CONTROLLED DATA REBALANCING IN MULTI-TASK LEARNING FOR REAL-WORLD IMAGE SUPER-RESOLUTION

Anonymous authors
Paper under double-blind review

ABSTRACT

Real-world image super-resolution (Real-SR) is a challenging problem due to the complex degradation patterns in low-resolution images. Unlike approaches that assume a broadly encompassing degradation space, we focus specifically on achieving an optimal balance in how SR networks handle different degradation patterns within a fixed degradation space. We propose an improved paradigm that frames Real-SR as a data-heterogeneous multi-task learning problem. Our work addresses degradation task imbalance in the paradigm through coordinated advancements in task definition, imbalance quantification, and adaptive data rebalancing. Specifically, we introduce a novel task definition framework that segments the degradation space by setting parameter-specific boundaries for degradation operators, effectively reducing the task quantity while maintaining task discrimination. We then develop a focal loss based multi-task weighting mechanism that precisely quantifies task imbalance dynamics during model training. Furthermore, to prevent sporadic outlier samples from dominating the gradient optimization of the shared multi-task SR model, we strategically convert the quantified task imbalance into controlled data rebalancing through deliberate regulation of task-specific training volumes. Extensive quantitative and qualitative experiments demonstrate that our method achieves consistent superiority across all degradation tasks.

1 INTRODUCTION

Image super-resolution aims to recover high-resolution (HR) images from degraded low-resolution (LR) inputs. Unlike conventional non-blind Dong et al. (2014); Zhang et al. (2018d;c) and blind Huang et al. (2020); Wang et al. (2021a); Gu et al. (2019) SR methods assuming simplified degradations, real-world SR (Real-SR) must contend with complex, unknown degradations, such as sensor noiseZhang et al. (2021), motion blur Bistroń & Piotrowski (2024), compression artifacts Wang et al. (2024a), and optical aberrations Zhang et al. (2021). These distortions corrupt high-frequency details and vary spatially, making traditional approaches inadequate for demanding applications Zhu et al. (2020); Liebel & Körner (2016); Wan et al. (2020). Consequently, Real-SR frameworks must accommodate diverse, unpredictable degradation patterns while preserving perceptual fidelity.

To bridge the simulation-to-reality gap, prior studies have prioritized the design of sophisticated degradation models to synthesize training pairs that mimic real-world conditions. Pioneering works like RealESRGAN Wang et al. (2021b) employ randomized degradation pipelines combining degradation operators like blur, noise, resize and compression in varying sequences. Such methods train a single SR network on synthetically degraded data, aiming for the broadest coverage of potential degradation cases.

On the other hand, our work shifts focus to examining the consistency of SR network optimization across different degradation patterns within a fixed degradation space. TGSR Zhang et al. (2023) has posited that the rebalancing of diverse degradation patterns can be fundamentally framed as a multi-task learning paradigm with data heterogeneity, where each task corresponds to a set of HR-LR image pairs generated under a specific degradation pattern. To intuitively demonstrate the phenomenon of task imbalance during the training of real-SR network, we sampled multiple degradation patterns following the degradation process of RealESRGAN. We then quantified the degree of task competition

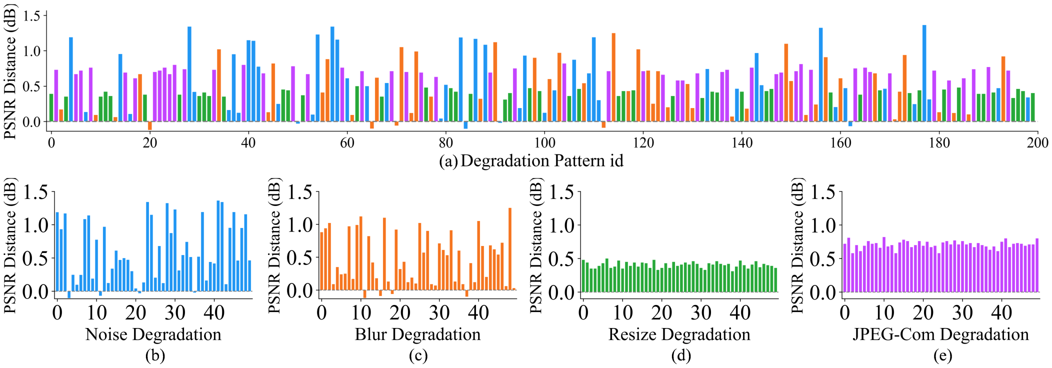


Figure 1: Task-defining capacity of each degradation operator.

capacity for specific patterns by comparing the performance between single-task networks, each trained exclusively on one degradation pattern, and the multi-task network trained on all patterns. As illustrated in Fig. 1(a), the learning progress for each task exhibits significant variance. This disparity reveals the inherent task imbalance problem the model confronts when attempting to address multiple degradation patterns simultaneously without external intervention.

In this paper, we delve further into this paradigm. We observe that not all degradation operators exhibit the same strong discriminative power when defining tasks, some operators contribute redundantly to task differentiation. We conduct a further validation experiment to empirically analyze the discriminative capacity of individual degradation operators. The results in Fig. 1 reveal that operators such as noise injection and blur artifacts exert significantly stronger influence on task formulation and imbalance, demonstrating a higher discriminative power compared to others like down-sampling or JPEG compression. We thus propose simplifying the paradigm definition, focusing primarily on the more impactful degradation operators as the defining factors for distinct tasks, thereby deprioritizing those less influential in task differentiation. Concurrently, we partition the degradation space by the parameter ranges of these influential operators, redefining tasks as sets of sample pairs from each resulting subspace.

Building upon an explicit definition of degradation tasks, we systematically investigate task imbalance within the multi-task SR paradigm and identify two critical issues. (1) Tasks possess various inherent difficulties. Methods relying on short-term absolute performance gains (like Zhang et al. (2023)) wrongly assume that equal gains reflect equivalent task competition, without accounting for this variance in difficulty. (2) External regulation to mitigate task imbalance inevitably introduce secondary imbalances. As both task imbalance and relative learning progress evolve significantly during training, static approaches are fundamentally limited. For example, the fixed task grouping in TGSR cannot adapt to these dynamic imbalances induced by the regulation itself. To address these challenges, we introduce a dynamic multi-task loss-weighting algorithm that both implicitly captures task difficulty by referencing single-task model performance and adapts to evolving training dynamics by periodically re-evaluating loss weights.

Previous studies Grégoire et al. (2024) have demonstrated that the gradient stability of multi-task loss re-weighting methods is susceptible to the negative impact of outliers. To address this stability challenge, we prove that under our proposed task definition framework, the weighting term of multi-task losses is equivalent to controlling the sample quantity for each task, and that this data rebalancing can provide a more stable training process.Consequently, we strategically regulate task-specific sample generation at different intervals, address task imbalance through controlled data rebalancing. Notably, the conventional practice of constructing LR training images from HR counterparts through degradation models enables straightforward manipulation of task-specific data quantities. Our findings reveals new potential for multi-task learning paradigms in real-world SR scenarios.

Our main contributions are: **(1)** We reframe the paradigm of multi-task learning for real-world image super-resolution by proposing a novel task definition framework. This definition removes redundant degradation parameters while preserving degradation model's capacity. **(2)** We develop a difficulty-aware loss weighting mechanism to dynamically quantify and address task imbalance during training. **(3)** We theoretically prove that task imbalance, as quantified by loss weighting, is equivalent to an adjustable data quantity imbalance , and that the latter provides a more stable optimization process. Notably, this strategy can be readily integrated with other Real-SR methods.

2 RELATED WORKS

Real-world super-resolution. Compared to non-blind SR methods, Real-SR requires handling complex degradations. Early works Kim et al. (2016); Ledig et al. (2017); Zhang et al. (2018a) model degradation spaces using Gaussian kernels and noise. BSRGAN Zhang et al. (2021) and RealESRGAN Wang et al. (2021b) advance explicit degradation modeling via shuffled and high-order strategies. Recently, diffusion-based models Wu et al. (2024a;b); Fan et al. (2024) address intricate degradations via iterative refinement but incur high inference costs. Somewhat similarly, DASR Liang et al. (2022) proposes three degradation patterns and introduces an additional network for degradation recognition. In contrast, our method prioritizes balanced training progression across distinct degradation patterns within a fixed degradation space, rather than expanding it to accommodate broader degradation variations.

Multi-task learning. Multi-task learning methodologies can generally be classified into three main types: (1) Task rebalancing Guo et al. (2018); Kendall et al. (2018) mitigates task imbalance via loss weighting or gradient manipulation. (2) Task grouping Zamir et al. (2018); Fifty et al. (2021) identifies synergistic tasks for joint learning, leveraging concepts like task taxonomy Zamir et al. (2018) or affinity scores Fifty et al. (2021). (3) Architecture design methods include hard parameter sharing (shared encoder, distinct decoders) Kokkinos (2017) and soft parameter sharing (separate networks with cross-talk) Misra et al. (2016). Our approach adopts the most relevant task rebalancing methods to quantify and address the issue of degradation task imbalance within SR models.

3 METHOD

Problem Formulation. Real-SR addresses the ill-posed inverse problem of reconstructing a high-resolution (HR) image $x \in \mathbb{R}^{H\times W\times c}$ from its degraded low-resolution (LR) observation $y \in \mathbb{R}^{h\times w\times c}$, where the degradation process $\mathcal{D}(\cdot)$ explicitly encapsulates the complexities inherent in physical imaging systems. Formally, the forward degradation model can be expressed as:

$$y = \mathcal{D}(x;\Theta) = \left(f_{\theta_n} \circ \cdots \circ f_{\theta_2} \circ f_{\theta_1}\right)(x) \tag{1}$$

Here, $\mathcal{D}(x;\Theta)$ is a parameterized degradation model with a set of empirically predefined parameters $\Theta = \{\theta_n, \cdots, \theta_2, \theta_1\}$, each function f_{θ_i} represents an individual degradation operator, such as blurring, noise addition, down-sampling, or JPEG compression, applied in a sequential manner. Existing Real-SR methods (e.g. RealESRGAN Wang et al. (2021b)) customize $\mathcal{D}$ through the binding of stochastic sampling parameters and optimize the sequential ordering to mimic real imaging pipelines.

