Augmenting Perceptual Super-Resolution via Image Quality Predictors

Supplementary Material

7. Complete Analysis of NR-IQA metrics

In §3.1 of the main paper, we present accuracy of only top 7 NR-IQA metrics on the subset of SBS180K [43] train set. Here, in Table 6, we present accuracy of 20 NR-IQA metrics and their variants (42 in total) on the same subset. We make two main observations. First, unsurprisingly, recent NR-IQA metrics (e.g. PaQ-2-PiQ [97], MUSIQ [41], Q-Align [88]) are more aligned with human preferences than the classical ones (e.g. NIQE [98] and BRISQUE [59]), calling for wider adaptation of more recent metrics in evaluating SR models. Second, the IQA dataset on which the metric is trained affects its accuracy in determining human preference for SR. For instance, TOPIQ [11] trained using KonIQ [35] dataset is more aligned with human judgement (73.06%) than the one trained using FLIVE [97] dataset (58.19%). Results indicate an opportunity to create a no-reference IQA dataset exclusively for training NR metrics for SR.

7.1. Remark on NR-IQA Choices

As discussed in §3, our choice of MUSIQ for weighted sampling and fine-tuning comes from several considerations. First, on SBS-180K, MUSIQ is highly performant (see Phase II analysis of §3.1). Second, on HGGT (§3.2), MUSIQ has the best positive misalignment, meaning it is the least likely to misrank a positive. It is also relatively efficient for both inference and back-propagation.

In our application, distinguishing between the quality of positives would seem to be more significant, since we are ideally training the SR model in a manner that focuses on the highest quality images (i.e., incorrect ordering of the lower-ranked images will not affect our method; hence, fine-grained differentiation between the high-ranked images is more important). Nevertheless, it is true that MUSIQ (like all NR-IQA models evaluated here) does not perform well on negative misalignment. However, we do not expect this to have a large impact on training, due to the *rarity* of negatives. Specifically, in HGGT-train, only $\sim$6% of tuples contain negatives and, among those, MUSIQ ranks a negative the highest in $\sim$34% of cases. Thus, our AMO model will be exposed to a negative in only $\sim$2% of examples. Hence, merely for numerical magnitude, discernment for positives is likely to be more impactful than for negatives. Of course, this reasoning is somewhat specific to the HGGT setup.

Further, in terms of evaluation, note that we choose NIMA and Q-Align specifically because they perform best on the SBS-180K samples on which MUSIQ fails. Ideally, this complementarity would help ensure that errors induced by shortcomings of MUSIQ could potentially be detected by the other NR-IQA metrics. Nevertheless, as seen in Table 7, our experiments with PaQ-2-PiQ (which was among the best models according to Table 2) show MUSIQ outperforms it.

Regardless, our method does not specifically require the use of MUSIQ. Indeed, we believe further advancements in NR-IQA models (e.g., approaches specific to SR image quality, adversarially robust models) will be applicable to our method as well.

8. Methodological Details

8.1. Sampling Details

The altered sampling (§4.2) is trained identically to the standard HGGT version, just replacing the uniform nature of the GT sampling. The only additional parameter is the temperature, τ, which we set to 10 for both SMA and SMP.

8.2. Hardware and Timing

Similar to HGGT, we train on four A100 GPUs for 300K iterations. This takes $\sim$23 and $\sim$32 hours for SwinIR and RealESRGAN, respectively, with an additional $\sim$3.5 hours for fine-tuning.

8.3. Fine-Tuning Details

Unless otherwise noted, we use the same training parameters as HGGT. We fine-tune for only 20,000 steps and set $\lambda_Q = 0.05$ as the FT MUSIQ weight. For SwinIR and RealESRGAN, respectively, we change the learning rate to 5×10^{-6} (halved at 5K steps) and 5×10^{-5}. Recall that, by default, the adversarial loss is not used (i.e., $\lambda_A = 0$) during FT (but see §9). Architecturally, LoRA weights are inserted slightly differently: on SwinIR [50], only the multilayer perceptions are altered (rank 48), while on the convolutional RealESRGAN [84, 85], only the layers in the Residual-in-Residual Dense Blocks (RRDBs) are altered (rank 24). This follows other works, including the original LoRA paper [36], which only apply LoRA-based fine-tuning to a subset of layers (e.g., see [26, 46, 53]). Recall that LoRA cannot increase the capacity (i.e., expressive capability) of the networks (as the new weights can simply be merged into the old ones at inference time, which also prevents any run-time penalty to inference), so comparisons to non-FT models are fair. Fine-tuning is run for 20K steps, as opposed to the 300K in stage two training. Brief exploration of hyper-parameters (beyond those considered in §5)

Method	Acc (%)	Method	Acc (%)	Method	Acc (%)	Method	Acc (%)	Method	Acc (%)
paq2piq	76.41	arniqa-kadid	71.48	tres	69.98	arniqa-clive	66.81	brisque_matlab	61.00
nima	74.91	arniqa-flive	71.30	clipiqa+_vitL14_512	69.98	arniqa-spaq	66.73	wadiqam_nr	60.30
musiq	74.47	topiq_nr-spaq	71.30	musiq-paq2piq	69.98	arniqa	66.46	topiq_nr-flive	58.19
liqe	74.03	arniqa-csiq	71.21	maniqa-pipal	69.63	musiq-ava	66.37	ilniqe	57.92
arniqa-tid	74.03	musiq-spaq	70.86	clipiqa+_rn50_512	69.10	nrqm	65.05	niqe	56.43
qalign	73.77	nima-vgg16-ava	70.77	dbcnn	68.49	cnniqa	63.73	brisque	55.11
topiq_nr	73.06	maniqa	70.51	clipiqa	68.40	tres-flive	63.29	niqe_matlab	51.94
hyperiqa	72.27	clipiqa+	70.25	arniqa-live	68.05	pi	62.41	piqe	46.21
liqe_mix	71.48	maniqa-kadid	70.16						

Table 6. **Phase I analysis on SBS180K dataset.** Accuracy of 20 NR-IQA metrics and their variants on the subset (1212 image pairs) of train set of SBS180K dataset. We denote a metric by its 'Model Name' as defined in `IQA-PyTorch` toolbox (`https://iqa-pytorch.readthedocs.io/en/latest/ModelCard.html`). We use the default configuration for all metrics and their variants.

