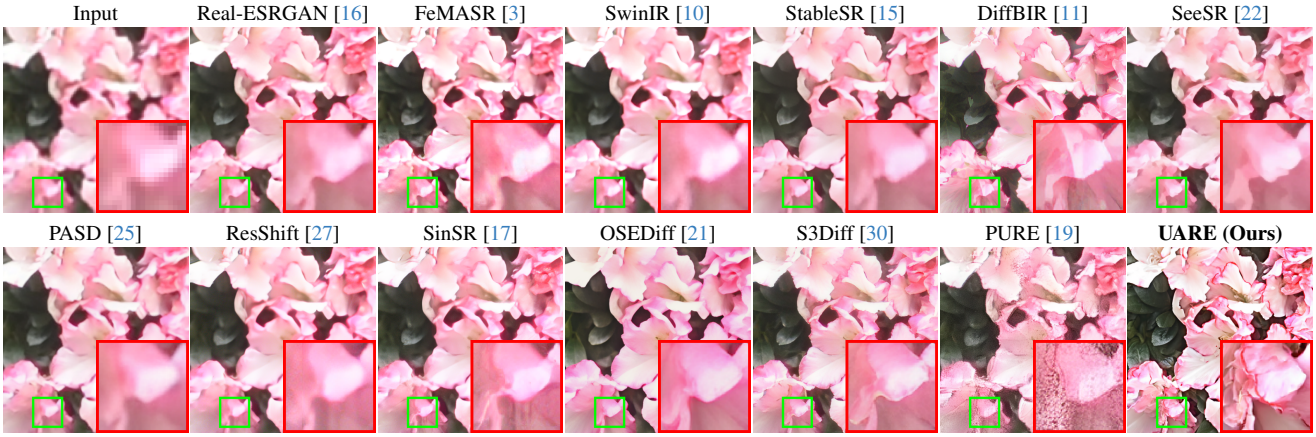


Figure B.3. Visual comparison on the image named "DSC_1425_x1" from the DRealSR dataset.

## C. Discussion and Limitations

Due to the large number of parameters in the unified model, UARE has a relatively large model size and slow inference speed, which limits its deployment on resource-constrained devices. In addition, although we demonstrate that IQA can boost restoration and enhancement performance, how restoration and enhancement, in turn, can better improve IQA remains an open question and requires further investigation.

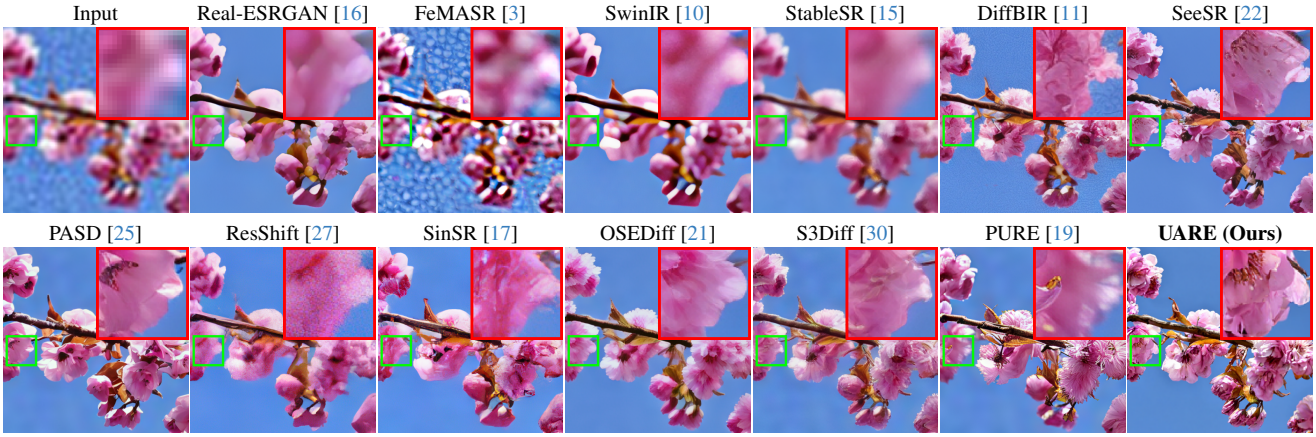


Figure B.4. Visual comparison on the image named "0000098" from the DIV2K dataset.