We recall the conception proposed by TGSR, mapping the Real-SR onto a multi-task learning problem. Since the degradation model $\mathcal{D}$ can be regarded as a vast degradation space, an SR task τ is defined as training pairs $(X, Y = d(X))$, where the degradation $d(\cdot)$ is sampled from the infinite degradation space $\mathcal{D}$ and applied to an HR image set X to produce an LR image set Y. Due to the infinite size of $\mathcal{D}$, TGSR chose to sampled a large number of fixed degradation parameters to form the task space, and tried to handle the task competition problem from a multi-task learning perspective.

The pipeline of our framework is shown in Fig. 2. The detailed introduction of each part is below.

3.1 TASK REDEFINATION VIA DEGRADATION SUBSPACE PARTITIONING

To quantitatively validate the task discrimination ability of different degradation operators, we conduct a validation experiment. For each operator (e.g., Gaussian blur, noise, JPEG compression and down-sampling), we create 50 degradation configurations by varying only the severity of the target operator while keeping others fixed. We then fine-tune a pre-trained Real-SR model on each configuration separately, yielding 50 operator-specific single-task models. The PSNR differences between the pre-trained model and each fine-tuned model are used to assess the task-specific learning progress. A larger variance in PSNR improvements indicates stronger task imbalance. As illustrated in Fig. 1(b) and Fig. 1(c), noise and blur operators exhibit substantial variance, indicating severe task imbalance during joint training of shared models. In stark contrast, Fig. 1(d) and Fig. 1(e) reveal a markedly stable PSNR distances variation curve while varying down-sampling (and JPEG compression) degradation parameters, implying weaker task imbalance under this set of configurations.

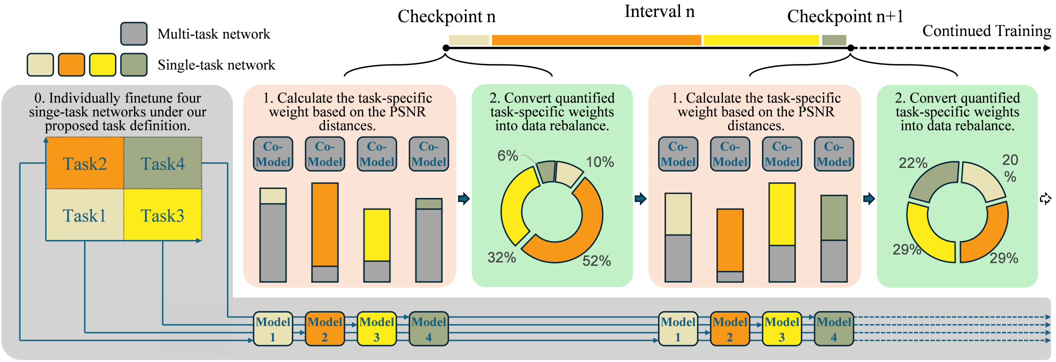
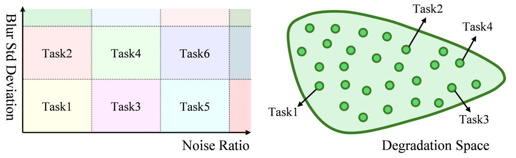


Figure 2: Overview of our proposed pipeline. At the beginning of each training interval, PSNR distances for each task are measured relative to the respective single-task networks to compute task weights, which are then converted into the task data volume for the duration of that interval.

The results in Fig. 1 show that different degradation operators f_{θ_i} demonstrate heterogeneous contribution weights in task formulation, as their representational capacities vary significantly across the degradation space $\mathcal{D}$. In other words, degradation task clusters formed by varying intensity levels of noise contamination operators (and blur artifacts operators) demonstrate significantly stronger task imbalance and negative transfer problems compared to those formed by other degradation operators (e.g., down-sampling or JPEG compression artifacts).

Building upon these insights, we strategically shift our task definition focus to noise contamination operators and blur artifacts operators while excluding redundant hyperparameters (e.g. fixed down-sampling ratios or predetermined compression levels), thereby refining the task definition into a more practical and performance-impactful paradigm. Meanwhile, unlike Zhang et al. (2023) samples a number of fixed degradation task parameters to represent the whole degradation space - this random and discrete task sampling approach inevitably compromises the expressiveness of the original degradation space $\mathcal{D}$ - we establish the parameter values of noise and blur operators as degradation task boundaries to partition the degradation space $\mathcal{D}$, redefine tasks as training data pairs generated from distinct continuous degradation subspaces.

Specifically, we partition the degradation space $\mathcal{D}$ into n distinct subspaces based on the parametric values of noise and blur operators, $\mathcal{D}_{\text{sub}} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n\}$. Given the whole set of HR images $\mathcal{X}$ ($|\mathcal{X}| \gg n$), we construct n SR tasks $\mathcal{T} = \{\tau_i = (X_i, Y_i)\}_{\text{i=1}}^{\text{n}}$, where each $X_i \in \mathcal{X}$ corresponds to an LR counterpart Y_i generated by applying degradation parameters from the subspace $\mathcal{D}_i \in \mathcal{D}_{\text{sub}}$. Therefore, without compromising the expressive power of degradation models, Real-SR is reformulated as a multi-task learning framework, where a shared model collaboratively addresses degradation tasks derived from distinct degradation subspaces. Fig. 3 illustrates the differences in task definitions between our method and TGSR.



(a) Ours task definition (b) TGSR's task definition

Figure 3: Divergent task definition between our method and TGSR.

Analysis for Task Imbalance. From the multi-task learning perspective, task imbalance occurs when tasks exhibit uneven learning progress or disparate resource allocation during joint optimization, leading to compromised convergence or sub-optimal performance for individual tasks Yu et al. (2020). For instance, in tasks combining semantic segmentation and depth estimation, gradients from one task may dominate parameter updates, suppressing the learning of other tasks.

In the context of conventional multi-task learning, task imbalance stems primarily from task diversity, reflecting the absence of adequately shared latent representations across different tasks. On the other hand, within our task definition framework, the involved tasks are semantically homogeneous as they all belong to the super-resolution domain, yet exhibit heterogeneity in their degradation prior distributions. As we illustrated in Fig. 1(b) and (c), the degradation tasks partitioned by varying severity levels may compete for the shared capacity of jointly trained models. This competition could induce an optimization bias where the model preferentially learns to restore textures with higher-frequency distortions (e.g. severely degraded regions requiring intensive reconstruction), while insufficiently addressing the restoration needs of mildly degraded images.

Most real-world SR algorithms utilize randomized degradation parameters to generate LR training data. This unbiased generation approach is equivalent to giving all degradation tasks equal optimization priority. However, when the entire task cluster exhibits severe task imbalance, owing to lack of external regulation, such random task selection mechanism tends to impede the optimization efficiency of shared models and ultimately restricts the performance of under-optimized tasks.

3.2 ADAPTIVE LOSS WEIGHTING FOR MULTI-TASK SUPER-RESOLUTION

Loss function re-weighting is an effective external regulatory mechanism to address task imbalance in multi-task field. In this section, we present our methodology for quantifying task imbalance to determine adaptive loss weights. As analyzed in Sec. 1, our loss weighting strategy must account for two critical aspects: (1) inherent task difficulty discrepancies arising from varying degradation severity levels, (2) dynamic evolution of task imbalance throughout the training process. Drawing inspiration from Yun & Cho (2023), we formulate our adaptive SR task weighting scheme based on the focal loss.

Focal loss was proposed to address class imbalance in dense detection scenarios, and it modifies the standard cross-entropy loss by introducing a focusing weighting term $(1 - p_t)$ to prioritize hard-classified examples. The loss function is formulated as:

$$FL\left(p_t\right) = \left(1 - p_t\right)^{\gamma} CE = -\left(1 - p_t\right)^{\gamma} \log\left(p_t\right) \tag{2}$$

where p_t denotes the model's estimated probability for the true class. The focusing weighting term $(1 - p_t)$ dynamically scales loss contributions by suppressing well-classified samples while preserving higher weights for ambiguous cases, forcing optimization to prioritize challenging examples.

Building upon the focusing weighting term, we define the loss weight for task τ as follows:

$$w_\tau = \mathrm{DP}\left(N^{\tau}_{\text{single}}, N^{\tau}_{\text{multi}}\right) \tag{3}$$

where $\mathrm{DP}\left(N^{\tau}_{\text{single}}, N^{\tau}_{\text{multi}}\right)$ denotes the PSNR distance on validation set generated by degradation task τ between the dedicated single-task network trained on task τ and multi-task network built on joint training. Compared with solely relying on the absolute PSNR values of the shared model, this relative measurement mitigates the impact of inherent task difficulty variations. Tasks exhibiting smaller PSNR distances are identified as dominant competitors in current shared training process, and we accordingly assign them lower weights to balance gradient updates across tasks.

