HETEROGENEOUS-NEGATIVE PREFERENCE OPTIMIZATION FOR REAL-WORLD SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

We present a new approach to explore Direct preference optimization (DPO) technique for real-world image supre-resolution (SR). First, we propose a preference-pair construction strategy in which the ground-truth high-quality image serves as the preferred sample, while predictions from a heterogeneous pool of existing SR models serve as negative samples. This strategy requires neither human annotations nor external reward models, while capturing diverse failure modes and providing more informative supervision for preference learning. Second, based on the constructed preference pairs, we derive a general DPO formulation applicable to arbitrary representation spaces. Further, we instantiate this formulation in the feature space of a pretrained DINOv2 model and introduce an adaptive layer-weighting mechanism to integrate representations across different semantic levels. Third, we decouple the shared temperature coefficient in DPO into two independent coefficients that separately control attraction toward preferred reconstructions and repulsion from negative predictions, thereby accounting for the asymmetric optimization difficulty of these two opposite directions. Extensive experiments on both synthetic and real-world datasets demonstrate the effectiveness and generality of the proposed framework.

1 INTRODUCTION

Image super-resolution (SR) aims to recover a high-quality (HQ) image from its low-quality (LQ) observation. This task is fundamentally ill-posed, as a single LQ image can correspond to multiple plausible HQ reconstructions. This ambiguity becomes more pronounced in real-world scenarios, where the underlying degradation process is typically complex and unknown. Despite substantial progress over recent decades (Dong et al., 2015; Zhang et al., 2021; Wang et al., 2024), real-world SR remains an active research topic, with significant challenges yet to be addressed.

Direct Preference Optimization (DPO) is an effective preference-alignment technique originally designed for language models (Rafailov et al., 2023) and has subsequently been extended to visual generation (Wallace et al., 2024). These advances motivate a natural question: can preference optimization also facilitate image SR? A key challenge in applying DPO to SR lies in constructing reliable preference pairs. Unlike language generation, SR results may exhibit different relative qualities across local image regions, making it difficult to establish a fair global preference. Moreover, large-scale human annotation is also expensive and time-consuming. Several recent studies have therefore explored automatic preference construction using existing image-quality assessment metrics or separately trained reward models (Wu et al., 2026; Cai et al., 2025; Yi et al., 2026), as illustrated in Fig.1 (b). However, these evaluators are themselves imperfect proxies for restoration quality, leading to unexpected overfitting onto the biases of the selected metric or reward model.

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

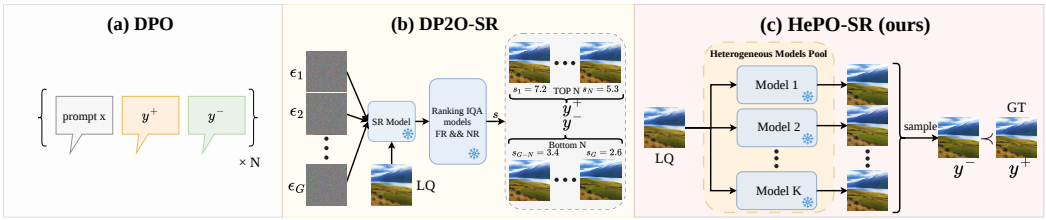

Figure 1: **Preference-pair construction in various methods.** (a) DPO uses human-annotated positive and negative samples. (b) DP^{2}O-SR (Wu et al., 2026) scores multiple predictions of the same model using manually selected IQA metrics to form preference order. (c) HePO-SR pairs the ground truth with a negative sampled from heterogeneous pretrained SR models pool.

In this work, we first consider the preference construction for SR. Given an LQ input, the goal is to obtain a positive reconstruction that aligns with user preferences, together with a corresponding negative reconstruction. Paired SR dataset already provides an ideal positive target, namely the ground-truth HQ image. To construct negative samples, we build a heterogeneous model pool comprising multiple existing SR models and randomly sample one of them to generate a negative example, as shown in Fig.1 (c). Different SR models exhibit distinct inductive biases and failure modes, such as over-smoothing, hallucinated textures, and structural distortion. Compared with existing DPO-based SR methods, our strategy contains a more diverse set of negative samples, providing richer and more informative supervision for subsequent preference-learning.

Building upon this preference construction strategy, we derive a DPO-based preference optimization framework tailored to SR. Our formulation aligns preferences in arbitrary representation spaces rather than being restricted to the image space, motivated by recent advances in representation learning and image generation (Yu et al., 2024; Yao et al., 2025). In this paper, we explore the preference optimization in the feature space of a pretrained DINOv2 model (Oquab et al., 2023). As different DINOv2 layers encode visual information at varying levels, ranging from low-level appearance to higher-level semantics, we further introduce an adaptive layer-weighting mechanism to dynamically balance their contributions during preference learning. Empirically, we find that both transferring preference optimization from pixel space to the DINOv2 feature space and adaptively aggregating multi-layer representations lead to consistent performance gains.

We further examine the role of the temperature coefficient in DPO. The temperature coefficient β jointly controls the model's attraction toward the preferred results and its repulsion from the negative predictions. We observe an inherent asymmetry between these two optimization directions in SR. Moving away from an imperfect negative result is often relatively easy because many directions in the output space can increase the repulsion distance. In contrast, approaching the unique ground-truth target requires recovering precise structures and high-frequency details, which is substantially more difficult. A shared temperature coefficient cannot adequately account for this asymmetry and may cause optimization to be dominated by the easier repulsive direction. We therefore decouple the temperature coefficient into β_1 and β_2, which independently control the attraction toward the preferred sample and the repulsion from the negative sample, respectively. This decoupling enables more precise control over the relative strength of the two preference directions and provides greater flexibility during optimization.

Our contributions can be summarized as follows:

- We propose a new strategy for constructing preference pairs for SR. Ground-truth HQ images serve as naturally preferred samples, while predictions from multiple existing SR models form a heterogeneous negative pool. This design captures diverse failure modes and provides richer supervision than existing single-source construction approaches.
- Based on the constructed preference pairs, we derive an SR-specific preference optimization framework that operates in arbitrary representation spaces. We instantiate this framework in the feature space of a pretrained DINOv2 model and demonstrate its effectiveness for SR.
- We decouple the temperature coefficient in DPO into two independent coefficients, thereby accounting for the asymmetric optimization difficulty of attraction toward preferred reconstructions and repulsion from negative predictions.

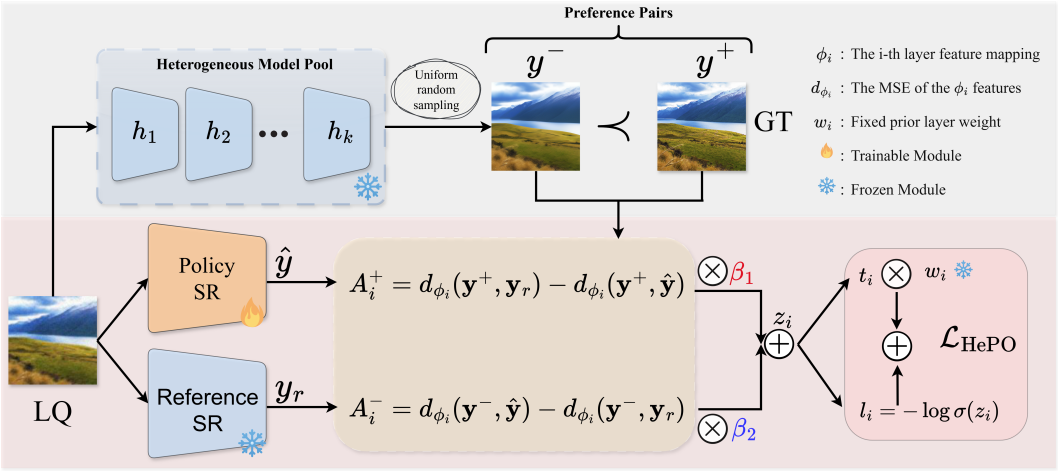

Figure 2: **Overview of HePO-SR.** A negative sample is uniformly drawn from a heterogeneous pool of pretrained SR models and paired with the ground truth. The trainable policy and frozen reference produce SR outputs from the same LQ input. Layer-wise DINOv2 feature distances define attraction and repulsion terms, controlled separately by β_1 and β_2. The resulting preference losses are aggregated using normalized adaptive layer weights and fixed prior weights.