Model	NR	λ_A		FR Low-Lev. Dist.		FR Mid-Lev. Dist.			NR High-Lev. Perceptual Quality			
				PSNR ↑	SSIM ↑	LPIPS ↓	LPIPS-ST ↓	DISTS ↓	MUSIQ ↑	NIMA ↑	Q-Align ↑	TOPIQ ↑
Gold Standard	–	–	✗	–	–	–	–	–	69.64	5.28	3.78	0.69
SwinIR-UPos*	–	–	✗	22.30	0.647	0.169	0.129	0.123	66.39	5.16	3.56	0.62
SwinIR-UPos + $\mathrm{FT}_{\mathrm{HP}}$	M	0	✗	22.17	0.642	0.166	0.123	0.122	68.38	5.23	3.64	0.65
SwinIR-UPos + $\mathrm{FT}_{\mathrm{IG}}$	M	0.1	✗	22.03	0.635	0.168	0.122	0.123	69.37	5.24	3.69	0.66
SwinIR-UPos + $\mathrm{FT}_{\mathrm{NNR,IG}\times 2}$	–	0.2	✗	22.25	0.646	0.171	0.130	0.124	66.61	5.16	3.56	0.61
SwinIR-UPos + $\mathrm{FT}_{\mathrm{NNR,IG}\times 5}$	–	0.5	✗	22.20	0.644	0.174	0.134	0.125	66.61	5.16	3.56	0.61
SwinIR-UPos + $\mathrm{FT}_{\mathrm{PaQ2PiQ}}$	P	0	✗	22.29	0.649	0.166	0.120	0.121	67.29	5.18	3.58	0.62
SwinIR-UPos + FT	M	0	✗	22.01	0.633	0.169	0.123	0.124	69.70	5.26	3.70	0.67
SwinIR-AMO + FT	M	0	✓	21.77	0.624	0.174	0.121	0.128	70.81	5.29	3.75	0.70
RESRGAN-UPos*	–	–	✗	21.54	0.608	0.233	0.192	0.158	65.93	5.25	3.47	0.63
RESRGAN-UPos + $\mathrm{FT}_{\mathrm{HP}}$	M	0	✗	21.30	0.595	0.226	0.175	0.158	70.28	5.32	3.65	0.69
RESRGAN-UPos + $\mathrm{FT}_{\mathrm{IG}}$	M	0.1	✗	21.14	0.586	0.236	0.182	0.160	72.01	5.35	3.70	0.70
RESRGAN-UPos + $\mathrm{FT}_{\mathrm{NNR,IG}\times 2}$	–	0.2	✗	21.35	0.600	0.234	0.191	0.157	65.94	5.22	3.45	0.63
RESRGAN-UPos + $\mathrm{FT}_{\mathrm{NNR,IG}\times 5}$	–	0.5	✗	21.25	0.598	0.237	0.195	0.158	65.78	5.22	3.46	0.63
RESRGAN-UPos + $\mathrm{FT}_{\mathrm{PaQ2PiQ}}$	P	0	✗	21.46	0.605	0.228	0.182	0.157	67.26	5.22	3.51	0.64
RESRGAN-UPos + FT	M	0	✗	21.09	0.580	0.235	0.179	0.163	72.69	5.37	3.69	0.71
RESRGAN-AMO + FT	M	0	✓	21.02	0.581	0.228	0.169	0.161	71.67	5.35	3.68	0.71

Table 7. **Additional evaluation on held-out HGGT Test-100.** As in Table 5 in the main paper, "FR Low-Lev Dist" refers to full-reference low-level distance metrics; "FR Mid-Lev Dist" and "NR High-Lev. Perceptual Quality" refer to full-reference and no-reference perceptual metrics, respectively. Second column ( ) indicates that a method works with *no human GT ranking data* (✓), or requires such GT annotations (✗). "Gold Standard" shows the average of best metric value per quintuplet of test GTs. "UPos" denotes the "positives-only" scenario (uniform sampling from human-ranked positives), the SoTA baseline method from HGGT (marked by *). "FT" refers to fine-tuning (direct optimization): "$\mathrm{FT}_{\mathrm{IG}}$" includes the adversarial loss during FT, "$\mathrm{FT}_{\mathrm{NNR,IG}\times 2}$" and "$\mathrm{FT}_{\mathrm{NNR,IG}\times 5}$" have no NR term during FT, but increase the GAN loss (two and five times, respectively), and finally "$\mathrm{FT}_{\mathrm{PaQ2PiQ}}$" replaces MUSIQ with PaQ-2-PiQ. The NR column denotes which NR-IQA model is used (M: MUSIQ, P: PaQ-2-PiQ, –: None), while λ_A is the adversarial loss weight (the standard HGGT default for training is 0.1). We also show our best method: AMO+FT, which combines IQA-based sampling with our standard FT settings, for comparison. Note that AMO+FT is the only method here that *does not use human annotations*. We remark also that the NR-IQA models have the following ranges: MUSIQ (0-100), NIMA (0-10), Q-Align (1,5), and TOPIQ (0-1).

yielded minimal changes, likely due to rapid convergence of the low-rank (i.e., low capacity) weights ϕ.

9. Detailed Results on Ablations and Variations

In this section, we consider additional FT variations: (i) using a GAN discriminator instead of an NR-IQA model (using two different loss weights) and (ii) replacing Q (set as MUSIQ) with a different NR-IQA model (PaQ-2-PiQ). The point of (i) is to check whether the GAN critic, which is effectively an NR-IQA model that has been specialized to the SR model in question, can be used for fine-tuning, instead of a separate NR-IQA model. For (ii), we wish to check if our choice of optimized NR metric, MUSIQ, is reasonable.