However, since the shared model serves as a mediator for cross-task knowledge transfer during joint training, the validation PSNR of the shared model may surpass that of the single-task network, resulting in negative PSNR distances. To address this phenomenon, we reformulate the task weights through applying an exponential transformation. Meanwhile, to further accommodate the evolving task imbalance dynamics during joint optimization, we introduce temporal partitioning of the training schedule into K intervals, $T = \{T_1, T_2, \ldots, T_k, \ldots, T_K\}$, each spanning a fixed number of iterations. At the onset of every interval T_k, we dynamically calibrate all task weights by recomputing the exponential weighting formula given by Equ. 3 using the latest validation metrics. After the final weight normalization, the proposed multi-task SR loss can be formulated as:

$$\begin{aligned} &\text{For each } T_k \in T, \quad \mathcal{L}_{\text{SR}}\left(T_k\right) = \sum_{\tau \in \mathcal{T}} w^k_\tau \mathcal{L}_\tau \\ &w^k_\tau = \exp\left(\mathrm{DP}\left(N^{\tau}_{\text{single}}, N^{\tau}_{\text{multi}}\right)\right) \end{aligned} \tag{4}$$

3.3 CONTROLLED DATA REBALANCING FOR ADDRESSING TASK IMBALANCE

In this section, we theoretically prove the equivalence between loss function re-weighting and sample rebalancing . We further demonstrate that the latter enhances stability by providing a smaller gradient variance. As the proof is intricate and relatively self-contained, we only present the conclusion here due to space constraints. The detailed derivation is provided in Appendix A.2.

Conclusion 1 (Equivalence of Loss Reweighting and Data Rebalancing).

$$N_i = \frac{w^k_{\tau_i}}{\alpha} = N \cdot w^k_{\tau_i}, \qquad \alpha = \frac{1}{N}.$$

where N_i denotes the number of samples of task τ_i within interval k, and $w^k_{\tau_i}$ is the normalized loss weight assigned to task τ_i in interval k.

Conclusion 2 (Variance Comparison between Re-weighting to Re-sampling).

$$\boxed{\Delta \;=\; \sum_{j=1}^{n} \Delta_j = \sum_{j=1}^{n} w_j (n w_j - 1)\, \mathbb{E}[G_j^2].}$$

Here $\Delta = \mathrm{Var}(Z_{\mathrm{rw}}) - \mathrm{Var}(Z_{\mathrm{rs}})$ measures the variance reduction, and Δ_j denotes the contribution from task j. Tasks with $nw_j > 1$ contribute positively, demonstrating that reweighting amplifies their variance, while resampling mitigates this effect. This explains why data rebalancing leads to more stable training dynamics in practice.

4 EXPERIMENT

Training Datasets. We employ DIV2K Agustsson & Timofte (2017), Flickr2K Agustsson & Timofte (2017) and OutdoorSceneTraining Wang et al. (2018) datasets for training. We use the same degradation pipeline as RealESRGAN Wang et al. (2021b) to synthesize LR-HR pairs.

Test Datasets. For evaluation, we construct the DIV2K4Level dataset from the DIV2K validation set, comprising four distinct validation subsets aligned with the four degradation subspaces defined in our task formulation. Each subset consists of 100 image pairs. Beyond this, we evaluate our model on a synthetic test set, DIV2K-random, as well as two real-world datasets, RealSR Cai et al. (2019) and DRealSR Wei et al. (2020). The synthetic dataset comprises 100 image pairs. The LQ images are synthesized by applying random degradations using Real-ESRGAN Wang et al. (2021b) to DIV2K_valid Agustsson & Timofte (2017).

Evaluation Metrics. For evaluating our method, we apply both fidelity and perceptual quality metrics. Fidelity metrics include PSNR and SSIM Wang et al. (2004) (calculated on the Y channel in YCbCr space). Perceptual quality metrics include LPIPS Zhang et al. (2018b).

Degradation Task Definition Settings. In the domain of Real-SR, partitioning the degradation space based on the semantic properties of operators to apply a "divide and conquer" approach is not uncommon. Chen et al. (2023a) employ partitioning schemes with blur and noise biased degradation spaces to train scenario-specific data augmentation models. Wang et al. (2021a) defines three degradation types and introduces an auxiliary network to estimate the degradation of the input image.

Broadly referencing the overall degradation ranges adopted by these methods, we partition the overall degradation space into 4 subspaces by uniformly dividing the value ranges of two parameters: blur kernel standard deviations and noise injection rate. These subspaces are termed mild, blur, noise, and severe. We adopt the high-order degradation model proposed by Real-ESRGAN as the Real-SR degradation model in our experiments. Detailed degradation settings and further analysis of the partitioning boundaries are provided in Appendix A.1.

4.1 COMPARISON WITH STATE-OF-THE-ART

We compare our method with the state-of-the-art methods, contrastive learning method include DASR Liang et al. (2022), Gan-based methods include RealESRGAN Wang et al. (2021b), SwinIR Liang et al. (2021), and MM-RealSR Mou et al. (2022). Diffusion-based methods include ResShift Yue et al. (2023), SinSR Wang et al. (2024b), TSD-SR Dong et al. (2025), and AdcSR Chen et al. (2025). Officially released pre-trained models are used for the compared methods.

Quantitative Comparisons on DIV2K4Level. DASR defines three degradation types for image degradation estimation. It achieves exceptionally high PSNR metrics, which is often attributed to specific hyperparameter configurations in its loss function weighting. However, its performance on LPIPS is notably poor (the worst and exhibiting a significant disparity). In contrast, MM-RealSR

	DIV2K_mild			DIV2K_blur			DIV2K_noise			DIV2K_severe		
	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM
DASR(CVPR'21) Liang et al. (2022)	23.72	0.5581	0.6508	23.40	0.6466	0.5677	23.68	0.6403	0.5782	23.46	0.6344	0.5863
RealESRGAN(ICCV'21) Wang et al. (2021b)	23.62	0.4273	0.6382	23.44	0.4343	0.6314	23.58	0.4330	0.6364	23.39	0.4403	0.6292
RealSwinIR(ICCV'21) Liang et al. (2021)	23.52	0.4372	0.6374	23.46	0.4416	0.6337	23.45	0.4424	0.6337	23.39	0.4471	0.6299
MM-RealSR(ECCV'22) Mou et al. (2022)	23.44	0.4293	0.6468	23.45	0.4352	0.6430	23.44	0.4343	**0.6448**	23.45	0.4412	0.6340
ResShift(NIPS'23) Yue et al. (2023)	23.71	0.4646	0.6174	23.41	0.4711	0.6110	23.57	0.4728	0.6090	23.48	0.4796	0.6031
SinSR(CVPR'24) Wang et al. (2024b)	23.39	0.4835	0.5654	23.29	0.4901	0.5585	23.24	0.4938	0.5524	23.13	0.5004	0.5456
TSD-SR(CVPR'25) Dong et al. (2025)	22.16	0.4256	0.5753	22.04	**0.4308**	0.5694	22.16	0.4380	0.5727	22.03	**0.4333**	0.5666
AdcSR(CVPR'25) Chen et al. (2025)	23.27	0.4272	0.6050	23.18	0.4332	0.6002	23.22	0.4336	0.6000	23.13	0.4398	0.5949
TGSR(NIPS'23) Zhang et al. (2023)	23.64	0.4283	0.6425	23.49	0.4354	0.6359	23.54	0.4339	0.6395	23.39	0.4415	0.6330
Ours	**23.79**	**0.4190**	**0.6511**	**23.66**	**0.4308**	**0.6483**	**23.73**	**0.4272**	0.6424	**23.54**	0.4383	**0.6396**

Table 1: Quantitative results of different methods on DIV2K4Level.

	RealSRset-Nikon			RealSRset-Canon			DRealSR			DIV2K-random		
	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM	PSNR	LPIPS	SSIM
DASR(CVPR'21) Liang et al. (2022)	**27.15**	0.4391	0.7598	**27.67**	0.4236	0.7893	**30.41**	0.4387	0.8261	**24.43**	0.5726	0.6438
RealESRGAN(ICCV'21) Wang et al. (2021b)	25.62	0.3820	0.7607	26.06	0.3629	0.7864	28.64	0.3766	0.8052	23.50	0.4341	0.6338
RealSwinIR(ICCV'21) Liang et al. (2021)	25.72	0.3768	0.7663	26.33	**0.3573**	0.7945	28.24	0.3819	0.7983	23.45	0.4424	0.6335
MM-RealSR(ECCV'22) Mou et al. (2022)	23.54	0.3822	0.7425	24.06	0.3606	0.7708	26.97	0.3723	0.7972	23.43	0.4353	0.6434
ResShift(NIPS'23) Yue et al. (2023)	25.11	0.4736	0.6999	25.87	0.4626	0.7501	27.12	0.4689	0.7407	23.69	0.4724	0.6101
SinSR(CVPR'24) Wang et al. (2024b)	25.68	0.4702	0.6843	26.35	0.4596	0.7279	28.45	0.4496	0.7516	23.25	0.4921	0.5552
TSD-SR(CVPR'25) Dong et al. (2025)	23.71	0.4023	0.6849	23.89	0.3956	0.7073	26.23	0.4019	0.7160	22.10	0.4338	0.5709
AdcSR(CVPR'25) Chen et al. (2025)	25.49	0.3925	0.7187	25.71	0.3796	0.7457	28.21	0.3822	0.7723	23.20	0.4343	0.5996
TGSR(NIPS'23) Zhang et al. (2023)	25.88	0.3862	0.7688	26.35	0.3660	0.7954	28.91	0.3839	0.8142	23.51	0.4348	0.6376
Ours	26.16	**0.3763**	**0.7728**	26.54	0.3580	**0.7960**	29.27	**0.3697**	**0.8393**	23.72	**0.4254**	**0.6441**

Table 2: Quantitative Comparisons on Synthetic and Real-World Test Sets.

and RealESRGAN utilize higher-order degradation models. These approaches typically sacrifice some PSNR performance to achieve enhanced perceptual quality. SwinIR, which leverages the Swin Transformer architecture for image super-resolution, demonstrates strong performance across selected evaluation metrics. Diffusion-based methods generally excel on non-reference metrics, indicating their capacity to generate perceptually realistic images. However, the outputs from these models can sometimes exhibit considerable differences from the ground truth images, consequently leading to lower performance on reference-based metrics. The result is shown in Tab. 1. Our proposed method consistently demonstrates either the best or second-best performance across all degradation settings, underscoring the effectiveness of our balanced approach to these degradation tasks.