Together, these designs establish a simple and general framework for adapting preference optimization to real-world image super-resolution, without relying on human annotations or a potentially biased external reward model.

2 Methodology

We propose a **He**terogeneous-negative **P**reference **O**ptimization, named HePO-SR, for Real-World Super-Resolution. Given an LQ image $\mathbf{x}$ and its ground-truth HQ counterpart $\mathbf{y}^+$, the trainable target model produces the reconstruction $\hat{\mathbf{y}} = f_\theta(\mathbf{x})$. We further use a frozen reference model, f_{ref}, to generate the reference solution $\mathbf{y}_r = f_{\text{ref}}(\mathbf{x})$. The reference model is initialized as an exact copy of the trainable model and remains fixed throughout fine-tuning. Notably, it is distinct from the external model pool used to construct negative samples. Fig. 2 illustrates the whole pipeline.

2.1 Heterogeneous Preference Optimization in a General Feature Space

Heterogeneous-negative preference construction. For each LQ image $\mathbf{x}$, the paired ground-truth image $\mathbf{y}^+$ naturally serves as the preferred reconstruction. To generate a corresponding negative sample, we define a heterogeneous model pool containing multiple existing SR methods, *i.e.*, $\mathcal{H} = \{h_1, h_2, \ldots, h_K\}$. We randomly sample h_k from $\mathcal{H}$ and use its output as the negative reconstruction:

$$\mathbf{y}^- = h_k(\mathbf{x}). \tag{1}$$

The source-model outputs can either be generated online during training or precomputed and cached offline. Combined with the ground-truth preferred target, this procudure constructs a preference dataset of triplets, namely $\mathcal{D}_{\text{pref}} = \{(\mathbf{x}, \mathbf{y}^+, \mathbf{y}^-)\}$.

The heterogeneous model pool is designed to capture the diverse failure modes of existing SR methods. Different source models exhibit distinct inductive biases and may produce artifacts such as over-smoothed regions, hallucinated textures, and structural distortions. By exposing the target model to these heterogeneous negative reconstructions, HePO-SR provides a richer preference signal than a single-source negative construction strategy and encourages the model to move away from these undesirable reconstructions while aligning its predictions with the preferred ground-truth ones.

Auxiliary distributions in a general feature space. Standard DPO (Rafailov et al., 2023) optimizes likelihood ratios between trainable and reference models evaluated on preferred and negative samples. However, most existing SR models are deterministic and do not define output likelihoods.

To enable likelihood-based preference optimization, we introduce auxiliary conditional distributions in a general representation space.

Let $\phi : \mathcal{Y} \to \mathbb{R}^N$ denote a feature mapping. We define its vectorized and dimension-normalized representation as

$$\psi_\phi(\mathbf{y}) = \frac{\mathrm{vec}(\phi(\mathbf{y}))}{\sqrt{N}}. \tag{2}$$

The mapping ϕ may be the identity mapping for pixel-space comparison or a representation extracted by a visual encoder. We define the corresponding squared feature distance as

$$d_\phi(\mathbf{a}, \mathbf{b}) = \|\psi_\phi(\mathbf{a}) - \psi_\phi(\mathbf{b})\|_2^2 . \tag{3}$$

For a candidate target representation $\mathbf{u}$, we define the auxiliary conditional distributions associated with the trainable and reference outputs as

$$q_\theta^\phi(\mathbf{u} \mid \mathbf{x}) = \mathcal{N}\left(\mathbf{u}; \psi_\phi(\hat{\mathbf{y}}), s^2\mathbf{I}\right), \;\; q_{\mathrm{ref}}^\phi(\mathbf{u} \mid \mathbf{x}) = \mathcal{N}\left(\mathbf{u}; \psi_\phi(\mathbf{y}_r), s^2\mathbf{I}\right). \tag{4}$$

The means of these distributions are given by the representations of the trainable and reference reconstructions, respectively. Both distributions share a parameter-independent variance s^2, which places their likelihoods on a common feature-distance scale.

Feature-space preference objective. As the shared variance scale cancels in the log-likelihood ratio, we thus yield

$$\Delta_\phi(\mathbf{y}; \mathbf{x}) = \log \frac{q_\theta^\phi\left(\psi_\phi(\mathbf{y}) \mid \mathbf{x}\right)}{q_{\mathrm{ref}}^\phi\left(\psi_\phi(\mathbf{y}) \mid \mathbf{x}\right)} = \frac{d_\phi(\mathbf{y}, \mathbf{y}_r) - d_\phi(\mathbf{y}, \hat{\mathbf{y}})}{2s^2}. \tag{5}$$

Here, $\Delta_\phi(\mathbf{y}; \mathbf{x})$ quantifies the relative preference of the trainable model over the reference model for the target image $\mathbf{y}$ in the representation space induced by ϕ.

Motivated by DPO, we define the following feature-space preference objective:

$$\mathcal{L}_{\phi\text{-DPO}} = -\mathbb{E}_{(\mathbf{x}, \mathbf{y}^+, \mathbf{y}^-) \sim \mathcal{D}_{\mathrm{pref}}} \left[\log \sigma\left(\beta\left[\Delta_\phi(\mathbf{y}^+; \mathbf{x}) - \Delta_\phi(\mathbf{y}^-; \mathbf{x})\right]\right)\right], \tag{6}$$

where $\sigma(\cdot)$ is the sigmoid function and β is the temperature coefficient.

To explicitly characterize the two directions induced by the preference update, we rewrite Eq. (6) by decomposing its logit into positive and negative components:

$$\mathcal{L}_{\phi\text{-DPO}} = -\mathbb{E}\left[\log \sigma\left(\frac{\beta}{2s^2}(A^+ - A^-)\right)\right], \tag{7}$$

where

$$A^+ = d_\phi(\mathbf{y}^+, \mathbf{y}_r) - d_\phi(\mathbf{y}^+, \hat{\mathbf{y}}), \;\; A^- = d_\phi(\mathbf{y}^-, \mathbf{y}_r) - d_\phi(\mathbf{y}^-, \hat{\mathbf{y}}). \tag{8}$$

The term A^+ measures the relative improvement of the trainable reconstruction over the reference one with respect to the preferred image. In contrast, A^- measures the relative proximity of the trainable reconstruction to the negative sample. Minimizing Eq. (7) thus encourages the model to approach the preferred solution while separating from heterogeneous negative ones.

2.2 Multi-Level DINOv2 Features and Adaptive Aggregation

Multi-level DINOv2 features. We instantiate the feature mapping ϕ in Eq. (2) using a pretrained DINOv2 encoder (Oquab et al., 2023; Darcet et al., 2024). Features from different layers encode complementary spatial and semantic information at varying scales (Zhang et al., 2024). We thus align preference on a set of selected layers $\mathcal{I}$ rather than relying on a single representation.

For each layer $i \in \mathcal{I}$, we define ϕ_i, ψ_i, and d_i following Eqs. (2) and (3). We further compute the preferred and negative components, $A_{b,i}^+$ and $A_{b,i}^-$, for sample b according to Eq. (8). The layer-wise preference loss is given by

$$\ell_{b,i} = -\log \sigma(z_{b,i}), \;\; z_{b,i} = \frac{\beta}{2s^2}\left(A_{b,i}^+ - A_{b,i}^-\right). \tag{9}$$

Adaptive layer aggregation. The contribution of each feature layer can vary across training samples. We introduce a set of prior weights $\{\omega_i\}_{i\in\mathcal{I}}$, satisfying $\sum_{i\in\mathcal{I}}\{\omega_i\} = 1$, and adapt them according to the relative layer responses of each sample. Specifically, we define the normalized layer response as

$$t_{b,i} = \frac{\mathrm{sg}(z_{b,i}) - \mu_b}{\nu_b + \epsilon}, \tag{10}$$

where $\mathrm{sg}(\cdot)$ is the stop-gradient operator, ϵ is a small constant for numerical stability, and μ_b and ν_b are the mean and standard deviation of the detached logits $\{\mathrm{sg}(z_{b,i})\}_{i\in\mathcal{I}}$, respectively.