Our results on these variations are in Table 7. Since FT optimizes MUSIQ, we focus on the other NR metrics, especially Q-Align and NIMA (since they perform the best on examples where MUSIQ fails; see §3.1). First, we find that including the GAN loss in the standard scenario has a slight negative effect on the NR metrics; however, removing the NR metric term and strengthening the adversarial term (i.e., "$\mathrm{FT}_{\mathrm{NNR,IG}\times 2}$" and "$\mathrm{FT}_{\mathrm{NNR,IG}\times 5}$") has a significantly more negative impact on the NR evaluations. This suggests that

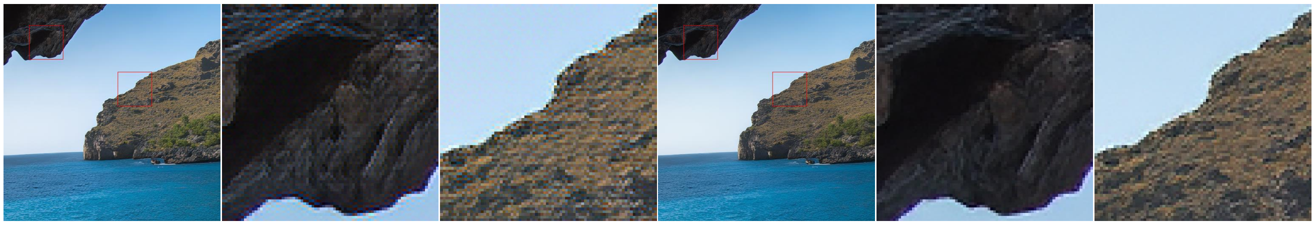

Figure 5. **Structured Noise due to naive NR-IQA optimization.** The left three insets show an image and two close-ups that was fine-tuned *without* LoRA, whereas the right three show the effect of using LoRA. Note the patterns that form in the sky and the strangely coloured pixels that appear around certain edges (e.g., the blue/red grid in the second inset) when LoRA is not used.

the critic network *cannot* replace the NR-IQA model, even though it is intuitively similar to one (in that it evaluates the image quality of a single input, which can be used as a learning signal). We conjecture this is because the critic is trained to detect the idiosyncrasies of its associated generator (at a specific point in time), rather than match human quality estimates; hence, optimizing it more aggressively may reduce those specific issues that the critic has detected, but not necessarily increase general quality.

Second, we tried to replace MUSIQ with PaQ-to-PiQ. We find that this tends to improve low and mid level distortion (though the relation is less clear for RealESRGAN, especially with LPIPS-ST), but worsens NIMA and Q-Align. We therefore choose to stay with MUSIQ for our main results. In general, we do not wish to claim that MUSIQ is an optimal starting point for FT; however, it does suggest our analysis is a useful approach to initially identifying a good NR-IQA network. Nevertheless, we suspect that using an alternative NR-IQA model (with sufficient hyper-parameter exploration), fine-tuning a new model, combining multiple models, or training a model specific to SR could all be potentially useful future approaches to improving results.

10. Additional Qualitative Examples

10.1. Additional Comparative Samples

Additional comparisons are shown in Fig. 6 (as in Fig. 4). Our method (AMO or AMO+FT) is universally sharper and more detailed than UPos (e.g., see the hair in row three). Further, it can occasionally remove some of the noise present in the UPos scenario (see the tongue of the red panda). Importantly, our approach may not generate details that are identical to the GT, but it does construct sharp image content without jarring unrealistic artifacts (e.g., see rows one and four; the plants, rocks, and bricks have slightly different details, but they are plausible and of similar aesthetic quality nonetheless).

10.2. Additional Naive Optimization Visualizations

In Fig. 5, as in Fig. 3, we show the subtle "grid-like" artifacts that appear when naive NR-IQA optimization is performed. In particular, we see spatial patterns form in homogeneous areas (e.g., stripes in the sky or on the tan coloured island), while other areas exhibit highly unnatural colours (e.g., the alternating blue-red pixels on the dark rock). These small, pixel-scale artifacts are akin to an adversarial attack on MUSIQ; hence, much of this structured noise is alleviated by applying LoRA (right insets). Other methods of handling such artifacts, such as an adversarially robust NR-IQA model, may also be effective, but we leave this to future work.

11. Additional Results on RealSR

We provide results on the RealSRv3 [9] dataset in Table 8. Similar to the HGGT test dataset, we find that *our method is superior in terms of every NR-IQA metric*, at the expense of the exact pixel-level details measured by PSNR and SSIM (following the perception-distortion tradeoff [5]). However, according to mid-level FR metrics, our method *also* performs well, obtaining the best scores on LPIPS-ST, and even on LPIPS and DISTS for SwinIR. This suggests our method can improve image quality, while maintaining the most salient perceptual details (e.g., mid-level textures) of the underlying GT.

12. Comparative Evaluation via User Study

Similar to the HGGT user study, we invite 12 volunteers to evaluate their preference between SwinIR-AMO+FT and SwinIR-UPos, using the HGGT Test-100 dataset. Each volunteer evaluates 25 image pairs (25% of the dataset), with each image in Test-100 being seen an equal number of times (namely, three). For each pair (SwinIR-AMO+FT vs. SwinIR-UPos), we employ an image comparison slider. This tool places two images on top of each other, and allows volunteers to use a slider to alternate between them (see Fig. 7 for a visualization). The order of presentation of the two methods (left vs. right) is randomized to eliminate bias. For each individual, we obtain a single score, which is the percentage of the time that they prefer our method (across those 25 images). The average score across raters is **69.7%** (median: 68.0%; empirical standard error of the