Quantitative Comparisons on Synthetic and Real-World Test Sets. Our method demonstrates consistent superiority not only on specialized task-specific benchmarks but also across the synthetic DIV2K-random dataset and real-world test sets (RealSRset & DRealSR). The result is shown in Tab. 2. These results demonstrate that our method does not merely achieve a performance balance among the delineated degradation tasks, but rather fulfills our foundational goal: achieving an optimal balance in how SR networks handle different degradation patterns within a fixed degradation space.

Qualitative Comparisons. The qualitative comparisons in Fig. 4 demonstrate the superiority of our framework over mainstream Real-SR methods. Under heavy and mixed degradations (first two rows), competing approaches either fail to remove blur and noise or introduce unnatural textures, whereas our model effectively restores structures with minimal artifacts. In high-contrast regions (last row), our results exhibit crisper edges and more faithful texture reconstruction compared to the overly smooth or artifact-ridden outputs of prior methods.

4.2 ANALYSIS AND ABLATION STUDY

Task Imbalance Dynamics Analysis during Training. Fig. 5 visualizes the adaptability of TGSR and our proposed method to dynamic task imbalance. For Fig. 5(a), we selected two checkpoints during the training process and employed the grouping indicator from TGSR to evenly group 100 randomly sampled discrete tasks into 4 groups. The resulting changes in grouping outcomes reveal how task imbalance evolves. Because TGSR's task grouping is executed only once at initialization and remains fixed throughout training, this static strategy cannot adapt to shifting the secondary imbalances that arise from the optimization process itself. This limitation is evidenced by the marked disparity in the grouping results between the two intervals. In contrast, Fig. 5(b) demonstrates the stable optimization and convergence of task weights in our method, showcasing its ability to dynamically track and adapt to the evolving task imbalance.

Algorithm Integrability Verification. To validate the algorithmic integrability of our method, we conduct extensive experiments by integrating our framework into mainstream real-world image

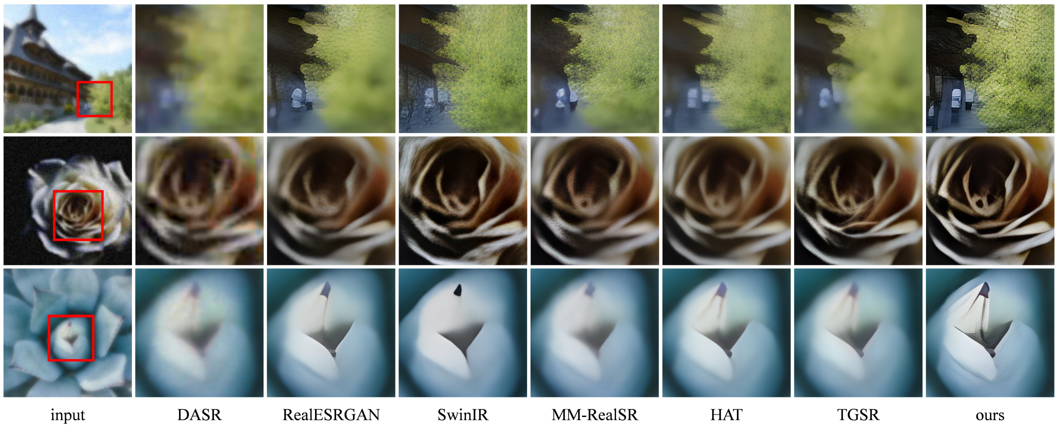
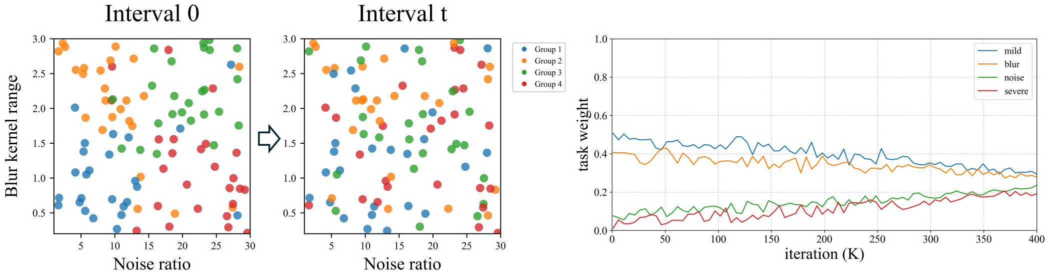


Figure 4: Qualitative results of different methods. Zoom in for details.



(a) Task Grouping Result for TGSR

(b) Evolution of Task Weights

Figure 5: A comparison of the adaptability of TGSR and our method to the dynamic evolution of task imbalance. Tasks grouped into same groups are represented with identical colors in (a).

super-resolution models. As our approach builds a multi-task learning framework based on training data heterogeneity modeling, it imposes no assumptions on the specific architecture of the base SR model. We select 4 mainstream models-RealSRGAN Ledig et al. (2017), RealESRGAN Wang et al. (2021b), SwinIR Liang et al. (2021), and HAT Chen et al. (2023b)—as integration baselines. Without modifying their original architectures or training configurations, we apply our task definition framework and dynamic loss weighting mechanism to their training pipelines. As shown in Tab. 3, our method achieves significant performance gains across real-world datasets, Fig. 6 shows the visual improvements of our method compared to the baselines, for regions with composite degradations and complex textures, baseline methods tend to uniformly smear details, whereas our method demonstrates superior detail recovery.

Study of the performance upper bound. We train 4 dedicated single-task networks for each degradation subspace, and compare their performance against our multi-task network on respective DIV2K4Level validation sets. As shown in Fig. 7, our multi-task network achieved performance metrics comparable to those of the single-task networks across all degradation tasks. Notably, for the 'noise' and 'severe' degradation tasks, our multi-task network achieved PSNR improvements of 0.08 dB and 0.07 dB, respectively, surpassing the performance of the single-task models. These results highlight the capability of our framework to leverage cross-task information interaction, demonstrating its potential in addressing highly degraded real-world image super-resolution scenarios.

Ablation Study for Loss Function Weighting Algorithms. To investigate the effectiveness of our proposed loss weighting method in addressing task imbalance, we conduct an ablation study on the loss weighting component. We compare our approach with several classical multi-task loss weighting algorithms, including RLW Lin et al. (2021), DWA Liu et al. (2019) and GLSChennupati et al. (2019). As shown in Tab. 4, our task weighting algorithm achieves superior performance in real-world image super-resolution applications. In the table, "Diff" denotes difficulty-aware algorithms, while "Dyna" denotes dynamic algorithms. This improvement can be attributed by the fact that the loss function weighting of traditional multi-task learning is mainly oriented to heterogeneous task semantics (e.g.,

	RealSRset-Nikon		RealSRset-Canon	
	PSNR	LPIPS	PSNR	LPIPS
RealSRGAN	24.71	0.4159	25.42	0.3902
RealSRGAN+ours	**25.27**	**0.3997**	**25.86**	**0.3752**
RealESRGAN	25.62	0.3820	26.06	0.3629
RealESRGAN+ours	**26.16**	**0.3763**	**26.54**	**0.3580**
RealSwinIR	25.72	0.3768	26.33	0.3573
RealSwinIR+ours	**26.68**	**0.3652**	**27.01**	**0.3415**
RealHAT	27.17	**0.3612**	27.72	**0.3360**
RealHAT+ours	**27.65**	0.3691	**28.01**	0.3417

Table 3: The quantitative results of applying proposed framework to mainstream Real-SR methods.

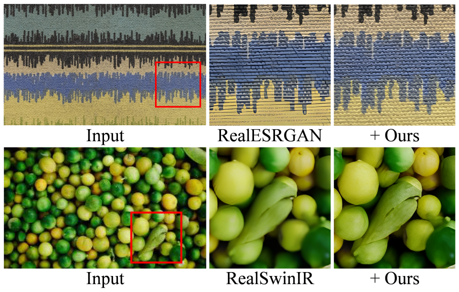


Figure 6: The qualitative results of algorithm integrability.

Diff	Dyna	Methods	RealSRset-Nikon		RealSRset-Canon	
			PSNR	LPIPS	PSNR	LPIPS
✓		RLW	26.02	0.3833	26.14	0.3699
	✓	DWA	25.82	0.3759	26.38	0.3594
✓	✓	GLS	25.68	0.3801	26.10	0.3606
✓	✓	Ours	**26.16**	**0.3763**	**26.54**	**0.3580**

Table 4: Ablation study for loss weighting algorithms.

Strategy	#Tasks	RealSRset-Nikon		RealSRset-Canon	
		PSNR	LPIPS	PSNR	LPIPS
re-sampling	2	25.89	0.3759	26.29	0.3580
re-sampling	3	26.02	0.3783	26.40	0.3615
re-sampling	5	26.14	**0.3749**	26.48	**0.3552**
re-weighting	4	25.83	0.3788	26.31	0.3597
re-sampling	4	**26.16**	0.3763	**26.54**	0.3559

Table 5: Ablation study for the quantity of tasks and data rebalancing strategy.

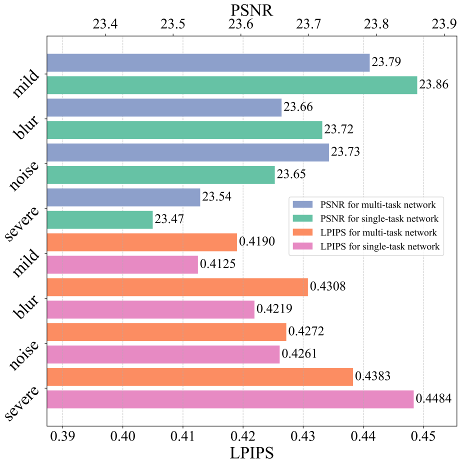


Figure 7: Comparison of Single-task Networks and Proposed Multi-task Network.

cross-domain tasks like segmentation and depth estimation), while our proposed method aims to better handle task imbalance under different data distribution definitions.