Let $S_b = \sum_{j\in\mathcal{I}} \omega_j |t_{b,j}|$. We define the aggregation weights as

$$\alpha_{b,i} = \begin{cases} \dfrac{\omega_i |t_{b,i}|}{S_b}, & S_b > \epsilon, \\ \omega_i, & S_b \le \epsilon, \end{cases} \tag{11}$$

where ϵ is a predefined threshold. When the layer responses have negligible variation, we retain the prior weights. Otherwise, we reweight the layers according to the magnitudes of their sample-specific normalized responses. Finally, we aggregate the layer-wise losses in Eq. (9) as

$$\mathcal{L}_{\mathrm{HePO}} = \frac{1}{B}\sum_{b=1}^{B}\sum_{i\in\mathcal{I}} \alpha_{b,i}\ell_{b,i}, \tag{12}$$

where B is the batch size.

2.3 Decoupled Temperature Coefficients

The shared-temperature mechanism in Eq. (7) optimizes the coupled quantity $A^+ - A^-$. Consequently, it does not guarantee simultaneous improvement in both components, where our desired behavior is to increase A^+ while decreasing A^-. Fig. 3 illustrates this issue during training. Although A^- decreases steadily, meaning increasing separation between the prediction and the negative sample, A^+ may also decrease. This indicates that the prediction fails to move closer to the preferred solution relative to the reference output. A reduction in A^- can compensate for an undesired reduction in A^+ while still decreasing the DPO loss. Thus, the model may exploit a reward-hacking-like shortcut that prioritizes separation from negative samples without improving attraction toward the preferred ground truth.

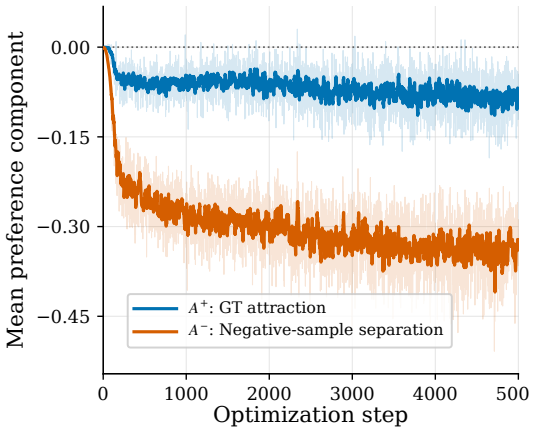

Figure 3: Training dynamics with a shared $\beta = 4$.

This imbalance arises partly because a shared temperature coefficient jointly scales A^+ and A^- without independently controlling their relative contributions. We therefore replace the shared temperature with two independent coefficients, β_1 and β_2, and redefine the preference objective of Eq. (7) as

$$\mathcal{L}_{\mathrm{dec}} = -\mathbb{E}\left[\log\sigma\left(\beta_1 A^+ - \beta_2 A^-\right)\right]. \tag{13}$$

Note that the common factor $1/(2s^2)$ is absorbed into the two coefficients. Here, β_1 controls the contribution of attraction toward the ground-truth target, whereas β_2 controls the strength of separation from negative samples. Decoupling these coefficients enables explicit control over the relative strengths of the two optimization directions and mitigates the observed component imbalance.

2.4 Training Objectives

HePO-SR is applied to both OSEDiff and OMGSR while preserving their respective reconstruction objectives. For either backbone, the training objective is

$$\mathcal{L} = \mathcal{L}_{\mathrm{Others}} + \lambda_{\mathrm{HePO}}\mathcal{L}_{\mathrm{HePO}}, \tag{14}$$

where $\mathcal{L}_{\text{Others}}$ denotes the losses inherited from the corresponding backbone, and λ_{HePO} controls the contribution of the HePO-SR preference loss relative to these reconstruction losses.

The value of λ_{HePO}, together with the specific loss terms, weighting coefficients, and implementation details for OSEDiff and OMGSR, is provided in Appendix B. The preference loss $\mathcal{L}_{\text{HePO}}$ is defined in Eq. (12).

2.5 Discussion

We do not report direct comparisons between HePO and GDPO/DP^{2}O for two reasons. First, applying GDPO or DP^{2}O to OSEDiff and OMGSR baselines is incompatible with their native one-step generation process. Both methods rely on generating multiple candidates for the same LQ input using different diffusion noise realizations, such that meaningful group-wise or pairwise preferences can be constructed. In OSEDiff and OMGSR, however, the adapted VAE encoder directly maps an LQ image to a noisy intermediate latent associated with a predefined diffusion timestep, which is then restored through a single UNet prediction. Resampling the VAE posterior provides only limited output diversity and is unlikely to produce sufficiently distinct candidates for reliable preference ranking. Conversely, injecting additional diffusion noise would perturb a latent that already represents a noisy intermediate state, causing double noising and a mismatch between its actual signal-to-noise ratio and the timestep supplied to the UNet. This would alter the original training distribution and generation pipeline, making the resulting variants unsuitable for a controlled comparison. Second, applying HePO to the original GDPO and DP^{2}O baselines cannot be reproduced under matched conditions because the baseline checkpoint required by GDPO has not been publicly released, while the training data used by DP^{2}O are unavailable.

3 Experiments

3.1 Experimental Setup

Training Details. We employ HePO-SR for post-training on two diffusion pre-trained backbones, OSEDiff (Wu et al., 2024) and OMGSR (Wu et al., 2025). Both variants are trained on the same crops from LSDIR (Li et al., 2023) while retaining their model-specific reconstruction objectives. Our negative pool consists of the classical SR results by SwinIR, Real-ESRGAN, StableSR, and DiffBIR (Liang et al., 2021; Wang et al., 2021; 2024; Lin et al., 2024). In the main text, we use offline training, i.e., GT images, LQ inputs, and negative samples are prepared and cached in advance, while model predictions and frozen-reference outputs are computed during training. Results for online training are provided in Appendix C. We extract DINOv2 features from layers $\{0, 3, 7, 9, 11\}$, aggregate the layer-wise preference signals adaptively, and set $(\beta_1, \beta_2) = (5, 3)$. Optimization and implementation details are provided in Appendix B.

Testing Datasets. We evaluate $4\times$ SR on the paired real-world datasets DRealSR (Wei et al., 2020) and RealSR (Cai et al., 2019), the synthetically degraded DIV2K benchmark (Agustsson & Timofte, 2017), and the unpaired real-world dataset RealSet65 (Yue et al., 2023).

Compared Methods. We compare with BSRGAN (Zhang et al., 2021), Real-ESRGAN (Wang et al., 2021), ResShift (Yue et al., 2023), DiffBIR (Lin et al., 2024), InvSR (Yue et al., 2025), PiSA-SR (Sun et al., 2025), OSEDiff (Wu et al., 2024), and OMGSR (Wu et al., 2025). OSEDiff* and OMGSR* denote the corresponding models fine-tuned with HePO-SR.

Metrics. PSNR, SSIM (Wang et al., 2004), LPIPS (Zhang et al., 2018), and DISTS (Ding et al., 2020) are used as full-reference metrics, while CLIP-IQA (Wang et al., 2023), MUSIQ (Ke et al., 2021), MANIQA (Yang et al., 2022), TOPIQ (Chen et al., 2024), and AFINE-NR (Chen et al., 2025) are used as no-reference metrics.

3.2 Comparison with State-of-the-Art Methods

Quantitative Comparison and Analysis. Tables 1 and 2 quantitatively compare HePO-SR with representative real-world SR methods on the paired benchmarks and RealSet65, respectively. As shown in Table 1, HePO-SR improves most no-reference perceptual metrics across the three paired benchmarks. In particular, CLIP-IQA, MUSIQ, and TOPIQ increase for both backbones on all three