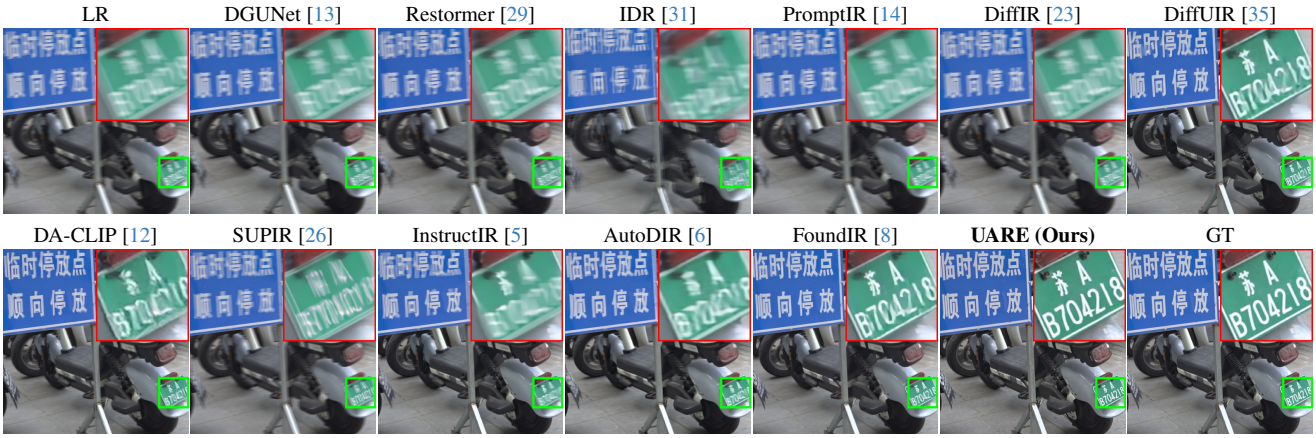


Figure B.5. Visual comparison on the image "0131" with blur from the FoundIR dataset.

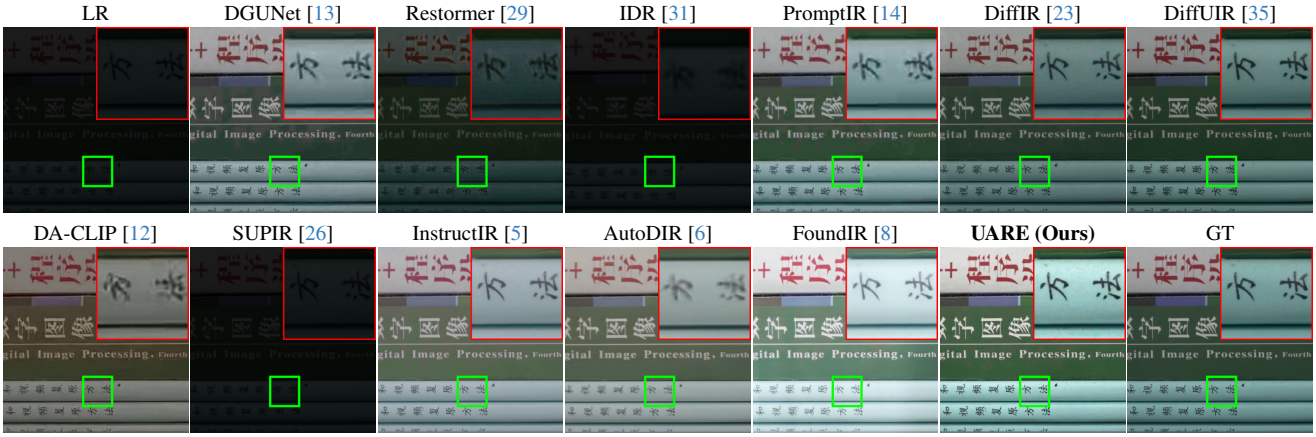


Figure B.6. Visual comparison on the image "1143" with low-light from the FoundIR dataset.