Other Ablation Study Results. We conducted supplementary ablation studies to validate the effectiveness of our task-imbalance-to-data-imbalance transformation approach described in Sec. 3.3, and investigating model performance under varying granularities of task grouping configurations. As demonstrated in Tab. 5, our algorithm consistently achieves stable performance gains across different task quantity settings. Notably, while finer task partitioning yields finite incremental gains, it incurs linearly increasing training costs proportional to the number of task groups.

5 CONCLUSION

This work enhances multi-task real-world image super-resolution by refining task-space modeling and task imbalance management. We propose a degradation operator-aware task definition framework that segments the degradation space with parameter-specific boundaries, balancing task discrimination and efficiency. A focal loss-based weighting mechanism dynamically quantifies task imbalance, while a task-to-data imbalance conversion strategy stabilizes optimization by regulating task-specific training volumes. Experiments demonstrate consistent superiority across various degradation scenarios, with seamless integration into existing architectures. A potential limitation of our method is that the training cost increases proportionally with the number of defined tasks, this inherently restricts the extent to which tasks can be subdivided.

REFERENCES

Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pp. 126–135, 2017.

Marta Bistroń and Zbigniew Piotrowski. Optimization of imaging reconnaissance systems using super-resolution: Efficiency analysis in interference conditions. *Sensors (Basel, Switzerland)*, 24 (24):7977, 2024.

Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 3086–3095, 2019.

Bin Chen, Gehui Li, Rongyuan Wu, Xindong Zhang, Jie Chen, Jian Zhang, and Lei Zhang. Adversarial diffusion compression for real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2025.

Du Chen, Jie Liang, Xindong Zhang, Ming Liu, Hui Zeng, and Lei Zhang. Human guided ground-truth generation for realistic image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 14082–14091, 2023a.

Xiangyu Chen, Xintao Wang, Wenlong Zhang, Xiangtao Kong, Yu Qiao, Jiantao Zhou, and Chao Dong. Hat: Hybrid attention transformer for image restoration. *arXiv preprint arXiv:2309.05239*, 2023b.

Sumanth Chennupati, Ganesh Sistu, Senthil Yogamani, and Samir A Rawashdeh. Multinet++: Multi-stream feature aggregation and geometric loss strategy for multi-task learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition workshops*, pp. 0–0, 2019.

Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pp. 184–199. Springer, 2014.

Linwei Dong, Qingnan Fan, Yihong Guo, Zhonghao Wang, Qi Zhang, Jinwei Chen, Yawei Luo, and Changqing Zou. Tsd-sr: One-step diffusion with target score distillation for real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2025.

Yuanting Fan, Chengxu Liu, Nengzhong Yin, Changlong Gao, and Xueming Qian. Adadiffsr: Adaptive region-aware dynamic acceleration diffusion model for real-world image super-resolution. In *European Conference on Computer Vision*, pp. 396–413. Springer, 2024.

Chris Fifty, Ehsan Amid, Zhe Zhao, Tianhe Yu, Rohan Anil, and Chelsea Finn. Efficiently identifying task groupings for multi-task learning. *Advances in Neural Information Processing Systems*, 34: 27503–27516, 2021.

Emilie Grégoire, Muhammad Hafeez Chaudhary, and Sam Verboven. Sample-level weighting for multi-task learning with auxiliary tasks. *Applied Intelligence*, 54(4):3482–3501, 2024.

Jinjin Gu, Hannan Lu, Wangmeng Zuo, and Chao Dong. Blind super-resolution with iterative kernel correction. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1604–1613, 2019.

Michelle Guo, Albert Haque, De-An Huang, Serena Yeung, and Li Fei-Fei. Dynamic task prioritization for multitask learning. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 270–287, 2018.

Yan Huang, Shang Li, Liang Wang, Tieniu Tan, et al. Unfolding the alternating optimization for blind super resolution. *Advances in Neural Information Processing Systems*, 33:5632–5643, 2020.

Alex Kendall, Yarin Gal, and Roberto Cipolla. Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 7482–7491, 2018.

Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 1646–1654, 2016.

Iasonas Kokkinos. Ubernet: Training a universal convolutional neural network for low-, mid-, and high-level vision using diverse datasets and limited memory. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 6129–6138, 2017.

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 *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 4681–4690, 2017.

Jie Liang, Hui Zeng, and Lei Zhang. Efficient and degradation-adaptive network for real-world image super-resolution. In *European Conference on Computer Vision*, pp. 574–591. Springer, 2022.

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

Lukas Liebel and Marco Körner. Single-image super resolution for multispectral remote sensing data using convolutional neural networks. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, 41:883–890, 2016.

Baijiong Lin, Feiyang Ye, Yu Zhang, and Ivor W Tsang. Reasonable effectiveness of random weighting: A litmus test for multi-task learning. *arXiv preprint arXiv:2111.10603*, 2021.

Shikun Liu, Edward Johns, and Andrew J Davison. End-to-end multi-task learning with attention. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 1871–1880, 2019.

Ishan Misra, Abhinav Shrivastava, Abhinav Gupta, and Martial Hebert. Cross-stitch networks for multi-task learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3994–4003, 2016.

Chong Mou, Yanze Wu, Xintao Wang, Chao Dong, Jian Zhang, and Ying Shan. Metric learning based interactive modulation for real-world super-resolution. In *European Conference on Computer Vision*, pp. 723–740. Springer, 2022.

Robin Rombach, Andreas Blattmann, Dominik Lorenz, Patrick Esser, and Björn Ommer. High-resolution image synthesis with latent diffusion models. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10684–10695, 2022.

Liangtian Wan, Yuchen Sun, Lu Sun, Zhaolong Ning, and Joel JPC Rodrigues. Deep learning based autonomous vehicle super resolution doa estimation for safety driving. *IEEE Transactions on Intelligent Transportation Systems*, 22(7):4301–4315, 2020.

Boyang Wang, Bowen Liu, Shiyu Liu, and Fengyu Yang. Vcisr: Blind single image super-resolution with video compression synthetic data. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 4302–4312, 2024a.

Longguang Wang, Yingqian Wang, Xiaoyu Dong, Qingyu Xu, Jungang Yang, Wei An, and Yulan Guo. Unsupervised degradation representation learning for blind super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 10581–10590, 2021a.

Xintao Wang, Ke Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 606–615, 2018.

Xintao Wang, Liangbin Xie, Chao Dong, and Ying Shan. Real-esrgan: Training real-world blind super-resolution with pure synthetic data. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 1905–1914, 2021b.

Yufei Wang, Wenhan Yang, Xinyuan Chen, Yaohui Wang, Lanqing Guo, Lap-Pui Chau, Ziwei Liu, Yu Qiao, Alex C Kot, and Bihan Wen. Sinsr: diffusion-based image super-resolution in a single step. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25796–25805, 2024b.

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 *Computer Vision–ECCV 2020: 16th European Conference, Glasgow, UK, August 23–28, 2020, Proceedings, Part VIII 16*, 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, 2024a.

Rongyuan Wu, Tao Yang, Lingchen Sun, Zhengqiang Zhang, Shuai Li, and Lei Zhang. Seesr: Towards semantics-aware real-world image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 25456–25467, 2024b.

Tianhe Yu, Saurabh Kumar, Abhishek Gupta, Sergey Levine, Karol Hausman, and Chelsea Finn. Gradient surgery for multi-task learning. *Advances in neural information processing systems*, 33: 5824–5836, 2020.

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.

Hayoung Yun and Hanjoo Cho. Achievement-based training progress balancing for multi-task learning. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pp. 16935–16944, 2023.

Amir R Zamir, Alexander Sax, William Shen, Leonidas J Guibas, Jitendra Malik, and Silvio Savarese. Taskonomy: Disentangling task transfer learning. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3712–3722, 2018.

Kai Zhang, Wangmeng Zuo, and Lei Zhang. Learning a single convolutional super-resolution network for multiple degradations. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 3262–3271, 2018a.

Kai Zhang, Jingyun Liang, Luc Van Gool, and Radu Timofte. Designing a practical degradation model for deep blind image super-resolution. In *Proceedings of the IEEE/CVF international conference on computer vision*, pp. 4791–4800, 2021.

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

Wenlong Zhang, Xiaohui Li, Guangyuan Shi, Xiangyu Chen, Yu Qiao, Xiaoyun Zhang, Xiao-Ming Wu, and Chao Dong. Real-world image super-resolution as multi-task learning. *Advances in Neural Information Processing Systems*, 36:21003–21022, 2023.

Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European conference on computer vision (ECCV)*, pp. 286–301, 2018c.

Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 2472–2481, 2018d.

Yongpei Zhu, Zicong Zhou, Guojun Liao, and Kehong Yuan. Csrgan: medical image super-resolution using a generative adversarial network. In *2020 IEEE 17th international symposium on biomedical imaging workshops (ISBI Workshops)*, pp. 1–4. IEEE, 2020.

A APPENDIX

A.1 DETAILED DEGRADATION SETTINGS AND ANALYSIS OF TASK BOUNDARIES

A.1.1 DETAILED PARAMETER BOUNDARIES FOR TASK DEFINITION.

The detailed parameter settings for our degradation task are reported in Tab. 6 7 8 9. We partition the entire degradation space $\mathcal{D}$ into 4 distinct levels: mild, blur, noise, and severe.

For the blur operator, we employ isotropic and anisotropic blur kernels with probabilities of 0.69 and 0.31, respectively. When an isotropic blur kernel is applied, its standard deviations along both axes are set equal. In the second degradation stage, following the practice of Real-ESRGAN Wang et al. (2021b), we skip the blur operation with a probability of 0.2 and apply sinc kernel filtering with a probability of 0.8.

The primary distinctions in our configuration pertain to the blur kernel standard deviations and noise levels. We highlight lower values for blur kernel standard deviations and noise levels in blue, and higher values for these parameters in red.

Table 6: Degradation Operator Parameters for Task "mild".