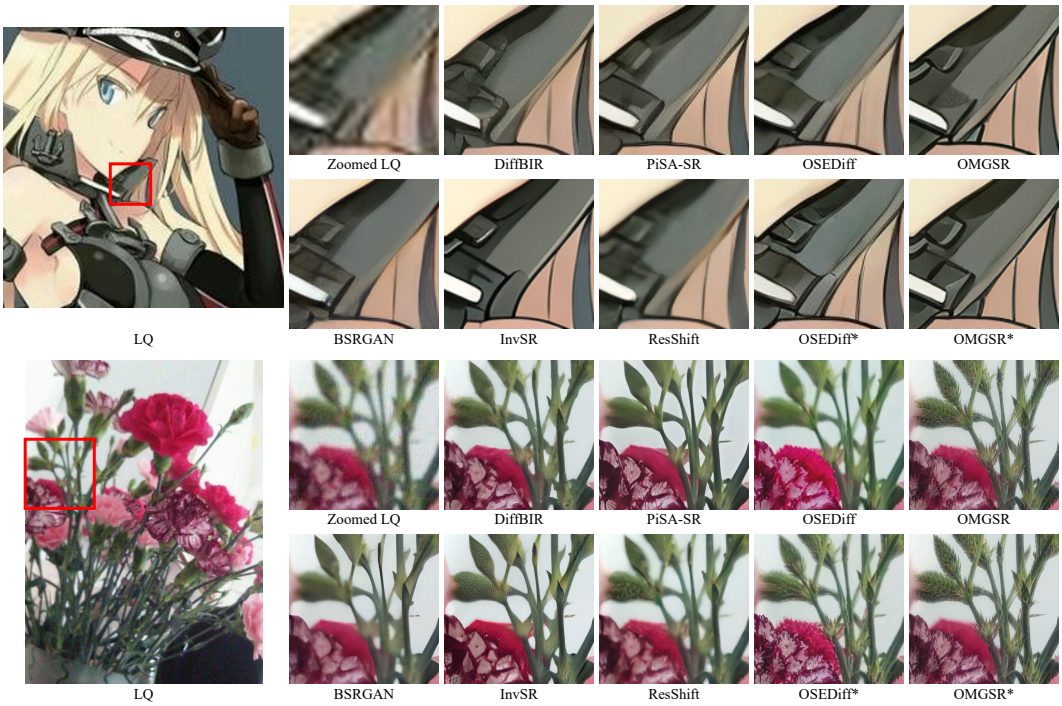

Figure 4: Visual comparison on a real-world flower image. The left panel is the LQ input with the selected crop, and the remaining panels show the corresponding enlarged region. Asterisks denote models fine-tuned with HePO-SR.

datasets; MANIQA follows the same trend with one negligible exception for OMGSR on DIV2K. AFINE-NR is also generally favorable, although its changes are less consistent for OSEDiff. Full-reference metrics show a more dataset-dependent pattern: PSNR decreases for both backbones, while SSIM, LPIPS, and DISTS improve for some model–dataset pairs. For example, OMGSR* improves SSIM on all three datasets, and OSEDiff* obtains lower LPIPS and DISTS on RealSR and DIV2K, despite declines on DRealSR. The same trend is observed on the unpaired RealSet65 benchmark. OMGSR* improves all five no-reference metrics and achieves the strongest overall scores among the compared methods. OSEDiff* improves four of the five metrics, with a slight decline in AFINE-NR. In summary, the results show that HePO-SR most consistently improves no-reference perceptual quality, while not significantly changing full-reference measures.

Qualitative Analysis and Visualization. Fig. 4 presents qualitative comparisons on two images: a real-world flower photograph and a cartoon-style illustration. On the flower image, the fine-tuned models exhibit clear improvements over their respective base models. Relative to OSEDiff, OSEDiff* recovers more visible texture around the petals and buds, whereas OMGSR* suppresses visible artifacts and produces cleaner bud and stem structures than OMGSR. On the cartoon-style illustration, OSEDiff* yields sharper boundaries and more discernible local structures in the gray object than OSEDiff, while OMGSR* produces cleaner and more clearly separated contour and panel boundaries, with less local smoothing than OMGSR. These examples show that HePO-SR fine-tuning improves local detail and structural clarity across both photographic and stylized content.

Table 1: Quantitative comparison on the paired SR benchmarks. Arrows indicate the preferred metric direction. Bold and underlined entries indicate the best and second-best results within each dataset.

Dataset	Method	Metrics								
		PSNR↑	SSIM↑	LPIPS↓	DISTS↓	CLIP-IQA↑	MUSIQ↑	MANIQA↑	TOPIQ↑	AFINE-NR↓
DRealSR	BSRGAN	**28.70**	0.8028	0.2858	0.2144	0.5093	57.17	0.4845	0.5062	-0.7630
	Real-ESRGAN	28.62	**0.8052**	**0.2819**	**0.2089**	0.4512	54.27	0.4904	0.4624	-0.7875
	InvSR	25.67	0.7131	0.3537	0.2459	0.7136	65.99	0.6326	0.6106	-0.9326
	DiffBIR	26.82	0.6633	0.4497	0.2731	0.6374	60.86	0.5918	0.5831	-0.8047
	PiSA-SR	28.31	0.7804	0.2960	0.2169	0.6971	66.11	0.6161	0.6332	-0.9325
	ResShift	27.05	0.7404	0.3870	0.2632	0.5394	51.24	0.4666	0.4530	-0.6586
	OMGSR	27.56	0.7465	0.3266	0.2335	0.6957	66.67	0.6365	0.6558	-0.9808
	OMGSR*	27.50	0.7518	0.3378	0.2512	**0.7205**	**67.35**	**0.6375**	**0.6842**	**-1.006**
	OSEDiff	27.92	0.7835	0.2968	0.2165	0.6963	64.65	0.5895	0.6004	-0.9178
	OSEDiff*	27.73	0.7440	0.3101	0.2207	0.7138	66.36	0.6266	0.6237	-0.9171
RealSR	BSRGAN	**26.37**	**0.7651**	**0.2656**	0.2124	0.5113	63.29	0.5420	0.5506	-0.8729
	Real-ESRGAN	25.68	0.7614	0.2710	0.2061	0.4484	60.37	0.5505	0.5149	-0.9133
	InvSR	24.13	0.7125	0.2871	0.2123	0.6790	68.54	0.6618	0.5900	-1.094
	DiffBIR	24.81	0.6595	0.3607	0.2311	0.6446	64.94	0.6232	0.5728	-0.9660
	PiSA-SR	25.50	0.7418	0.2672	**0.2044**	0.6700	70.15	0.6551	0.6372	-1.070
	ResShift	25.66	0.7360	0.3279	0.2475	0.5352	56.89	0.5108	0.4527	-0.8153
	OMGSR	24.60	0.7060	0.3009	0.2259	0.6713	69.95	0.6754	0.6700	-1.132
	OMGSR*	24.55	0.7068	0.3142	0.2416	0.6943	**70.53**	**0.6772**	**0.7001**	**-1.139**
	OSEDiff	25.15	0.7341	0.2921	0.2128	0.6693	69.08	0.6331	0.6254	-1.048
	OSEDiff*	25.08	0.7101	0.2872	0.2100	**0.6946**	69.88	0.6593	0.6310	-1.048
DIV2K	BSRGAN	24.58	0.6269	0.3351	0.2275	0.5247	61.20	0.5041	0.5460	-0.7551
	Real-ESRGAN	24.29	**0.6372**	0.3112	0.2141	0.5277	61.06	0.5485	0.5297	-0.8270
	InvSR	22.90	0.5910	0.3187	0.2058	**0.7180**	68.89	0.6403	0.6506	-0.9697
	DiffBIR	23.66	0.5652	0.3516	0.2124	0.6693	65.77	0.6194	0.6148	-0.8840
	PiSA-SR	23.86	0.6058	**0.2823**	**0.1934**	0.6928	**69.68**	0.6375	0.6619	-0.9448
	ResShift	**24.59**	0.6232	0.3077	0.2136	0.5716	58.90	0.5276	0.5059	-0.7482
	OMGSR	23.13	0.5745	0.3078	0.2005	0.6906	69.16	**0.6555**	0.6839	-0.9996
	OMGSR*	23.02	0.5811	0.3049	0.2080	0.7114	69.54	0.6554	**0.7049**	**-1.013**
	OSEDiff	23.72	0.6109	0.2942	0.1975	0.6681	67.97	0.6131	0.6186	-0.9199
	OSEDiff*	23.11	0.5737	0.2899	0.1936	0.6868	69.39	0.6457	0.6283	-0.9550

Table 2: Quantitative comparison on RealSet65. The dataset has no paired ground-truth images, so only no-reference metrics are reported. Bold and underlined entries indicate the best and second-best results, respectively.

Dataset	Method	Metrics				
		CLIP-IQA↑	MUSIQ↑	MANIQA↑	TOPIQ↑	AFINE-NR↓
RealSet65	BSRGAN	0.6163	65.59	0.5370	0.5812	-0.9649
	Real-ESRGAN	0.5991	63.22	0.5496	0.5367	-0.9662
	InvSR	0.7040	68.80	0.6209	0.6244	-0.9999
	DiffBIR	0.6684	62.41	0.5735	0.5739	-0.9201
	PiSA-SR	0.6959	69.68	0.6159	0.6240	-1.0060
	ResShift	0.6202	58.57	0.5024	0.4878	-0.8727
	OMGSR	0.7400	70.29	0.6360	0.6911	-1.0230
	OMGSR*	**0.7556**	**70.54**	**0.6370**	**0.7124**	**-1.0353**
	OSEDiff	0.6871	68.91	0.6013	0.5957	-0.9725
	OSEDiff*	0.7082	69.88	0.6226	0.6248	-0.9710

3.3 Ablation Study

We examine negative-pool expansion, adaptive layer weighting, and coefficient allocation on RealSet65 using OMGSR. All variants use the five no-reference metrics in Table 2. Table 3 summarizes these comparisons. The reconstruction-only continued training results are reported in Appendix C.1, while the additional quantitative and qualitative comparisons with alternative preferences spaces are provided in Appendix C.2.