## References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 126–135, 2017. 2

[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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 3086–3095, 2019. 1

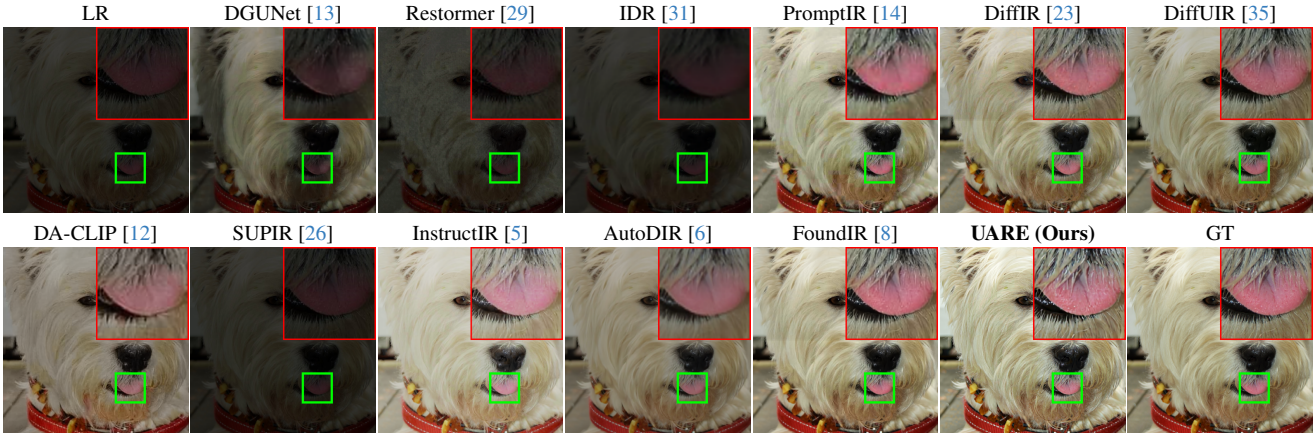


Figure B.7. Visual comparison on the image "1304" with low-light and JPEG compression from the FoundIR dataset.