Operator	Parameter	Stage 1	Stage 2
Blur	Kernel size $2m+1$	$m \in [3, 11]$	$m \in [3, 11]$
	Kernel list	iso, aniso	iso, aniso
	Kernel list probability	$[0.69, 0.31]$	$[0.69, 0.31]$
	Sinc kernel probability	0.1	0.1
	Standard deviation θ	$\theta \in [1.6, 2.3]$	$\theta \in [0.8, 1.2]$
Resize	Resize list	down, same, up	down, same, up
	Resize list probability	$[0.2, 0.7, 0.1]$	$[0.2, 0.7, 0.1]$
	Resize range ϕ	$\phi \in [0.15, 1.5]$	$\phi \in [0.15, 1.5]$
	Resize mode	area, bilinear, bicubic	area, bilinear, bicubic
Noise	Noise list	Gaussian, Poisson	Gaussian, Poisson
	Noise list probability	$[0.5, 0.5]$	$[0.5, 0.5]$
	Sigma of Gaussian σ	$\sigma \in [1, 4.5]$	$\sigma \in [1, 4]$
	Scale of Poisson γ	$\gamma \in [0.1, 0.4]$	$\gamma \in [0.1, 0.35]$
JPEG	Quality factor α	$\alpha \in [30, 95]$	$\alpha \in [30, 95]$

Table 7: Degradation Operator Parameters for Task "blur".

Operator	Parameter	Stage 1	Stage 2
Blur	Kernel size $2m+1$	$m \in [3, 11]$	$m \in [3, 11]$
	Kernel list	iso, aniso	iso, aniso
	Kernel list probability	$[0.69, 0.31]$	$[0.69, 0.31]$
	Sinc kernel probability	0.1	0.1
	Standard deviation θ	$\theta \in [2.3, 3]$	$\theta \in [1.2, 1.6]$
Resize	Resize list	down, same, up	down, same, up
	Resize list probability	$[0.2, 0.7, 0.1]$	$[0.2, 0.7, 0.1]$
	Resize range ϕ	$\phi \in [0.15, 1.5]$	$\phi \in [0.15, 1.5]$
	Resize mode	area, bilinear, bicubic	area, bilinear, bicubic
Noise	Noise list	Gaussian, Poisson	Gaussian, Poisson
	Noise list probability	$[0.5, 0.5]$	$[0.5, 0.5]$
	Sigma of Gaussian σ	$\sigma \in [1, 4.5]$	$\sigma \in [1, 4]$
	Scale of Poisson γ	$\gamma \in [0.1, 0.4]$	$\gamma \in [0.1, 0.35]$
JPEG	Quality factor α	$\alpha \in [30, 95]$	$\alpha \in [30, 95]$

Table 8: Degradation Operator Parameters for Task "noise".

Operator	Parameter	Stage 1	Stage 2
Blur	Kernel size $2m+1$	$m \in [3, 11]$	$m \in [3, 11]$
	Kernel list	iso, aniso	iso, aniso
	Kernel list probability	$[0.69, 0.31]$	$[0.69, 0.31]$
	Sinc kernel probability	0.1	0.1
	Standard deviation θ	$\theta \in [1.6, 2.3]$	$\theta \in [0.8, 1.2]$
Resize	Resize list	down, same, up	down, same, up
	Resize list probability	$[0.2, 0.7, 0.1]$	$[0.2, 0.7, 0.1]$
	Resize range ϕ	$\phi \in [0.15, 1.5]$	$\phi \in [0.15, 1.5]$
	Resize mode	area, bilinear, bicubic	area, bilinear, bicubic
Noise	Noise list	Gaussian, Poisson	Gaussian, Poisson
	Noise list probability	$[0.5, 0.5]$	$[0.5, 0.5]$
	Sigma of Gaussian σ	$\sigma \in [4.5, 8]$	$\sigma \in [4, 7]$
	Scale of Poisson γ	$\gamma \in [0.4, 0.7]$	$\gamma \in [0.35, 0.6]$
JPEG	Quality factor α	$\alpha \in [30, 95]$	$\alpha \in [30, 95]$

Table 9: Degradation Operator Parameters for Task "severe".

Operator	Parameter	Stage 1	Stage 2
Blur	Kernel size $2m+1$	$m \in [3, 11]$	$m \in [3, 11]$
	Kernel list	iso, aniso	iso, aniso
	Kernel list probability	$[0.69, 0.31]$	$[0.69, 0.31]$
	Sinc kernel probability	0.1	0.1
	Standard deviation θ	$\theta \in [2.3, 3]$	$\theta \in [1.2, 1.6]$
Resize	Resize list	down, same, up	down, same, up
	Resize list probability	$[0.2, 0.7, 0.1]$	$[0.2, 0.7, 0.1]$
	Resize range ϕ	$\phi \in [0.15, 1.5]$	$\phi \in [0.15, 1.5]$
	Resize mode	area, bilinear, bicubic	area, bilinear, bicubic
Noise	Noise list	Gaussian, Poisson	Gaussian, Poisson
	Noise list probability	$[0.5, 0.5]$	$[0.5, 0.5]$
	Sigma of Gaussian σ	$\sigma \in [4.5, 8]$	$\sigma \in [4, 7]$
	Scale of Poisson γ	$\gamma \in [0.4, 0.7]$	$\gamma \in [0.35, 0.6]$
JPEG	Quality factor α	$\alpha \in [30, 95]$	$\alpha \in [30, 95]$

A.1.2 ANALYSIS OF PARTITIONING STRATEGY

In our multi-task SR framework, the choice of a partitioning strategy for the task subspaces should follow the principle of maximum task dissimilarity. This involves defining tasks that are as distinct as possible to maximize the efficacy of our algorithms in mitigating task imbalance. From an empirical standpoint, since our degradation process involves uniformly sampling parameters within a specified range, adopting a uniform partition of the degradation space is the most natural and unbiased strategy. This approach ensures that each task covers a parametric region of equal size, thus maintaining a high degree of differentiability between tasks.

To validate the reasonableness of this uniform partitioning, we conducted an additional validation experiment. The overall protocol is largely identical to the experiment shown in Fig.1(b) and (c). The key distinction is that when selecting values for the noise (and blur) degradation operator, we uniformly sample 30 discrete values from within our defined degradation subspaces. The parameters for the second-stage degradation are chosen proportionally to those of the first stage. We then calculate the PSNR distances between the shared model and the single-task networks for each of these 30 degradation settings. This validation was performed on 6 model checkpoints saved randomly throughout the training process. The results are illustrated in the Fig.8.

The experimental results shown in Fig.8 indicate that for the task clusters generated by uniformly sampling the noise (and blur) operators within our defined ranges, the degree of task imbalance exhibits a roughly linear trend. The curve of the performance gap remains relatively stable and shows no pronounced non-linearities or inflection points. This finding validates the reasonableness of our uniform partitioning strategy, as it confirms that this approach provides substantial and consistent differentiability among the tasks across the predefined degradation space.

Furthermore, we performed a follow-up experiment to determine the effective range of this linearity by expanding the overall degradation parameter space. As illustrated in Fig.9, the results indicate that within the RealESRGAN degradation model, the linear trend in task imbalance is largely preserved for blur standard deviation values within the range of [0.4, 3.8], while the corresponding linear range for the noise ratio is [1, 19].

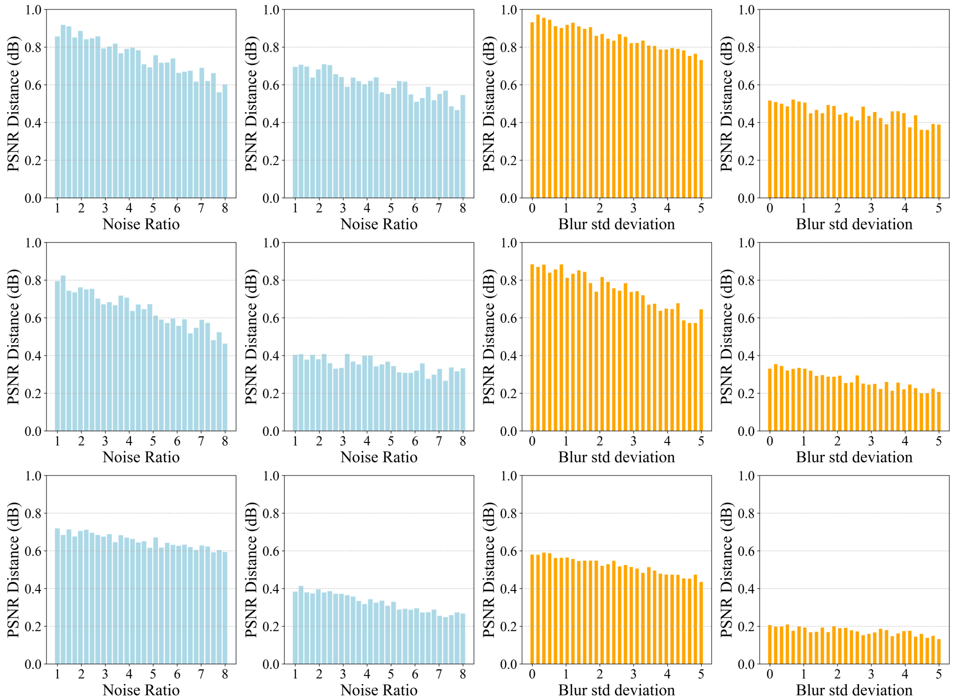


Figure 8: Validation of task imbalance at six random training checkpoints. The plots show the PSNR distance between the shared and single-task models across 30 uniformly sampled Noise ratio (and blur standard deviation) values. A consistently linear trend is observed at all stages.