Table 3: Ablation study of HePO-SR with OMGSR on RealSet65. For Expanded Pool, $\times$ denotes the two-model pool instead of the four-model pool. For Adaptive Weights, $\times$ denotes the fixed base layer weights. Full (Ours) denotes the configuration used for the principal results.

Variant	Configuration			Metrics				
	Expanded Pool	Adaptive Weights	(β_1, β_2)	CLIP-IQA↑	MUSIQ↑	MANIQA↑	TOPIQ↑	AFINE-NR↓
Variant 1	$\times$	✓	$(5, 3)$	<u>0.7513</u>	**70.5940**	<u>0.6407</u>	<u>0.7040</u>	<u>-1.0323</u>
Variant 2	✓	$\times$	$(5, 3)$	0.7408	70.1252	0.6376	0.6868	-1.0095
Variant 3	✓	✓	$(4, 4)$	0.7393	70.0483	0.6397	0.6865	-1.0058
Variant 4	✓	✓	$(3, 5)$	0.7382	70.0019	**0.6409**	0.6869	-1.0076
Full (Ours)	✓	✓	$(5, 3)$	**0.7556**	<u>70.5496</u>	0.6370	**0.7124**	**-1.0353**

Impact of Negative-Pool Expansion. We compare a two-model pool containing SwinIR and StableSR with the four-model pool, which also includes Real-ESRGAN and DiffBIR. The expanded pool improves three no-reference metrics, including CLIP-IQA, TOPIQ, and AFINE-NR, while resulting in slightly lower MUSIQ and MANIQA.

Impact of Adaptive Layer Weights. Adaptive weighting modulates the fixed base weights according to each sample's layer logits. We compare it with the same base weights used without modulation, keeping the selected layers unchanged. Adaptive weighting improves most of the no-reference metrics, except for a slight decrease in MANIQA.

Impact of Decoupled Temperature Coefficients. We compare $(\beta_1, \beta_2) = (4, 4)$, $(5, 3)$, and $(3, 5)$. These three pairs have the same coefficient sum, $\beta_1 + \beta_2 = 8$, which keeps the nominal coefficient budget and the overall scale of the preference objective fixed while changing the relative emphasis between the two components. Among these settings, $(5, 3)$ performs the best in most of the metrics, including CLIP-IQA, MUSIQ, TOPIQ, and AFINE-NR, while only worse than $(3, 5)$ in MANIQA. These results support the effectiveness of placing greater weight on ground-truth attraction, which is consistent with the motivation to counter excessive negative-sample separation in Section 2.3. Appendix C.3 provides a qualitative comparison of the three coefficient allocations and further analyzes these settings through component-wise training dynamics, illustrating how the coefficient allocation changes the balance between ground-truth attraction and negative-sample separation.

In summary, these ablation results support the reasonableness of the proposed method's hyperparameter settings.

4 Conclusion

We presented HePO-SR, a preference optimization framework for real-world image super-resolution. HePO-SR constructs preference pairs by treating the ground-truth image as the preferred target and outputs from a heterogeneous pool of pretrained SR models as negative samples, thereby eliminating the need for human annotations or external reward models. We formulated DPO in a general representation space and instantiated it using adaptively aggregated multi-level DINOv2 features. We further found that, with a shared temperature coefficient, the model can improve the preference objective primarily by increasing its separation from negative samples. Decoupling the coefficient into β_1 and β_2 prevents this one-sided optimization by independently controlling attraction toward the ground-truth image and repulsion from negative samples. Experiments on synthetic and real-world benchmarks demonstrate that HePO-SR improves both OSEDiff and OMGSR. The ablation studies further support the use of heterogeneous negative samples, the DINOv2 feature space, adaptive multi-level aggregation, and the decoupled coefficient allocation. Overall, the results demonstrate the effectiveness of heterogeneous-negative preference learning for improving pretrained real-world SR models.

AI Use Statement

Generative AI was used to assist with the English-language drafting and revision of this manuscript, including improvements to grammar, clarity, organization, and academic presentation.

REFERENCES

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1122–1131, 2017. doi: 10.1109/CVPRW.2017.150.

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 *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 3086–3095, 2019. doi: 10.1109/ICCV.2019.00318.

Miaomiao Cai, Simiao Li, Wei Li, Xudong Huang, Hanting Chen, Jie Hu, and Yunhe Wang. Dspo: Direct semantic preference optimization for real-world image super-resolution. *arXiv preprint arXiv:2504.15176*, 2025.

Mathilde Caron, Hugo Touvron, Ishan Misra, Hervé Jégou, Julien Mairal, Piotr Bojanowski, and Armand Joulin. Emerging properties in self-supervised vision transformers. In *2021 IEEE/CVF international conference on computer vision (ICCV)*, pp. 9630–9640. IEEE, 2021.

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*, 33:2404–2418, 2024. doi: 10.1109/TIP.2024.3378466.

Du Chen, Tianhe Wu, Kede Ma, and Lei Zhang. Toward generalized image quality assessment: Relaxing the perfect reference quality assumption. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 12742–12752, 2025. URL https://openaccess.thecvf.com/content/CVPR2025/html/Chen_Toward_Generalized_Image_Quality_Assessment_Relaxing_the_Perfect_Reference_Quality_CVPR_2025_paper.html.

Timothée Darcet, Maxime Oquab, Julien Mairal, and Piotr Bojanowski. Vision transformers need registers. In B. Kim, Y. Yue, S. Chaudhuri, K. Fragkiadaki, M. Khan, and Y. Sun (eds.), *International Conference on Learning Representations*, volume 2024, pp. 2632–2652, 2024. URL https://proceedings.iclr.cc/paper_files/paper/2024/file/0b408293619f725fd30162af057e531a-Paper-Conference.pdf.

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*, 44(5):2567–2581, 2020.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2): 295–307, 2015.

Junjie Ke, Qifei Wang, Yilin Wang, Peyman Milanfar, and Feng Yang. Musiq: Multi-scale image quality transformer. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 5128–5137. IEEE, 2021.

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 *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1775–1787. IEEE, 2023.

Jingyun Liang, Jiezhang Cao, Guolei Sun, Kai Zhang, Luc Van Gool, and Radu Timofte. Swinir: Image restoration using swin transformer. In *2021 IEEE/CVF international conference on computer vision workshops (ICCVW)*, pp. 1833–1844. IEEE, 2021.

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 *European conference on computer vision*, pp. 430–448. Springer, 2024.

Runtao Liu, Haoyu Wu, Ziqiang Zheng, Chen Wei, Yingqing He, Renjie Pi, and Qifeng Chen. Videodpo: Omni-preference alignment for video diffusion generation. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8009–8019. IEEE, 2025.

Maxime Oquab, Timothée Darcet, Théo Moutakanni, Huy Vo, Marc Szafraniec, Vasil Khalidov, Pierre Fernandez, Daniel Haziza, Francisco Massa, Alaaeldin El-Nouby, et al. Dinov2: Learning robust visual features without supervision. *arXiv preprint arXiv:2304.07193*, 2023.

Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. Training language models to follow instructions with human feedback. *Advances in neural information processing systems*, 35: 27730–27744, 2022.

Rafael Rafailov, Archit Sharma, Eric Mitchell, Christopher D Manning, Stefano Ermon, and Chelsea Finn. Direct preference optimization: Your language model is secretly a reward model. *Advances in neural information processing systems*, 36:53728–53741, 2023.

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 *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2333–2343. IEEE, 2025.

Bram Wallace, Meihua Dang, Rafael Rafailov, Linqi Zhou, Aaron Lou, Senthil Purushwalkam, Stefano Ermon, Caiming Xiong, Shafiq Joty, and Nikhil Naik. Diffusion model alignment using direct preference optimization. In *2024 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 8228–8238. IEEE, 2024.