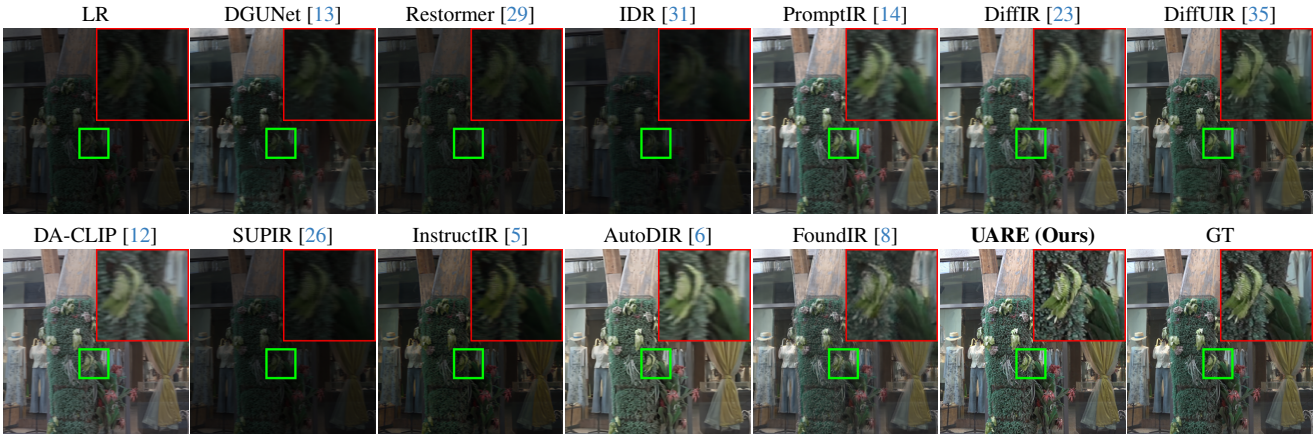


Figure B.8. Visual comparison on the image "1397" with low-light, blur and noise from the FoundIR dataset.

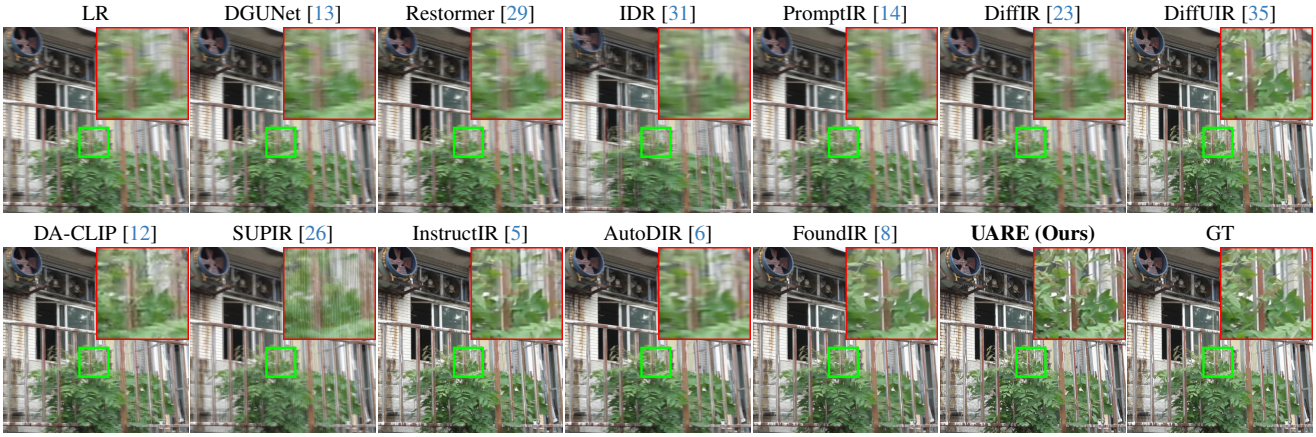


Figure B.9. Visual comparison on the image "0243" with blur and JPEG compression from the FoundIR dataset.

[3] Chaofeng Chen, Xinyu Shi, Yipeng Qin, Xiaoming Li, Xiaoguang Han, Tao Yang, and Shihui Guo. Real-world blind super-resolution via feature matching with implicit high-resolution priors. In *Proceedings of the ACM International Conference on Multimedia (ACM MM)*, pages 1329–1338, 2022. 2, 3, 5, 6

[4] 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 Transactions on Image Processing (TIP)*, 33:2404–2418, 2024. 2

[5] Marcos V Conde, Gregor Geigle, and Radu Timofte. In-

structir: High-quality image restoration following human instructions. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. 4, 6, 7

[6] Yitong Jiang, Zhaoyang Zhang, Tianfan Xue, and Jinwei Gu. Autodir: Automatic all-in-one image restoration with latent diffusion. In *Proceedings of the European Conference on Computer Vision (ECCV)*, 2024. 4, 6, 7

[7] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 5148–5157, 2021. 2

[8] Hao Li, Xiang Chen, Jiangxin Dong, Jinhui Tang, and Jinshan Pan. Foundir: Unleashing million-scale training data to advance foundation models for image restoration. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. 4, 6, 7