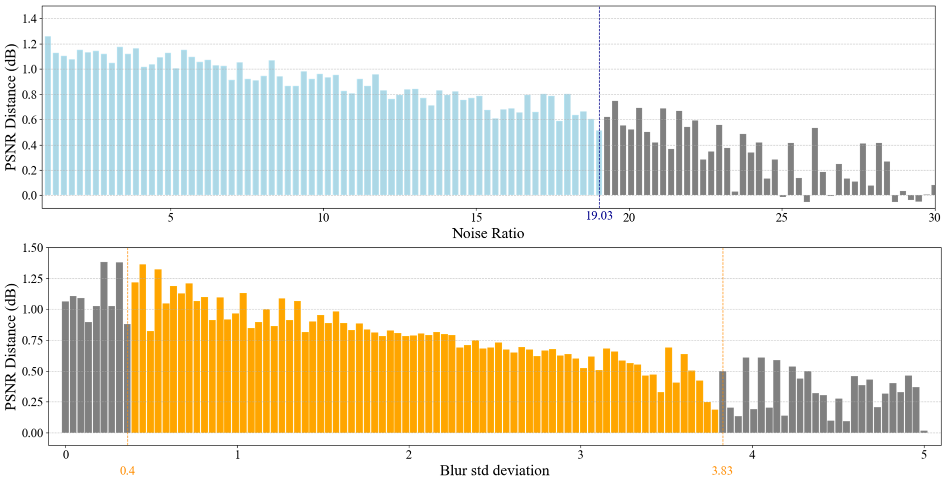


Figure 9: Determining the effective range of linearity for the noise operator. The blue and orange bars show the region where the task imbalance (PSNR distance) maintains a linear trend. The gray bars show the post-inflection region where this linearity breaks down. Dashed lines mark the approximate boundaries of the effective range.

A.2 THE DETAILED DERIVATION PROCESS

A.2.1 EQUIVALENCE OF LOSS RE-WEIGHTING AND DATA REBALANCING

We begin from Equ. 4 and reformulate the multi-task loss function into a sample re-weighting formulation. Given a set of degradation tasks $\mathcal{T} = \{\tau_i = (X_i, Y_i)\}_{i=1}^{n}$, each task corresponds to a degradation subspace $\mathcal{D}_{\text{sub}} = \{\mathcal{D}_1, \mathcal{D}_2, \ldots, \mathcal{D}_n\}$. Within a specific interval T_k (determined by a fixed number of iterations), the total number of training samples is N. We first assume that all tasks share an identical sample size, i.e., $N_1 = N_2 \ldots = N_n = N/n$. The loss function for task τ_i is defined as follows:

$$\mathcal{L}_{\tau_i} = \frac{1}{N_i} \sum_{(x_i, y_i) \in \tau_i} \ell(x_i, y_i) \tag{5}$$

where $\ell(x_i, y_i)$ denotes the single sample loss. Substituting this into the multi-task loss function Equ. 4 yields:

$$\mathcal{L} = \sum_{i=1}^{n} w_{\tau_i}^{k} \cdot \frac{1}{N_i} \sum_{(x_i, y_i) \in \tau_i} \ell(x_i, y_i) \tag{6}$$

By interchanging the order of summation, we derive a unified loss summation over all samples:

$$\mathcal{L} = \sum_{i=1}^{n} \sum_{(x_i, y_i) \in \tau_i} \frac{w_{\tau_i}^{k}}{N_i} \ell(x_i, y_i) \tag{7}$$

We define the sample weight $w(x_i, y_i)$ as:

$$w(x_i, y_i) = \frac{w_{\tau_i}^{k}}{N_i}, \quad \text{for } (x_i, y_i) \in \tau_i \tag{8}$$

Thus, the loss function can be reformulated as:

$$\mathcal{L} = \sum_{(x,y) \in \mathcal{T}} w(x, y) \ell(x, y) \tag{9}$$

Through this formulation, we convert the multi-task weighted loss into sample re-weighting, where samples from task τ_i are assigned weight of $w_{\tau_i}^{k}/N_i$.

Next, we enforce equal sample weights across all tasks, while adjusting the sample size of each task to preserve the task imbalance quantification results from Section 3.2. Specifically, we set the sample weight $w(x_i, y_i)$ to a uniform value:

$$\forall \tau_i \in \mathcal{T}, \quad w\left(x_i, y_i\right) = \frac{w_{\tau_i}^{k}}{N_i} = \alpha \tag{10}$$

We substitute $\sum_{i=1}^{n} N_i = N$ and $\sum_{i=1}^{n} w_{\tau_i}^{k} = 1$ into the above equation and solve for α:

$$\alpha = \frac{\sum_{i=1}^{n} w_{\tau_i}^{k}}{N} = \frac{1}{N} \tag{11}$$

Consequently, the adjusted sample size for each task in interval T_k is:

$$N_i = \frac{w_{\tau_i}^{k}}{\alpha} = N \cdot w_{\tau_i}^{k} \tag{12}$$

Through the above transformations, we equivalently convert the quantified task imbalance weights into controllable data rebalancing.

A.2.2 A VARIANCE REDUCTION ANALYSIS FROM LOSS REWEIGHTING TO DATA REBALANCING

To avoid confusion, and with a slight abuse of notation, we redefine the symbols as follows: we consider a multi-task setup with n tasks and normalized task weights $w_1, \ldots, w_n$ satisfying $\sum_{j=1}^{n} w_j = 1$. For task j let G_j denote the scalar random variable given by the (projected) single-sample stochastic gradient coming from a sample of task j. Define the conditional first and second moments

$$\mu_j := \mathbb{E}[G_j], \qquad E_j := \mathbb{E}[G_j^2] = \mathrm{Var}(G_j) + \mu_j^2.$$

The population target gradient component we wish to estimate is

$$G := \mathbb{E}_{j \sim q}[\mu_j] = \sum_{j=1}^{n} w_j \mu_j,$$

where q denotes the desired (normalized) task distribution $q_j = w_j$. Below we compare two single-sample estimators of G used in practice.

(1) Loss reweighting (importance-weighting baseline). Sample a task index J uniformly, $J \sim \mathrm{Unif}\{1, \dots, n\}$ (so $p_j^{\mathrm{rw}} = 1/n$). For a draw from task J use the importance-weighted estimator

$$Z_{\mathrm{rw}} \;=\; \frac{q_J}{p_J}\, G_J \;=\; n w_J\, G_J .$$

(2) Data rebalancing / resampling. Sample a task index J directly from q (so $p_j^{\mathrm{rs}} = w_j$) and use the direct estimator

$$Z_{\mathrm{rs}} \;=\; G_J .$$

Both estimators are unbiased for G:

$$\mathbb{E}[Z_{\mathrm{rw}}] = \sum_{j=1}^{n} \frac{1}{n}\,(n w_j)\,\mu_j = \sum_{j=1}^{n} w_j \mu_j = G, \qquad \mathbb{E}[Z_{\mathrm{rs}}] = \sum_{j=1}^{n} w_j \mu_j = G.$$

We compare their variances. Using $\mathrm{Var}(Z) = \mathbb{E}[Z^2] - \mathbb{E}[Z]^2$ and denoting $\mathbb{E}[G_j^2] = E_j$, the second moments are

$$\mathbb{E}[Z_{\mathrm{rw}}^2] = \sum_{j=1}^{n} \frac{1}{n}\,(n w_j)^2\, E_j = n \sum_{j=1}^{n} w_j^2 E_j, \qquad \mathbb{E}[Z_{\mathrm{rs}}^2] = \sum_{j=1}^{n} w_j E_j .$$

Hence the difference of single-sample variances (reweighting minus resampling) is

$$\begin{aligned}
\Delta := \mathrm{Var}(Z_{\mathrm{rw}}) - \mathrm{Var}(Z_{\mathrm{rs}}) &= \big(\mathbb{E}[Z_{\mathrm{rw}}^2] - G^2\big) - \big(\mathbb{E}[Z_{\mathrm{rs}}^2] - G^2\big) \\
&= \mathbb{E}[Z_{\mathrm{rw}}^2] - \mathbb{E}[Z_{\mathrm{rs}}^2] = n \sum_{j=1}^{n} w_j^2 E_j - \sum_{j=1}^{n} w_j E_j \\
&= \sum_{j=1}^{n} \big(n w_j^2 - w_j\big) E_j = \sum_{j=1}^{n} w_j \big(n w_j - 1\big)\, E_j .
\end{aligned}$$

Therefore, defining the per-task contribution

$$\Delta_j := w_j \big(n w_j - 1\big)\, E_j ,$$

we obtain the exact decomposition

$$\boxed{\Delta \;=\; \sum_{j=1}^{n} \Delta_j \;=\; \sum_{j=1}^{n} w_j (n w_j - 1)\, \mathbb{E}[G_j^2].}$$

We continue from the decomposition

$$\Delta \;=\; \mathrm{Var}_{\mathrm{rw}} - \mathrm{Var}_{\mathrm{rs}} \;=\; \sum_{j=1}^{n} \Delta_j \;=\; \sum_{j=1}^{n} w_j (n w_j - 1)\, E_j , \qquad E_j := \mathbb{E}[G_j^2],$$

which exactly expresses the single-sample variance gap between the *pure loss-reweighting* scheme (uniform task sampling with importance weights, denoted “rw”) and the *pure data-rebalancing / resampling* scheme (sampling tasks directly from q, denoted “rs”) as a sum of task-wise contributions Δ_j.

Now partition the task index set into two disjoint subsets according to the sign of the factor $n w_j - 1$:

$$H \;:=\; \{\, j : \; n w_j > 1 \,\}, \qquad L \;:=\; \{\, j : \; n w_j < 1 \,\},$$

(we may treat indices with $n w_j = 1$ as boundary cases for which $\Delta_j = 0$). Then the global decomposition separates into the two groups:

$$\Delta \;=\; \sum_{h \in H} \Delta_h \;+\; \sum_{j \in L} \Delta_j \;=\; \sum_{h \in H} w_h (n w_h - 1) E_h \;+\; \sum_{j \in L} w_j (n w_j - 1) E_j .$$

The above separation suggests a *hybrid* (mixed) strategy: treat the two groups differently—

- for tasks in H (those with $nw_h > 1$) adopt *data rebalancing / resampling* (i.e. draw task indices J from q for these tasks), and
- for tasks in L (those with $nw_j < 1$) retain *loss reweighting* (i.e. continue to sample uniformly among those tasks and apply importance weights).