Jianyi Wang, Kelvin CK Chan, and Chen Change Loy. Exploring clip for assessing the look and feel of images. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pp. 2555–2563, 2023.

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*, 132(12):5929–5949, 2024.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW)*, pp. 1905–1914. IEEE, 2021.

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*, 13(4):600–612, 2004.

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 *European conference on computer vision*, pp. 101–117. Springer, 2020.

Rongyuan Wu, Lingchen Sun, Zhiyuan Ma, and Lei Zhang. One-step effective diffusion network for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 37: 92529–92553, 2024.

Rongyuan Wu, Lingchen Sun, Zhengqiang Zhang, Shihao Wang, Tianhe Wu, Qiaosi Yi, Shuai Li, and Lei Zhang. Dp^{2}o-sr: Direct perceptual preference optimization for real-world image super-resolution. *Advances in Neural Information Processing Systems*, 38:172015–172041, 2026.

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.

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 *2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 1190–1199. IEEE, 2022.

Jingfeng Yao, Bin Yang, and Xinggang Wang. Reconstruction vs. generation: Taming optimization dilemma in latent diffusion models. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 15703–15712. IEEE, 2025.

Qiaosi Yi, Shuai Li, Rongyuan Wu, Lingchen Sun, Zhengqiang Zhang, and Lei Zhang. Gdpo-sr: Group direct preference optimization for one-step generative image super-resolution. *arXiv preprint arXiv:2603.16769*, 2026.

Sihyun Yu, Sangkyung Kwak, Huiwon Jang, Jongheon Jeong, Jonathan Huang, Jinwoo Shin, and Saining Xie. Representation alignment for generation: Training diffusion transformers is easier than you think. *arXiv preprint arXiv:2410.06940*, 2024.

Zongsheng Yue, Jianyi Wang, and Chen Change Loy. Resshift: Efficient diffusion model for image super-resolution by residual shifting. *Advances in neural information processing systems*, 36: 13294–13307, 2023.

Zongsheng Yue, Kang Liao, and Chen Change Loy. Arbitrary-steps image super-resolution via diffusion inversion. In *2025 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 23153–23163. IEEE, 2025.

Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *2021 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 4771–4780. IEEE, 2021.

Richard Zhang, Phillip Isola, Alexei A Efros, Eli Shechtman, and Oliver Wang. The unreasonable effectiveness of deep features as a perceptual metric. In *2018 IEEE/CVF conference on computer vision and pattern recognition*, pp. 586–595. IEEE, 2018.

Yongjian Zhang, Longguang Wang, Kunhong Li, Yun Wang, and Yulan Guo. Learning representations from foundation models for domain generalized stereo matching. In *European Conference on Computer Vision*, pp. 146–162. Springer, 2024.

A Related Work

A.1 Image Super-Resolution and Heterogeneous Reconstruction Sources

Image SR has developed from reconstruction-oriented convolutional networks to transformer-based and generative models. Residual architectures such as EDSR and RCAN improve reconstruction through deep residual transformations and channel attention, while SwinIR and HAT introduce window-based attention and channel aggregation for stronger spatial modeling (Liang et al., 2021). These methods establish feed-forward reconstruction as an important foundation for both conventional and real-world SR. Real-world SR further requires models to handle unknown degradations. BSRGAN and Real-ESRGAN address this issue through richer synthetic degradation processes and adversarial or perceptual supervision (Zhang et al., 2021; Wang et al., 2021). Diffusion-based restoration models introduce pretrained generative priors into SR, as demonstrated by StableSR and DiffBIR (Wang et al., 2024; Lin et al., 2024). More recent one-step methods, including OSEDiff and OMGSR, reduce the sampling cost while retaining diffusion-based image priors (Wu et al., 2024; 2025). Other approaches explore alternative intermediate states or separate reconstruction and semantic enhancement, such as InvSR (Yue et al., 2025) and PiSA-SR (Sun et al., 2025).

Differences in architecture, degradation modeling, generative prior, and reconstruction objective lead to different reconstruction behaviors on the same input. This diversity provides the sources for the negative pool in our preference-pair construction strategy.

A.2 Preference Optimization for Generative Models and Super-Resolution

Preference optimization was first developed as a way to align model outputs with human judgments. Reinforcement learning from human feedback learns a reward model from ranked responses, while DPO directly optimizes a reference-relative preference objective using preferred and rejected samples (Ouyang et al., 2022; Rafailov et al., 2023). DPO has subsequently been extended to visual generation. Diffusion-DPO derives a tractable preference objective for diffusion models, and VideoDPO extends preference optimization to video generation with automatically constructed preference pairs (Wallace et al., 2024; Liu et al., 2025).

Preference optimization for real-world SR is a more recent research direction. DP^{2}O-SR constructs preference pairs by sampling multiple candidates from a common generator and ranking them with a combination of full-reference and no-reference image-quality models (Wu et al., 2026). GDPO-SR studies group-based preference optimization for one-step generative SR, using noise-aware candidate generation and content-adaptive rewards (Yi et al., 2026). DSPO incorporates semantic instance alignment and textual feedback into preference learning for image restoration (Cai et al., 2025). These methods obtain preference supervision through candidate ranking or feedback. In this work, we instead use the paired ground truth to specify the preferred target and vary the reconstruction model supplying the negative sample.

A.3 Vision Foundation Models and Perceptual Metrics

Perceptual metrics provide alternatives to pixel-wise errors when comparing restored images. LPIPS measures distances in deep feature spaces calibrated against human perceptual judgments, while DISTS combines structure and texture similarities in a learned full-reference metric (Zhang et al., 2018; Ding et al., 2020). No-reference metrics such as MUSIQ and CLIP-IQA estimate image quality without requiring a paired target, using multi-scale transformer representations or vision-language priors (Ke et al., 2021; Wang et al., 2023). In preference learning, such metrics can be used to rank candidate images and construct preference pairs. A separate choice is the representation in which image differences are measured during optimization.

Vision foundation models offer pretrained representations for this comparison. DINO and DINOv2 learn transferable visual features from large-scale image data without manual semantic labels, and DINOv2 has shown strong transferability across image-level and dense prediction tasks (Caron et al., 2021; Oquab et al., 2023). Register tokens further improve the behavior of Vision Transformer feature maps by reducing high-norm artifacts (Darcet et al., 2024). Related work such as REPA also demonstrates the usefulness of pretrained visual representations for guiding generative model training (Yu et al., 2024).

In this work, we adopt frozen DINOv2 features define the preference distances in the objective, which connects perceptual comparison in feature space with preference optimization.

B Experimental Details

B.1 OMGSR Fine-Tuning Details

We fine-tune OMGSR at a resolution of 512×512, using Stable Diffusion 2.1 as the underlying model. LoRA parameters are optimized in the UNet with rank 32 and in the VAE encoder with rank 16. Training uses the offline preference dataset constructed from LSDIR, with random horizontal flipping and rotation as data augmentations. Negative samples are loaded from a precomputed cache, with SwinIR, Real-ESRGAN, StableSR, and DiffBIR specified as the source models.

The total OMGSR training objective is

$$\mathcal{L}_{\mathrm{OMGSR}} = \lambda_{\mathrm{LRR}}\mathcal{L}_{\mathrm{LRR}} + \lambda_{\mathrm{Dv3D}}\mathcal{L}_{\mathrm{Dv3D}} + \lambda_{\mathrm{L1}}\mathcal{L}_{1} + \lambda_{\mathrm{HePO}}\mathcal{L}_{\mathrm{HePO}}. \tag{15}$$

Here, $\mathcal{L}_{\mathrm{LRR}}$ is the latent reconstruction regularization term inherited from OMGSR. It compares the latent representation of the LQ input produced by the trainable VAE encoder with the GT latent representation after noise is added at the prescribed intermediate diffusion timestep. The comparison is performed using mean squared error, encouraging the encoded LQ input to remain consistent with the diffusion-space target used by OMGSR.

The term $\mathcal{L}_{\mathrm{Dv3D}}$ is the perceptual reconstruction loss inherited from OMGSR. It measures the distance between the restored output and the GT image using statistical features extracted by a pretrained DINOv3 ConvNeXt-Large model. This branch provides perceptual supervision through a pretrained visual representation and is distinct from the DINOv2 features used by $\mathcal{L}_{\mathrm{HePO}}$ to define the preference objective. The term $\mathcal{L}_1$ provides pixel-wise reconstruction supervision, while $\mathcal{L}_{\mathrm{HePO}}$ is the proposed DINOv2 feature-space preference loss.