[9] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2021. 2, 3

[10] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 1833–1844, 2021. 5, 6

[11] Xinqi Lin, Jingwen He, Ziyan Chen, Zhaoyang Lyu, Bo Dai, Fanghua Yu, Yu Qiao, Wanli Ouyang, and Chao Dong. Diffbir: Toward blind image restoration with generative diffusion prior. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 430–448, 2024. 2, 3, 4, 5, 6

[12] Ziwei Luo, Fredrik K Gustafsson, Zheng Zhao, Jens Sjölund, and Thomas B Schön. Controlling vision-language models for universal image restoration. In *Proceedings of the International Conference on Learning Representations (ICLR)*, 2024. 6, 7

[13] Chong Mou, Qian Wang, and Jian Zhang. Deep generalized unfolding networks for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 6, 7

[14] Vaishnav Potlapalli, Syed Waqas Zamir, Salman H Khan, and Fahad Shahbaz Khan. Promptir: Prompting for all-in-one image restoration. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 4, 6, 7

[15] Jianyi Wang, Zongsheng Yue, Shangchen Zhou, Kelvin CK Chan, and Chen Change Loy. Exploiting diffusion prior for real-world image super-resolution. *International Journal of Computer Vision (IJCV)*, 132(12):5929–5949, 2024. 2, 3, 4, 5, 6

[16] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2, 3, 5, 6

[17] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25796–25805, 2024. 2, 3, 4, 5, 6

[18] Zhou Wang, Alan C Bovik, Hamid R Sheikh, and Eero P Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing (TIP)*, 13(4):600–612, 2004. 2

[19] Hongyang Wei, Shuaizheng Liu, Chun Yuan, and Lei Zhang. Perceive, understand and restore: Real-world image super-resolution with autoregressive multimodal generative models. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, 2025. 2, 3, 4, 5, 6

[20] 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 *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 101–117. Springer, 2020. 1

[21] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. In *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 2024. 2, 3, 4, 5, 6

[22] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 25456–25467, 2024. 2, 3, 4, 5, 6

[23] 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. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 4, 6, 7

[24] 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 *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1191–1200, 2022. 2

[25] Tao Yang, Rongyuan Wu, Peiran Ren, Xuansong Xie, and Lei Zhang. Pixel-aware stable diffusion for realistic image super-resolution and personalized stylization. In *European conference on computer vision*, pages 74–91, 2024. 2, 3, 5, 6

[26] Fanghua Yu, Jinjin Gu, Zheyuan Li, Jinfan Hu, Xiangtao Kong, Xintao Wang, Jingwen He, Yu Qiao, and Chao Dong. Scaling up to excellence: Practicing model scaling for photo-realistic image restoration in the wild. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 4, 6, 7

[27] Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Proceedings of the Advances in Neural Information Processing Systems (NeurIPS)*, 36, 2024. 2, 3, 5, 6

[28] Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inver-

sion. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 23153–23163, 2025. 2, 3

[29] Syed Waqas Zamir, Aditya Arora, Salman Khan, Munawar Hayat, Fahad Shahbaz Khan, and Ming-Hsuan Yang. Restormer: Efficient transformer for high-resolution image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 4, 6, 7

[30] 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, 3, 4, 5, 6

[31] Jinghao Zhang, Jie Huang, Mingde Yao, Zizheng Yang, Hu Yu, Man Zhou, and Feng Zhao. Ingredient-oriented multi-degradation learning for image restoration. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 6, 7

[32] Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 4791–4800, 2021. 4

[33] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing (TIP)*, 24(8):2579–2591, 2015. 2

[34] Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14071–14081, 2023. 2

[35] Dian Zheng, Xiao-Ming Wu, Shuzhou Yang, Jian Zhang, Jian-Fang Hu, and Wei-Shi Zheng. Selective hourglass mapping for universal image restoration based on diffusion model. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 4, 6, 7