We now show rigorously that, compared to the *pure reweighting* baseline, this hybrid scheme eliminates the negative contribution from the L-group in the variance-difference expression and leaves only the strictly positive contributions from the H-group; consequently the hybrid variance is strictly smaller than the pure-reweighting variance whenever at least one $h \in H$ has $E_h > 0$.

Formal proof. Let $\mathrm{Var}_{\mathrm{rw}}$ denote the single-sample variance under the pure reweighting baseline (uniform sampling $p_j^{\mathrm{rw}} = 1/n$ with estimator $Z_{\mathrm{rw}} = nw_J G_J$), and let $\mathrm{Var}_{\mathrm{hyb}}$ denote the single-sample variance under the hybrid scheme defined above. We compute $\mathrm{Var}_{\mathrm{hyb}}$ by assembling the contribution of each task under the hybrid sampling/weighting rule.

- For $h \in H$: under the hybrid scheme we *resample* from H according to q for those tasks, so the marginal contribution to the second moment is $w_h E_h$ (exactly as in the pure-resampling case).
- For $j \in L$: under the hybrid scheme we *retain reweighting* for those tasks (i.e. keep uniform sampling and apply the same importance weights as in the baseline), hence the marginal contribution to the second moment from each such task is exactly the same as in the baseline, namely $nw_j^2 E_j$.

Therefore the hybrid second moment equals

$$\mathbb{E}[Z_{\mathrm{hyb}}^2] = \sum_{h \in H} w_h E_h \; + \sum_{j \in L} nw_j^2 E_j,$$

while the baseline second moment is

$$\mathbb{E}[Z_{\mathrm{rw}}^2] = \sum_{h \in H} nw_h^2 E_h \; + \sum_{j \in L} nw_j^2 E_j.$$

Subtracting (and cancelling the identical L-group terms) gives the exact difference of second moments (and therefore variances, since means coincide and cancel):

$$\mathrm{Var}_{\mathrm{rw}} - \mathrm{Var}_{\mathrm{hyb}} \; = \sum_{h \in H} \big(nw_h^2 - w_h\big) E_h \; = \sum_{h \in H} w_h (nw_h - 1) E_h.$$

But each summand satisfies $nw_h - 1 > 0$ by definition of H, and $E_h = \mathbb{E}[G_h^2] \geq 0$ with strict positivity whenever the task's per-sample gradient second moment is nonzero. Hence every $\Delta_h := w_h(nw_h - 1)E_h$ is nonnegative, and at least one is strictly positive whenever there exists $h \in H$ with $E_h > 0$. Consequently

$$\boxed{\mathrm{Var}_{\mathrm{rw}} - \mathrm{Var}_{\mathrm{hyb}} \; = \sum_{h \in H} \Delta_h \; \geq \; 0,}$$

with strict inequality whenever $\exists\, h \in H : \; E_h > 0$. Equivalently,

$$\mathrm{Var}_{\mathrm{hyb}} < \mathrm{Var}_{\mathrm{rw}} \quad \text{(when some } h \in H \text{ has } E_h > 0\text{)}.$$

Interpretation and practical remark. The algebra above shows the precise mechanism: because the hybrid scheme uses the same reweighting procedure as the baseline on the small-weight group L, the contributions of those tasks to the variance are identical and thus cancel in the variance difference; all remaining contributions come from the large-weight group H, for which replacing reweighting by resampling reduces each per-task term from $nw_h^2 E_h$ down to $w_h E_h$, yielding a positive reduction $w_h(nw_h - 1)E_h$ for every $h \in H$, this mixed strategy therefore provably reduces the stochastic gradient variance relative to pure loss reweighting, and so yields a more stable optimization procedure.

A.3 MORE VISUAL RESULTS.

A.3.1 MORE QUALITATIVE COMPARISONS ON DIV2K4LEVEL.

We present additional qualitative results on DIV2K4Level in Fig. 10. Our method demonstrates the ability to recover substantial clarity even from highly degraded low-resolution inputs.

A.3.2 QUALITATIVE COMPARISONS WITH DIFFUSION-BASED MODELS.

Recently, diffusion-based models have demonstrated considerable potential in real-world super-resolution (Real-SR). These models typically employ or fine-tune large pre-trained diffusion models, such as Stable Diffusion Rombach et al. (2022), to generate visually pleasing HR images. However, However, the large number of parameters in these pre-trained models can often lead to outputs that lack fidelity to the ground truth. We illustrate this phenomenon in Fig. 11: although diffusion-based models may achieve high scores on non-reference metrics, indicative of perceptual quality, they are prone to substantially altering the content of the LR input, resulting in significant deviations from the ground truth.

A.4 OTHER IMPLEMENTATION DETAILS.

To establish task difficulty, single-task networks were trained by fine-tuning a pre-trained RealESRNet model on each of the four degradation subspaces for 400K iterations. The entire training process was divided into intervals, with checkpoints set every 5K iterations. At each checkpoint, multi-task weights were calculated based on the PSNR distances between the single-task models (trained up to that checkpoint) and the current shared model. Finally, the training data volume for each task within an interval was dynamically allocated according to the normalized weights derived from the quantified task imbalance. We train our models with 8 NVIDIA GeForce RTX 2080 Ti GPUs with a total batch size of 48. We employ Adam optimizer.

A.5 THE USE OF LARGE LANGUAGE MODELS

In the preparation of this manuscript, we utilized Large Language Models as a writing assistant. The primary use of the LLM was for proofreading, including grammatical corrections, and rephrasing of sentences to improve clarity and readability. The core ideas, experimental design, results, and conclusions presented in this paper are entirely our own.

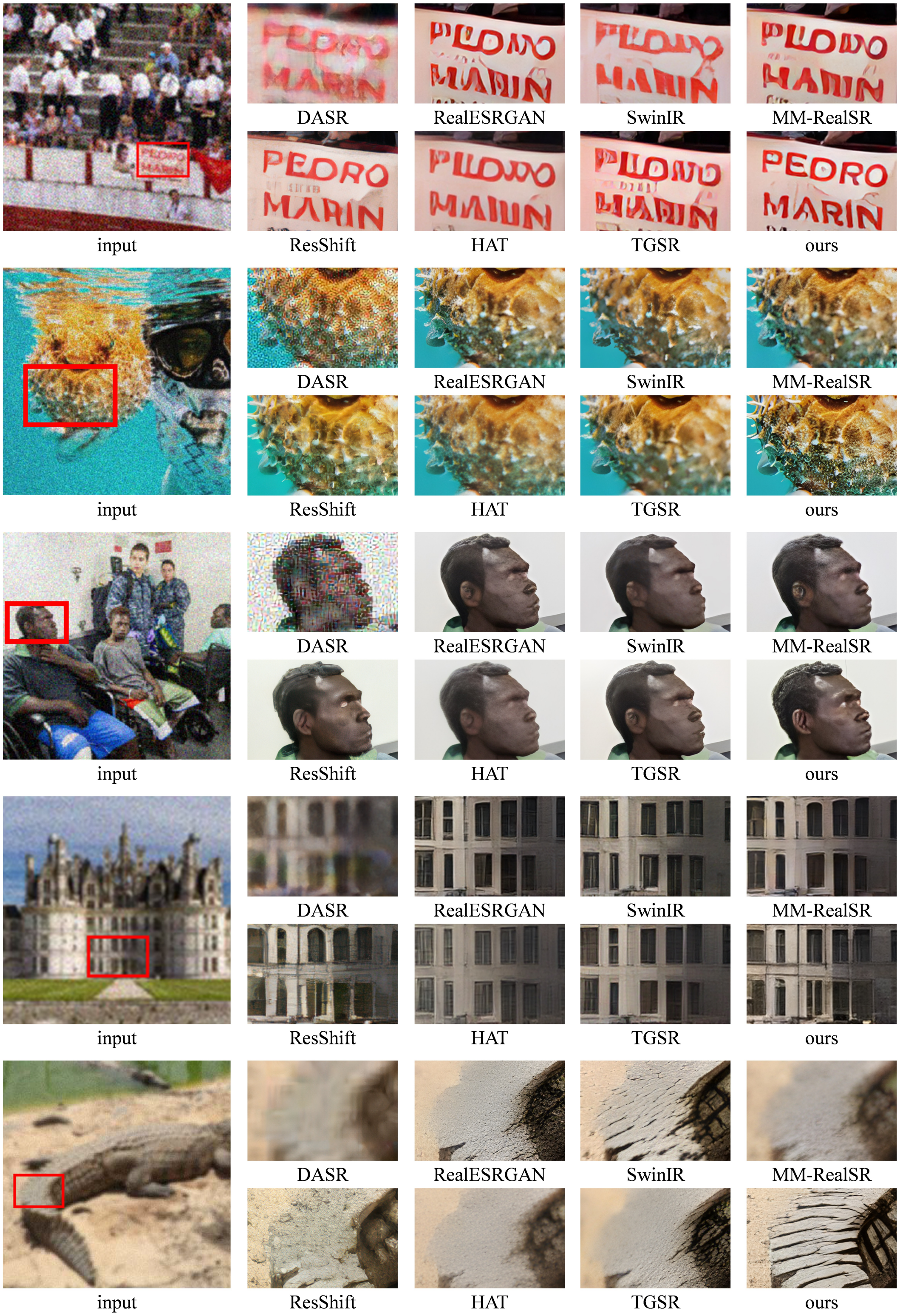


Figure 10: More Qualitative Comparisons on DIV2K4Level.

Figure 11: Qualitative Comparisons with Diffusion-based Models.