We set the loss weights to $\lambda_{\mathrm{LRR}} = 5$, $\lambda_{\mathrm{Dv3D}} = 5$, $\lambda_{\mathrm{L1}} = 0.5$, and $\lambda_{\mathrm{HePO}} = 0.5$. The latent reconstruction term uses diffusion timestep 273. The HePO-SR preference objective uses a frozen DINOv2 ViT-B/14 encoder with registers and features from layers $\{0, 3, 7, 9, 11\}$. The corresponding prior layer weights are $[0.40, 0.25, 0.20, 0.10, 0.05]$, and the decoupled coefficients are set to $(\beta_1, \beta_2) = (5, 3)$.

Training is conducted for 20,000 optimization steps on eight GPUs, with a per-device batch size of 4 and no gradient accumulation, giving a global batch size of 32. We use AdamW with a learning rate of 1×10^{-5}, weight decay of 1×10^{-4}, and a constant learning-rate schedule with 1,000 warm-up steps. The maximum gradient norm is 1.0, mixed-precision training uses bfloat16, and the random seed is 123. Model checkpoints are saved every 500 optimization steps. Exponential moving averages with decay 0.999 are maintained for the trainable UNet and VAE encoder parameters.

B.2 OSEDiff Fine-Tuning Details

We fine-tune OSEDiff from its pretrained initialization using rank-4 LoRA adapters. A frozen reference model is initialized from the same checkpoint. For each paired forward pass, the trainable model and reference model use the same VAE posterior noise. Training examples are loaded from a precomputed cache containing the low-quality (LQ) input, ground-truth (GT) image, and an offline negative sample. The negative sample is selected uniformly from the outputs of SwinIR, Real-ESRGAN, StableSR, and DiffBIR. Image-conditioned prompts are generated from the GT images using a frozen RAM/DAPE captioning model.

The training objective combines an L2 reconstruction loss, an LPIPS loss, and $\mathcal{L}_{\mathrm{HePO}}$, with respective weights of 1, 2, and 1. The preference loss uses features from a frozen DINOv2 ViT-B/14 model with registers, extracted from layers $\{0, 3, 7, 9, 11\}$. The corresponding prior layer weights are $[0.40, 0.25, 0.20, 0.10, 0.05]$, and the decoupled preference coefficients are $\beta_1 = 5$ and $\beta_2 = 3$.

Training runs for 20,000 steps on eight GPUs, with a per-GPU batch size of 4 and no gradient accumulation, yielding a global batch size of 32. We use mixed-precision FP16 and AdamW with a learning rate of 1×10^{-5}, $(\beta_1, \beta_2) = (0.9, 0.999)$, weight decay 10^{-2}, and $\epsilon = 10^{-8}$. The learning-rate schedule is constant with a 500-step warm-up setting; the training script scales this setting by

the number of processes. The maximum gradient norm is 1.0, and checkpoints are saved every 500 steps. Random horizontal flipping and rotation are enabled, and the random seed is 123.

C Additional Experimental Results

C.1 Reconstruction-Only Continued Training

To separate the effect of HePO-SR from that of additional fine-tuning, we construct a reconstruction-only control based on OMGSR. This variant removes the HePO-SR objective and continues training with the model-specific reconstruction objectives retained by OMGSR. We compare the original OMGSR checkpoint, reconstruction-only continued training, and the full HePO-SR configuration on RealSet65.

Table 4: Comparison of OMGSR variants on RealSet65. Reconstruction only denotes continued training with the model-specific reconstruction objectives and without HePO-SR. Bold and underlined values denote the best and second-best results, respectively.

Variant	CLIP-IQA↑	MUSIQ↑	MANIQA↑	TOPIQ↑	AFINE-NR↓
OMGSR baseline	<u>0.7400</u>	<u>70.29</u>	<u>0.6360</u>	<u>0.6911</u>	<u>-1.0230</u>
Reconstruction only	0.7365	69.88	0.6325	0.6834	-1.0121
HePO-SR	**0.7556**	**70.54**	**0.6370**	**0.7124**	**-1.0353**

Reconstruction-only continued training does not reproduce the improvements obtained with HePO-SR. Compared with the original OMGSR checkpoint, it decreases CLIP-IQA from 0.7400 to 0.7365, MUSIQ from 70.2900 to 69.8855, MANIQA from 0.6360 to 0.6325, and TOPIQ from 0.6911 to 0.6834. AFINE-NR also becomes worse, increasing from -1.0230 to -1.0121. In contrast, HePO-SR improves all five metrics over the original checkpoint, reaching 0.7556 CLIP-IQA, 70.5496 MUSIQ, 0.6370 MANIQA, 0.7124 TOPIQ, and -1.0353 AFINE-NR. Under this setting, extending optimization with reconstruction supervision alone is therefore insufficient to explain the quality gains; the gains arise when the reconstruction objectives are combined with the heterogeneous preference signal provided by HePO-SR.

C.2 Ablation on the Preference Feature Space

We further study how the feature space used to construct the preference objective affects restoration quality. In addition to the DINOv2 feature space used by HePO-SR, we evaluate VGG and pixel-space alternatives. The VGG variant extracts features from five layers and uses the same adaptive aggregation mechanism and base layer weights, $[0.40, 0.25, 0.20, 0.10, 0.05]$, as the main DINOv2 configuration. It therefore changes the feature representation while retaining the multi-level aggregation design. Pixel space does not provide a hierarchy of intermediate features, so its preference loss is computed directly from pixel distances without adaptive layer aggregation or layer weights.

Table 5: Ablation of the preference feature space on RealSet65 using OMGSR. Bold and underlined values denote the best and second-best results, respectively.

Feature Space	CLIP-IQA↑	MUSIQ↑	MANIQA↑	TOPIQ↑	AFINE-NR↓
OMGSR baseline	<u>0.7400</u>	<u>70.29</u>	<u>0.6360</u>	<u>0.6911</u>	<u>-1.0230</u>
VGG	0.7288	69.59	0.6298	0.6706	-1.0095
Pixel	0.7356	69.83	0.6313	0.6819	-1.0098
DINOv2	**0.7556**	**70.54**	**0.6370**	**0.7124**	**-1.0353**

DINOv2 achieves the best result on every evaluated metric, whereas both VGG and pixel space perform below the original OMGSR checkpoint. Compared with VGG under the matched five-level adaptive aggregation setting, DINOv2 improves CLIP-IQA by 0.0268, MUSIQ by 0.9515, MANIQA by 0.0072, and TOPIQ by 0.0418, while reducing AFINE-NR from -1.0095 to -1.0353. Because the number of feature levels, adaptive aggregation mechanism, and base layer weights are held fixed in this comparison, the difference is attributable to the feature representation rather than merely to the use of multi-level aggregation.

Pixel space performs better than VGG on all five metrics but remains below both the DINOv2 variant and the original OMGSR checkpoint. Since pixel space has no layer hierarchy, this comparison jointly changes the representation and removes adaptive aggregation; it should therefore be interpreted as a single-level preference-space alternative rather than a layer-matched ablation. Taken together, the matched VGG comparison and the pixel-space result show that the improvement is not obtained from an arbitrary perceptual or pixel distance. The multi-level DINOv2 representation provides the most effective feature space for the HePO-SR preference objective among the evaluated alternatives.

Fig. 5 provides a qualitative comparison of these preference spaces. Reconstruction-only continued training and the original OMGSR model leave the diagonal fabric seams relatively faint. The VGG and pixel-space variants recover parts of the seam pattern, but the lines have weaker contrast or less uniform continuity than those produced in DINOv2 space. The DINOv2 variant reconstructs clearer, more continuous, and more evenly spaced seams while preserving the smooth shading of the curved fabric surface. This example complements the quantitative results in Table 5: the benefit of DINOv2 is reflected in both the no-reference metrics and the recovery of coherent local structure rather than indiscriminate high-frequency enhancement.



Figure 5: Qualitative comparison of preference spaces using OMGSR. The left panel shows the LQ input and selected crop. The remaining panels compare reconstruction-only continued training (No HePO), the original OMGSR model, and HePO-SR optimized in VGG, pixel, and DINOv2 feature spaces.