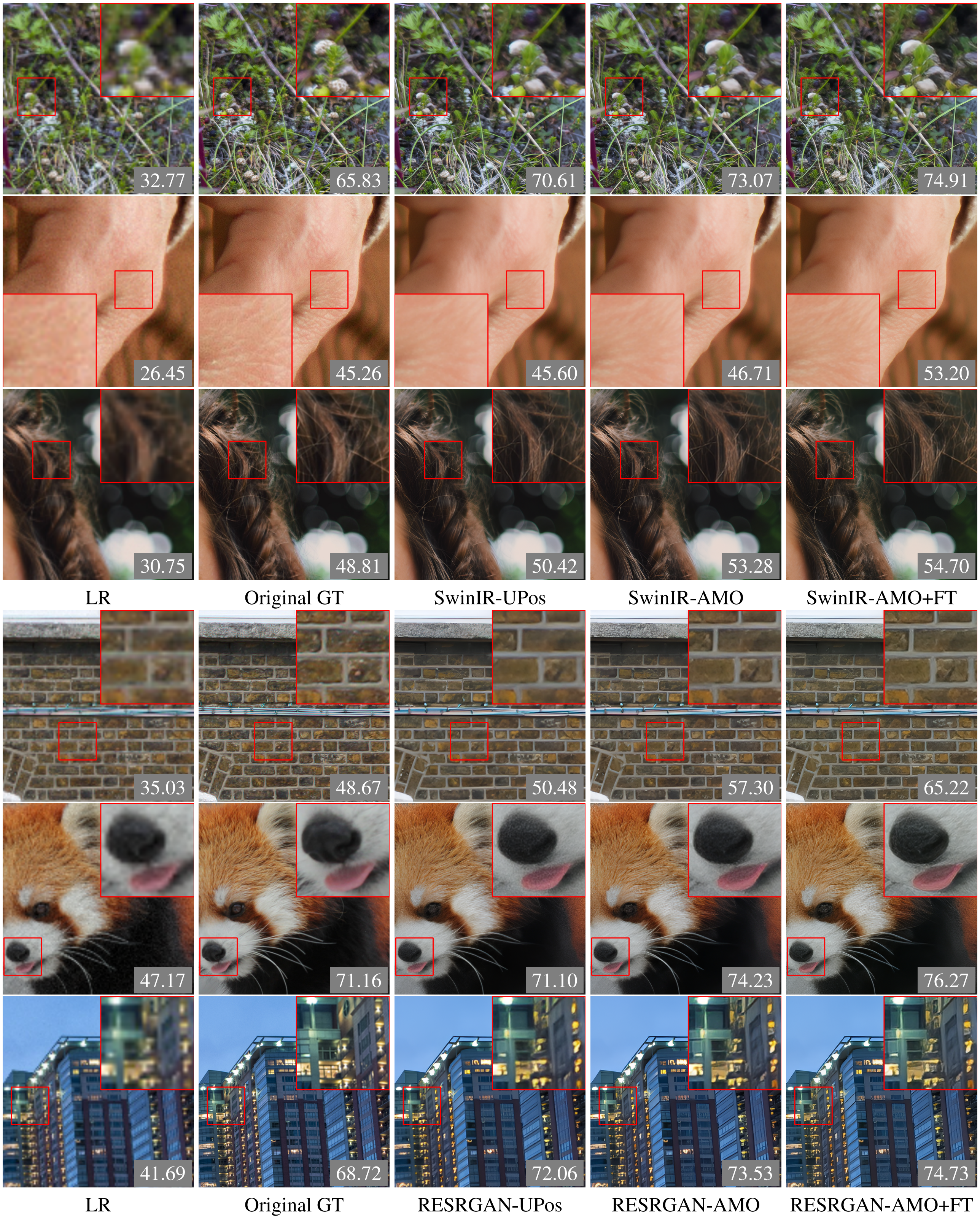

Figure 6. **Qualitative results with NR-IQA guidance.** Following the notation of Table 5, columns 3-5 are (top 3 rows) SwinIR-UPos, SwinIR-AMO, and SwinIR-AMO + FT, and (bottom 3 rows) Real-ESRGAN-UPos, Real-ESRGAN-AMO, and Real-ESRGAN-AMO + FT. We show MUSIQ scores in insets. Qualitatively, we see improved performance as we move across the 'UPos', 'AMO', and 'AMO-FT' methods, particularly in terms of sharpness and detail generation. Zoom in for details.

Model		FR Low-Lev. Dist.		FR Mid-Lev. Dist.			NR High-Lev. Perceptual Quality			
		PSNR ↑	SSIM ↑	LPIPS ↓	LPIPS-ST ↓	DISTS ↓	MUSIQ ↑	NIMA ↑	Q-Align ↑	TOPIQ ↑
SwinIR-OrigsOnly	✓	**26.05**	0.746	0.37	0.38	0.20	31.37	4.24	2.88	0.23
SwinIR-UPos*	✗	26.02	**0.747**	0.35	0.37	0.20	33.69	4.31	2.95	0.24
SwinIR-AMO	✓	25.99	**0.747**	0.34	0.37	**0.19**	34.86	4.32	2.96	0.25
SwinIR-AMO + FT	✓	25.96	0.742	**0.33**	**0.35**	**0.19**	**39.25**	**4.37**	**2.99**	**0.30**
RESRGAN-OrigsOnly	✓	**25.90**	**0.758**	**0.27**	0.27	**0.16**	46.11	4.80	3.40	0.32
RESRGAN-UPos*	✗	25.45	0.750	0.28	0.26	0.17	52.74	4.95	3.53	0.41
RESRGAN-AMO	✓	25.22	0.745	0.28	0.25	0.17	54.73	4.97	3.57	0.45
RESRGAN-AMO + FT	✓	24.71	0.718	0.32	**0.24**	0.19	**65.12**	**5.03**	**3.77**	**0.63**

Table 8. **Additional evaluation on the RealSRv3 [9].** Following Table 7, we evaluate the four main models on the RealSR V3 dataset, which consists of 100 test images captured using two DSLR cameras (Canon 5D3 and Nikon D810). Our methods ("AMO" and "AMO + FT") achieve the highest no-reference perceptual metric (i.e., NR-IQA) scores, outperforming both "OrigsOnly" (without enhanced GT) and "UPos" (the SOTA baseline from HGGT, marked by *).

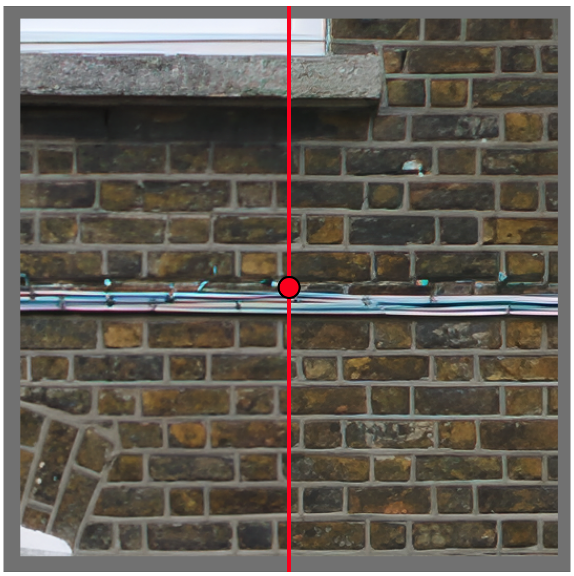

Figure 7. **User study example.** Users can move the slider to alternate between 2 images.

mean: 4.8%), suggesting our algorithm is preferred over the HGGT-based UPos approach at a more than 2:1 ratio, *despite their use of human annotations*, which ours does not use. Following similar image quality assessment protocols (e.g., [80]), a simple single-sample one-sided t-test finds the rater mean significantly above 50% ($p < 0.01$; 95% confidence interval: $[60.2\%, 79.1\%]$).

13. Remark on Evaluation Metric Types and Nomenclature

The perception-distortion tradeoff [5] necessitates a complex suite of evaluation metrics that consider different aspects of the SR outputs, including pixel-level fidelity to a GT image and standalone image quality. Some works (e.g., [90]) even utilize performance on downstream vision tasks (e.g., detection or segmentation) as a form of checking semantic preservation. In this work, we therefore also include a continuum of metrics, which we hope will cover various points along the perception-distortion frontier. These metrics are often categorized along two different axes: (i) the use of a reference and (ii) the level of visual abstraction (low vs mid vs high).

NR vs FR. The first form of metric categorization is full-reference (FR) vs no-reference (NR). In general, FR metrics (which have access to a GT) measure distortion, while NR metrics (which do not use a GT) measure perceptual quality. For NR metrics, there is no way to measure distortion; however, there are many different aspects of perceptual quality that can be considered, ranging from simple sharpness to differentiating aesthetic vs technical quality (e.g., [87]). Hence, it is common (e.g., [89, 90, 95]) to use a set of NR-IQA models, which presumably complement each other, as we do (see §3 and §7.1 for the discussion behind our metric choices). For FR metrics, there is more of a spectrum (i.e., they can include some aspects of perceptual information, in addition to measuring distortion). PSNR and other per-pixel distances have no notion of perception, operating directly on pixel values. SSIM is meant to be more perceptual, but is a simple, hand-crafted similarity operating on colours, limiting its perceptual modelling capabilities [60]. In contrast, LPIPS and DISTS utilize neural network features, aiming to capture certain aspects of human vision. They are therefore *more perceptual* than, e.g., PSNR, as they will tolerate some pixel differences (distortion) if they improve network activation similarity. Even further along this curve towards greater perceptual sensitivity is LPIPS-ST, a model designed specifically to ignore small spatial shifts (which are devastating to pixel-level distortion measures). Indeed, in many cases, we find that LPIPS-ST actually agrees with the NR-IQA perceptual metrics more closely than LPIPS or DISTS, despite being an FR metric. Hence, FR metrics can occupy a range across the perception-distortion curve.

Abstraction Level. A separate nomenclature arises based on the *type of information* that impacts the model. It is based on the hierarchical nature of *biological* vision (e.g., [64]), but is also commonly used throughout computer vision (e.g., [25]). Specifically, we divide visual processes into *low-level*, relating to raw colours and 2D geometry (e.g., edges); *mid-level*, encompassing "groupings" of more basic features into patterns and textures, as well as local 3D structures; and *high-level*, pertaining to semantics (e.g., scene classification) and representational abstraction (e.g., holistic interpretations of the image). For this reason, we refer to PSNR and SSIM, which operate directly on colours, as low-level, while LPIPS and DISTS are mid-level, as they respond best to textures, image "styles", and other regional "grouped" visual elements. We label neural NR-IQA models, such as MUSIQ, as high-level, as they process the image holistically, taking semantic context into account, as well as aesthetics, though they may also care about low-level issues, such as noise and blur. In general, including in our work, low-level metrics tend to measure distortion, while mid-level and high-level ones are more related to perceptual quality. However, there may be exceptions: for instance, measuring sharpness via a simple image filter is a low-level NR metric that targets perceptual quality rather than distortion (e.g., [78]).

14. Limitations

While our IQA-based method is able to sharpen SR outputs, as well as hallucinate aesthetically pleasing details in most cases, there are still several shortcomings to our approach. First, higher IQA model score does not guarantee improved human perceptual quality nor does it strictly ensure our outputs are artifact-free. This is related to the discussion in §4.3 and Fig. 3, where we postulate that some image changes can improve IQA score despite worsening perceptual quality (e.g., direct optimization being similar to an adversarial attack on the quality model). In Fig. 8, row two, for instance, we see that the SR model fails to predict the correct image details, leading to incorrect line orientations and aliasing-like artifacts (though the UPos baseline in column two arguably has worse artifacts). Second, from a semantic perspective, certain classes of image content may require different treatment, the requirements of which NR-IQA models are not naturally aware. For example, row one in Fig. 8 demonstrates how super-resolved *text* can become mangled. In terms of human preference, it can be argued that having a blurrier output in such uncertain cases may be more desirable (i.e., having blurred characters, rather than *wrong* characters, could be preferred for text). Nevertheless, text is notoriously challenging to super-resolve (prompting development of specialized methods for it [49, 103]); further, the UPos baseline suffers from similar artifacts as our outputs. Overall, we suspect better IQA models or more sophisticated regularized optimizations (i.e., beyond LoRA) can mitigate some of the artifacts incurred by our approach. Handling more semantic issues, such as text hallucination, may require more specialized models.

Figure 8. **Illustration of limitations.** We show examples of shortcomings of our method (see Fig. 14), with MUSIQ scores in insets. In row one, we show the shortcomings of our model with respect to text, a particularly difficult form of image content. In row two, we see that our model does still incur artifacts, such as the mangled lines in the zoomed inset.

References

[1] Lorenzo Agnolucci, Leonardo Galteri, Marco Bertini, and Alberto Del Bimbo. ARNIQA: Learning distortion manifold for image quality assessment. In *Winter Conference on Applications of Computer Vision (WACV)*, 2024. 3, 5

[2] Namhyuk Ahn, Byungkon Kang, and Kyung-Ah Sohn. Fast, accurate, and lightweight super-resolution with cascading residual network. In *European Conference on Computer Vision (ECCV)*, 2018. 2

[3] Yuval Bahat and Tomer Michaeli. Explorable super resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[4] Simon Baker and Takeo Kanade. Limits on super-resolution and how to break them. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2002. 1

[5] Yochai Blau and Tomer Michaeli. The perception-distortion tradeoff. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2, 8, 3, 5

[6] Yochai Blau and Tomer Michaeli. Rethinking lossy compression: The rate-distortion-perception tradeoff. In *International Conference on Machine Learning (ICML)*, 2019. 2, 8

[7] Yochai Blau, Roey Mechrez, Radu Timofte, Tomer Michaeli, and Lihi Zelnik-Manor. The 2018 PIRM challenge on perceptual image super-resolution. In *European*

Conference on Computer Vision Workshops (ECCVW), 2018. 3

[8] Sebastian Bosse, Dominique Maniry, Klaus-Robert Müller, Thomas Wiegand, and Wojciech Samek. Deep neural networks for no-reference and full-reference image quality assessment. *IEEE Transactions on Image Processing (TIP)*, 2017. 3

[9] 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 *International Conference on Computer Vision (ICCV)*, 2019. 3, 5

[10] Chaofeng Chen and Jiadi Mo. IQA-PyTorch: Pytorch toolbox for image quality assessment. [Online]. Available: https://github.com/chaofengc/IQA-PyTorch, 2022. 3