C.3 Qualitative Analysis of Decoupled Temperature Coefficients

Fig. 6 compares the component dynamics under the three coefficient allocations evaluated in Table 3. We denote the ground-truth attraction component by A^+ and define the negative-sample component as A^-. Thus, $A^+ > 0$ indicates that the current prediction is closer to the ground truth than the frozen-reference output, while $A^- < 0$ indicates greater separation from the negative sample. With equal allocation, $(\beta_1, \beta_2) = (4, 4)$, A^- becomes increasingly negative throughout optimization while A^+ remains negative and gradually decreases. Emphasizing negative-sample separation with $(3, 5)$ strengthens this imbalance: A^- decreases further, accompanied by a larger decline in A^+. In contrast, under the main setting $(5, 3)$, A^+ recovers from its initial decrease and becomes positive, while A^- remains negative. The decoupled coefficients therefore alter how the preference margin is obtained: greater emphasis on ground-truth attraction promotes the desired combination of attraction toward the ground truth and separation from the negative sample.

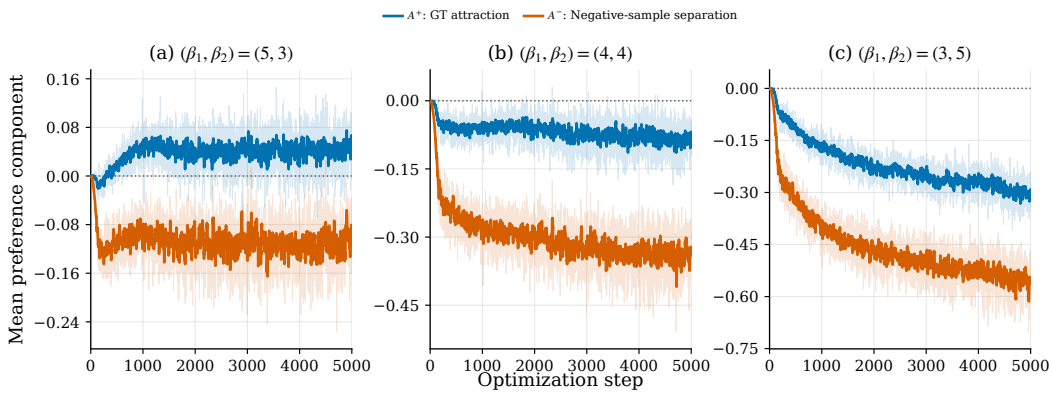

Figure 6: Training dynamics of the ground-truth attraction component A^+ and the negative-sample component A^- under different coefficient allocations. The three settings keep the overall coefficient scale fixed while changing their relative emphasis. Thin curves show unsmoothed per-step means, and thick curves show the corresponding smoothed trends.

Fig. 7 complements the coefficient ablation in Table 3 with a visual comparison. The original OMGSR result contains locally uneven and partially smeared fur around the cheek, neck, and body. Equal coefficient allocation, $(\beta_1, \beta_2) = (4, 4)$, sharpens the result but also introduces granular texture over the ears and face. Increasing the relative emphasis on negative-sample repulsion with $(3, 5)$ further amplifies irregular, spot-like patterns across the fur. By comparison, the main setting $(5, 3)$ produces smoother and more coherent fur transitions, retains clear facial and collar structures, and suppresses the spurious high-frequency patterns visible in the other coefficient settings. This visual trend supports the quantitative preference for $(5, 3)$: greater emphasis on ground-truth attraction prevents the objective from being improved primarily through excessive separation from negative samples.



Figure 7: Qualitative comparison of temperature-coefficient allocations using OMGSR. The main configuration, $(\beta_1, \beta_2) = (5, 3)$, places greater emphasis on ground-truth attraction; $(4, 4)$ uses equal allocation, and $(3, 5)$ places greater emphasis on negative-sample repulsion.

C.4 Offline and Online Training Comparison

HePO-SR can construct heterogeneous negative samples either before training and cache them for subsequent optimization (offline), or generate them on the fly during training (online). We compare the two strategies using OMGSR on all four evaluation benchmarks. Tables 6 and 7 report restoration quality, while Table 8 compares their training costs under the same global batch size.

Table 6: Comparison of offline and online negative construction on RealSet65. The dataset has no paired ground-truth images, so only no-reference metrics are reported. Bold and underlined entries indicate the best and second-best results, respectively.

Dataset	Strategy	Metrics				
		CLIP-IQA↑	MUSIQ↑	MANIQA↑	TOPIQ↑	AFINE-NR↓
RealSet65	Offline	<u>0.7556</u>	<u>70.54</u>	<u>0.6370</u>	**0.7124**	<u>-1.0353</u>
	Online	**0.7564**	**70.76**	**0.6386**	<u>0.7101</u>	**-1.0488**

Table 7: Comparison of offline and online negative construction on the three paired benchmarks. Bold and underlined entries indicate the best and second-best results, respectively.

Dataset	Strategy	Metrics								
		PSNR↑	SSIM↑	LPIPS↓	DISTS↓	CLIP-IQA↑	MUSIQ↑	MANIQA↑	TOPIQ↑	AFINE-NR↓
DRealSR	Offline	**27.50**	<u>0.7518</u>	<u>0.3378</u>	<u>0.2512</u>	**0.7205**	<u>67.35</u>	<u>0.6375</u>	**0.6842**	<u>-1.0058</u>
	Online	<u>27.45</u>	**0.7526**	**0.3293**	**0.2452**	<u>0.7197</u>	**67.67**	**0.6409**	<u>0.6791</u>	**-1.0205**
RealSR	Offline	**24.54**	<u>0.7068</u>	<u>0.3142</u>	<u>0.2416</u>	**0.6943**	**70.53**	<u>0.6772</u>	**0.7001**	<u>-1.1385</u>
	Online	<u>24.50</u>	**0.7077**	**0.3065**	**0.2359**	<u>0.6863</u>	<u>70.46</u>	**0.6797**	<u>0.6907</u>	**-1.1558**
DIV2K	Offline	**23.02**	<u>0.5811</u>	<u>0.3049</u>	<u>0.2080</u>	**0.7114**	<u>69.54</u>	<u>0.6554</u>	**0.7049**	<u>-1.0132</u>
	Online	<u>22.99</u>	**0.5835**	**0.2982**	**0.2035**	<u>0.7057</u>	**69.70**	**0.6586**	<u>0.6941</u>	**-1.0314**

The online strategy achieves better overall restoration quality. It improves four of the five no-reference metrics on RealSet65 and obtains better results on six of the nine metrics for each paired benchmark. In particular, online construction consistently improves SSIM, LPIPS, DISTS, MANIQA, and AFINE-NR across DRealSR, RealSR, and DIV2K. This advantage can be attributed to the greater negative-sample diversity available during online training: instead of repeatedly drawing preference supervision from a fixed offline cache, the model receives newly generated negative samples throughout optimization, which exposes it to a broader range of reconstruction errors.

Table 8: Training efficiency under a global batch size of 32, using a per-device batch size of 4 on eight GPUs without gradient accumulation. Relative overhead is measured against reconstruction-only training.

Training Objective	Efficiency			
	Global Batch	Time/Step (s)	Time/Sample (s)	Overhead
Reconstruction only	32	2.29	0.072	$1.00\times$
HePO-SR (offline)	32	3.94	0.123	$1.72\times$
HePO-SR (online)	32	15.60	0.488	$6.81\times$

The improved performance of online construction comes with a substantial computational cost because the heterogeneous SR model pool must be executed during training. Online construction increases the time per step from 3.94 seconds to 15.60 seconds, making it approximately $3.96\times$ slower than offline training and $6.81\times$ slower than reconstruction-only training. The offline strategy requires only a $1.72\times$ overhead over reconstruction-only training while remaining competitive in restoration quality. We therefore adopt offline negative construction in the main experiments to obtain a more practical balance between restoration performance and training efficiency.

D LIMITATIONS

The heterogeneous-negative strategy depends on the composition of the pretrained SR model pool. Different source models provide different reconstruction biases, so the resulting preference signal can vary with the selected pool and its sampling distribution. Our main experiments use offline negative construction, which reduces training cost but fixes the negative samples before optimization. Online construction provides continuously updated negative samples and achieves better overall performance in our comparison, but it incurs substantially higher training-time computation. This

creates a quality–efficiency trade-off between the two training modes. The current feature-space implementation uses DINOv2 with a fixed set of intermediate layers, prior layer weights, and decoupled coefficients. These choices are supported by our ablations, but their transferability to other feature encoders, layer selections, or restoration tasks has not been established.