[11] 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)*, 2024. 3, 5, 6, 1

[12] Du Chen, Jie Liang, Xindong Zhang, Ming Liu, Hui Zeng, and Lei Zhang. Human guided ground-truth generation for realistic image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2, 3, 4, 6

[13] Liangyu Chen, Xiaojie Chu, Xiangyu Zhang, and Jian Sun. Simple baselines for image restoration. In *European Conference on Computer Vision (ECCV)*, 2022. 2

[14] Xiangyu Chen, Xintao Wang, Jiantao Zhou, Yu Qiao, and Chao Dong. Activating more pixels in image super-resolution transformer. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2

[15] Zheng Chen, Zongwei Wu, Eduard Zamfir, Kai Zhang, Yulun Zhang, Radu Timofte, Xiaokang Yang, Hongyuan Yu, Cheng Wan, Yuxin Hong, Zhijuan Huang, Yajun Zou, Yuan Huang, Jiamin Lin, Bingnan Han, Xianyu Guan, Yongsheng Yu, Daoan Zhang, Xuanwu Yin, Kunlong Zuo, Jinhua Hao, Kai Zhao, Kun Yuan, Ming Sun, Chao Zhou, Hongyu An, Xinfeng Zhang, Zhiyuan Song, Ziyue Dong, Qing Zhao, Xiaogang Xu, Pengxu Wei, Zhi chao Dou, Gui ling Wang, Chih-Chung Hsu, Chia-Ming Lee, Yi-Shiuan Chou, Cansu Korkmaz, A. Murat Tekalp, Yubin Wei, Xiaole Yan, Binren Li, Haonan Chen, Siqi Zhang, Sihan Chen, Amogh Joshi, Nikhil Akalwadi, Sampada Malagi, Palani Yashaswini, Chaitra Desai, Ramesh Ashok Tabib, Ujwala Patil, Uma Mudenagudi, Anjali Sarvaiya, Pooja Choksy, Jagrit Joshi, Shubh Kawa, Kishor Upla, Sushrut Patwardhan, Raghavendra Ramachandra, Sadat Hossain, Geongi Park, S. M. Nadim Uddin, Hao Xu, Yanhui Guo, Aman Urumbekov, Xingzhuo Yan, Wei Hao, Minghan Fu, Isaac Orais, Samuel Smith, Ying Liu, Wangwang Jia, Qisheng Xu, Kele Xu, Weijun Yuan, Zhan Li, Wenqin Kuang, Ruijin Guan, Ruting Deng, Zhao Zhang, Bo Wang, Suiyi Zhao, Yan Luo, Yanyan Wei, Asif Hussain Khan, Christian Micheloni, and Niki Martinel. NTIRE 2024 challenge on image super-resolution ($\times 4$): Methods and results, 2024. 2

[16] Kevin Clark, Paul Vicol, Kevin Swersky, and David J Fleet. Directly fine-tuning diffusion models on differentiable rewards. *arXiv preprint arXiv:2309.17400*, 2023. 2, 3, 5

[17] Mauricio Delbracio and Peyman Milanfar. Inversion by direct iteration: An alternative to denoising diffusion for image restoration. *arXiv preprint arXiv:2303.11435*, 2023. 2

[18] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2020. 2

[19] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P Simoncelli. Comparison of full-reference image quality models for optimization of image processing systems. *International Journal of Computer Vision (IJCV)*, 2021. 2, 3

[20] Keyan Ding, Kede Ma, Shiqi Wang, and Eero P. Simoncelli. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2022. 6

[21] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision (ECCV)*, 2014. 1, 2

[22] Chao Dong, Chen Change Loy, and Xiaoou Tang. Accelerating the super-resolution convolutional neural network. In *European Conference on Computer Vision (ECCV)*, 2016. 1, 2

[23] Ying Fan, Olivia Watkins, Yuqing Du, Hao Liu, Moonkyung Ryu, Craig Boutilier, Pieter Abbeel, Mohammad Ghavamzadeh, Kangwook Lee, and Kimin Lee. DPOK: reinforcement learning for fine-tuning text-to-image diffusion models. In *Neural Information Processing Systems (NeurIPS)*, 2023. 2

[24] William T Freeman, Thouis R Jones, and Egon C Pasztor. Example-based super-resolution. *IEEE Computer graphics and Applications*, 2002. 1

[25] Stephanie Fu, Netanel Tamir, Shobhita Sundaram, Lucy Chai, Richard Zhang, Tali Dekel, and Phillip Isola. DreamSim: Learning new dimensions of human visual similarity using synthetic data. *Neural Information Processing Systems (NeurIPS)*, 2023. 6

[26] Kevin Galim, Wonjun Kang, Yuchen Zeng, Hyung Il Koo, and Kangwook Lee. Parameter-efficient fine-tuning of state space models. *arXiv preprint arXiv:2410.09016*, 2024. 1

[27] Sara Ghazanfari, Siddharth Garg, Prashanth Krishnamurthy, Farshad Khorrami, and Alexandre Araujo. R-LPIPS: An adversarially robust perceptual similarity metric. *arXiv preprint arXiv:2307.15157*, 2023. 2

[28] Abhijay Ghildyal and Feng Liu. Shift-tolerant perceptual similarity metric. In *European Conference on Computer Vision (ECCV)*, 2022. 6

[29] S Alireza Golestaneh, Saba Dadsetan, and Kris M Kitani. No-reference image quality assessment via transformers, relative ranking, and self-consistency. In *Winter Conference on Applications of Computer Vision (WACV)*, 2022. 3

[30] Ian Goodfellow, Jean Pouget-Abadie, Mehdi Mirza, Bing Xu, David Warde-Farley, Sherjil Ozair, Aaron Courville, and Yoshua Bengio. Generative adversarial nets. *Neural Information Processing Systems (NeurIPS)*, 2014. 2

[31] Ian J Goodfellow, Jonathon Shlens, and Christian Szegedy. Explaining and harnessing adversarial examples. *arXiv preprint arXiv:1412.6572*, 2014. 5

[32] Hang Guo, Jinmin Li, Tao Dai, Zhihao Ouyang, Xudong Ren, and Shu-Tao Xia. MambaIR: A simple baseline for image restoration with state-space model. In *European Conference on Computer Vision (ECCV)*, 2024. 2

[33] Alexander Gushchin, Khaled Abud, Georgii Bychkov, Ekaterina Shumitskaya, Anna Chistyakova, Sergey Lavrushkin, Bader Rasheed, Kirill Malyshev, Dmitriy Vatolin, and Anastasia Antsiferova. Guardians of image quality: Benchmarking defenses against adversarial attacks on image quality metrics. *arXiv preprint arXiv:2408.01541*, 2024. 5

[34] Martin Heusel, Hubert Ramsauer, Thomas Unterthiner, Bernhard Nessler, and Sepp Hochreiter. GANs trained by a two time-scale update rule converge to a local Nash equilibrium. *Neural Information Processing Systems (NeurIPS)*, 2017. 3

[35] Vlad Hosu, Hanhe Lin, Tamas Sziranyi, and Dietmar Saupe. KonIQ-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing (TIP)*, 2020. 3, 1

[36] 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. *International Conference on Learning Representations (ICLR)*, 2022. 5, 1

[37] Xiaozhong Ji, Yun Cao, Ying Tai, Chengjie Wang, Jilin Li, and Feiyue Huang. Real-world super-resolution via kernel estimation and noise injection. In *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2020. 4

[38] Younghyun Jo, Seoung Wug Oh, Peter Vajda, and Seon Joo Kim. Tackling the ill-posedness of super-resolution through adaptive target generation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 2

[39] Justin Johnson, Alexandre Alahi, and Li Fei-Fei. Perceptual losses for real-time style transfer and super-resolution. In *European Conference on Computer Vision (ECCV)*, 2016. 2

[40] Bahjat Kawar, Michael Elad, Stefano Ermon, and Jiaming Song. Denoising diffusion restoration models. *Neural Information Processing Systems (NeurIPS)*, 2022. 2

[41] Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. MUSIQ: Multi-scale image quality transformer. In *International Conference on Computer Vision (ICCV)*, 2021. 2, 3, 5, 6, 1

[42] Markus Kettunen, Erik Härkönen, and Jaakko Lehtinen. E-LPIPS: robust perceptual image similarity via random transformation ensembles. *arXiv preprint arXiv:1906.03973*, 2019. 2

[43] Valentin Khrulkov and Artem Babenko. Neural side-by-side: Predicting human preferences for no-reference super-resolution evaluation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2021. 3, 1

[44] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Deeply-recursive convolutional network for image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[45] Alexey Kurakin, Ian Goodfellow, and Samy Bengio. Adversarial machine learning at scale. *International Conference on Learning Representations (ICLR)*, 2017. 5

[46] Yongchan Kwon, Eric Wu, Kevin Wu, and James Zou. DataInf: Efficiently estimating data influence in LoRA-tuned LLMs and diffusion models. *International Conference on Learning Representations (ICLR)*, 2024. 1

[47] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep Laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[48] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. 2

[49] Xiaoming Li, Wangmeng Zuo, and Chen Change Loy. Learning generative structure prior for blind text image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 6

[50] Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. SwinIR: Image restoration using swin transformer. In *International Conference on Computer Vision (ICCV)*, 2021. 2, 6, 1

[51] Jie Liang, Hui Zeng, and Lei Zhang. Details or artifacts: A locally discriminative learning approach to realistic image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 2

[52] Youwei Liang, Junfeng He, Gang Li, Peizhao Li, Arseniy Klimovskiy, Nicholas Carolan, Jiao Sun, Jordi Pont-Tuset, Sarah Young, Feng Yang, et al. Rich human feedback for text-to-image generation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 3

[53] Liting Lin, Heng Fan, Zhipeng Zhang, Yaowei Wang, Yong Xu, and Haibin Ling. Tracking meets LoRA: Faster training, larger model, stronger performance. In *European Conference on Computer Vision (ECCV)*, 2024. 1

[54] Yue Liu, Yunjie Tian, Yuzhong Zhao, Hongtian Yu, Lingxi Xie, Yaowei Wang, Qixiang Ye, Jianbin Jiao, and Yunfan Liu. VMamba: Visual state space model. In *Neural Information Processing Systems (NeurIPS)*, 2024. 2

[55] Yujia Liu, Chenxi Yang, Dingquan Li, Jianhao Ding, and Tingting Jiang. Defense against adversarial attacks on no-reference image quality models with gradient norm regularization. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 5

[56] Andreas Lugmayr, Martin Danelljan, Luc Van Gool, and Radu Timofte. SRFlow: Learning the super-resolution space with normalizing flow. In *European Conference on Computer Vision (ECCV)*, 2020. 2

[57] Chao Ma, Chih-Yuan Yang, Xiaokang Yang, and Ming-Hsuan Yang. Learning a no-reference quality metric for single-image super-resolution. *Computer Vision and Image Understanding (CVIU)*, 2017. 3

[58] Cheng Ma, Yongming Rao, Yean Cheng, Ce Chen, Jiwen Lu, and Jie Zhou. Structure-preserving super resolution with gradient guidance. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2

[59] Anish Mittal, Anush Krishna Moorthy, and Alan Conrad Bovik. No-reference image quality assessment in the spatial domain. *IEEE Transactions on Image Processing (TIP)*, 2012. 3, 1

[60] Jim Nilsson and Tomas Akenine-Möller. Understanding SSIM. *arXiv preprint arXiv:2006.13846*, 2020. 5

[61] Xingang Pan, Xiaohang Zhan, Bo Dai, Dahua Lin, Chen Change Loy, and Ping Luo. Exploiting deep generative prior for versatile image restoration and manipulation. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2021. 2

[62] JoonKyu Park, Sanghyun Son, and Kyoung Mu Lee. Content-aware local GAN for photo-realistic super-resolution. In *International Conference on Computer Vision (ICCV)*, 2023. 2

[63] Seung Ho Park, Young Su Moon, and Nam Ik Cho. Perception-oriented single image super-resolution using optimal objective estimation. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2

[64] Jonathan W Peirce. Understanding mid-level representations in visual processing. *Journal of Vision (JOV)*, 2015. 6

[65] Zygmunt Pizlo. Perception viewed as an inverse problem. *Vision research*, 41(24):3145–3161, 2001. 1

[66] Nikolay Ponomarenko, Oleg Ieremeiev, Vladimir Lukin, Karen Egiazarian, Lina Jin, Jaakko Astola, Benoit Vozel, Kacem Chehdi, Marco Carli, Federica Battisti, et al. Color image database TID2013: Peculiarities and preliminary results. In *European workshop on visual information processing (EUVIP)*, pages 106–111. IEEE, 2013. 3

[67] Mihir Prabhudesai, Anirudh Goyal, Deepak Pathak, and Katerina Fragkiadaki. Aligning text-to-image diffusion models with reward backpropagation. *arXiv preprint arXiv:2310.03739*, 2023. 2

[68] Michele A Saad, Alan C Bovik, and Christophe Charrier. A DCT statistics-based blind image quality index. *IEEE Signal Processing Letters*, 2010. 2, 3

[69] Chitwan Saharia, Jonathan Ho, William Chan, Tim Salimans, David J Fleet, and Mohammad Norouzi. Image super-resolution via iterative refinement. *IEEE Transactions on Pattern Analysis and Machine Intelligence (PAMI)*, 2022. 2

[70] Divya Saxena and Jiannong Cao. Generative adversarial networks (GANs) challenges, solutions, and future directions. *ACM Computing Surveys (CSUR)*, 2021. 2

[71] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2016. 2

[72] Shaolin Su, Qingsen Yan, Yu Zhu, Cheng Zhang, Xin Ge, Jinqiu Sun, and Yanning Zhang. Blindly assess image quality in the wild guided by a self-adaptive hyper network. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3

[73] Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. In *International Conference on Learning Representations (ICLR)*, 2014. 5

[74] Hossein Talebi and Peyman Milanfar. NIMA: Neural image assessment. *IEEE Transactions on Image Processing (TIP)*, 2018. 2, 3, 5, 6

[75] Huixuan Tang, Neel Joshi, and Ashish Kapoor. Learning a blind measure of perceptual image quality. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2011. 2, 3

[76] Radu Timofte, Vincent De Smet, and Luc Van Gool. Anchored neighborhood regression for fast example-based super-resolution. In *International Conference on Computer Vision (ICCV)*, 2013. 2

[77] Radu Timofte, Vincent De Smet, and Luc Van Gool. A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Proceedings of the Asian Conference on Computer Vision (ACCV)*, 2015. 2

[78] Ilya Tolstikhin, Olivier Bousquet, Sylvain Gelly, and Bernhard Schoelkopf. Wasserstein auto-encoders. *arXiv preprint arXiv:1711.01558v3*, 2018. 6

[79] Rao Muhammad Umer and Christian Micheloni. Deep cyclic generative adversarial residual convolutional networks for real image super-resolution. In *European Conference on Computer Vision Workshops (ECCVW)*, 2020. 2

[80] International Telecommunication Union. Methodology for the subjective assessment of the quality of television pictures. *Recommendation ITU-R BT.500-13*, 2012. Geneva, Switzerland. 5

[81] Narasimhan Venkatanath, D Praneeth, Maruthi Chandrasekhar Bh, Sumohana S Channappayya, and Swarup S Medasani. Blind image quality evaluation using perception based features. In *2015 Twenty first national conference on communications (NCC)*. IEEE, 2015. 3

[82] Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring CLIP for assessing the look and feel of images. In *Proceedings of the National Conference on Artificial Intelligence (AAAI)*, 2023. 3

[83] 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)*, 2024. 2

[84] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. ESRGAN: Enhanced super-resolution generative adversarial networks. In *European Conference on Computer Vision Workshops (ECCVW)*, 2018. 2, 6, 1

[85] Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In *International Conference on Computer Vision (ICCV)*, 2021. 4, 6, 1

[86] 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)*, 2004. 2, 6

[87] Haoning Wu, Erli Zhang, Liang Liao, Chaofeng Chen, Jingwen Hou, Annan Wang, Wenxiu Sun, Qiong Yan, and Weisi Lin. Exploring video quality assessment on user generated contents from aesthetic and technical perspectives. In *International Conference on Computer Vision (ICCV)*, 2023. 5

[88] 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 *International Conference on Machine Learning (ICML)*, 2024. 3, 5, 6, 1

[89] Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *Neural Information Processing Systems (NeurIPS)*, 2024. 2, 5

[90] Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. SeeSR: Towards semantics-aware real-world image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 5

[91] 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 *International Conference on Computer Vision (ICCV)*, 2023. 2

[92] Jiazheng Xu, Xiao Liu, Yuchen Wu, Yuxuan Tong, Qinkai Li, Ming Ding, Jie Tang, and Yuxiao Dong. ImageReward: Learning and evaluating human preferences for text-to-image generation. *Neural Information Processing Systems (NeurIPS)*, 2024. 2, 3

[93] Jianchao Yang, Zhaowen Wang, Zhe Lin, Scott Cohen, and Thomas Huang. Coupled dictionary training for image super-resolution. *IEEE Transactions on Image Processing (TIP)*, 2012. 2

[94] 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 *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2022. 3

[95] 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 (ECCV)*, 2024. 2, 5

[96] Jie-En Yao, Li-Yuan Tsao, Yi-Chen Lo, Roy Tseng, Chia-Che Chang, and Chun-Yi Lee. Local implicit normalizing flow for arbitrary-scale image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 2

[97] Zhenqiang Ying, Haoran Niu, Praful Gupta, Dhruv Mahajan, Deepti Ghadiyaram, and Alan Bovik. From patches to pictures (PaQ-2-PiQ): Mapping the perceptual space of picture quality. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 3, 5, 1

[98] Lin Zhang, Lei Zhang, and Alan C Bovik. A feature-enriched completely blind image quality evaluator. *IEEE Transactions on Image Processing (TIP)*, 2015. 3, 1

[99] Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018. 2

[100] Weixia Zhang, Kede Ma, Jia Yan, Dexiang Deng, and Zhou Wang. Blind image quality assessment using a deep bilinear convolutional neural network. *IEEE Transactions on Circuits and Systems for Video Technology*, 2020. 3

[101] Weixia Zhang, Guangtao Zhai, Ying Wei, Xiaokang Yang, and Kede Ma. Blind image quality assessment via vision-language correspondence: A multitask learning perspective. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. 3, 5

[102] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *European Conference on Computer Vision (ECCV)*, 2018. 2, 6

[103] Yuzhe Zhang, Jiawei Zhang, Hao Li, Zhouxia Wang, Luwei Hou, Dongqing Zou, and Liheng Bian. Diffusion-based blind text image super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2024. 6

[104] Heliang Zheng, Huan Yang, Jianlong Fu, Zheng-Jun Zha, and Jiebo Luo. Learning conditional knowledge distillation for degraded-reference image quality assessment. In *International Conference on Computer Vision (ICCV)*, 2021. 3