Rethinking Imbalance in Image Super-Resolution for Efficient Inference

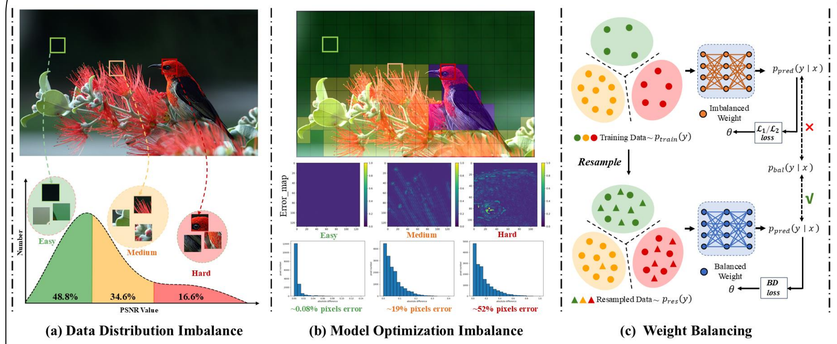
Wei Yu[1], Bowen Yang[1], Qinglin Liu[1], Jianing Li[2], Shengping Zhang[1], Xiangyang Ji[2]

[1]Harbin Institute of Technology, [2]Tsinghua University

Contact: Wei Yu (yuwei00828@gmail.com)

NEURAL INFORMATION PROCESSING SYSTEMS

Motivation



(a) Data Distribution Imbalance (b) Model Optimization Imbalance (c) Weight Balancing

Figure 1. Illustration of (a) the data distribution from the DIV2K training set, (b) the reconstruction results of RCAN, and (c) the proposed weight-balancing framework.

Limitation:

- **Distribution Imbalance:** Existing SR methods mostly use uniformly sampled LR-HR patch pairs and ignore the underlying fact that patch contents in images exhibit imbalanced distributions (i.e., the abundant easily reconstructed smooth flat patches (48.8 %) and rare hardly reconstructed edge texture patches (16.6 %), as shown in Figure 1 (a).
- **Model Optimization Imbalance:** Existing SR algorithms typically employ L1 or L2 losses to treat all patch areas and optimize each weight equally, which involve redundant calculations in flat areas, which leads to imbalanced inference performance where the model overfits in simple areas and underfits in complex ones and results in uneven distribution of model computational resources, as shown in Figure 1 (b).

Theoretical Analysis:

Let x and y denote LR and HR patches, the prediction $\hat{y} = f_\theta(x)$ from the SR network can be modeled as a Gaussian distribution

$$p(y|x;\theta) = \mathcal{N}(y;\hat{y},\sigma^2_{\text{noise}}\text{I})$$

The prediction $\hat{y}$ can be treated as the mean of a noisy prediction distribution.

- **Theorem: Distribution Transformation**

Given identical probability $p(y|x)$ across both train $p_{\text{train}}(y)$ and test $p_{\text{bal}}(y)$ sets

$$p_{\text{train}}(y|x) = p_{\text{train}}(y)\cdot\frac{p_{\text{bal}}(y|x)}{p_{\text{bal}}(y)}\cdot\frac{p_{\text{bal}}(x)}{p_{\text{train}}(x)}$$

This theorem reveals that the existence of imbalance issues stems from the direct proportionality between $p_{\text{train}}(y|x)$ and $p_{\text{train}}(y)$ with a ratio of $\frac{p_{\text{bal}}(x)}{p_{\text{train}}(x)}$.

Contributions:

- This paper is the first attempt to explore the imbalance in the image super-resolution field and gives a reasonable analysis from a perspective of probability statistics, i.e., the imbalance of data distribution and model optimization limits the algorithm performance.
- We propose a plug-and-play weight-balancing framework dubbed WBSR upon HES and BDLoss to achieve balance training without additional computation costs, which improves the restoration effect and inference efficiency of models without changing the original model structure and training data.

Methodology

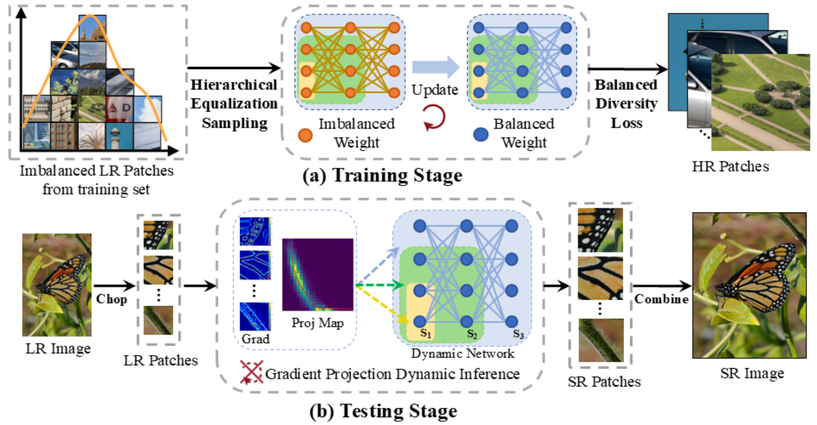


Figure 2. Illustration of the proposed weight-balancing framework WBSR. (a) The training stage combines hierarchical equalization sampling and balanced diversity loss to jointly train a supernet model with balanced weights. (b) The testing stage adopts the gradient projection dynamic inference with a gradient projection map and multiple dynamic subnets for efficient inference.

Weight-Balancing Training Framework

We consider attaining a robust model representation with balanced weights from the perspective of two aspects: data sampling and optimization function. Figure 2 (a) illustrates the training process of our WBSR consisting of Hierarchical Equalization Sampling (HES) and Balanced Diversity Loss $\mathcal{L}_{bd}$, the optimization objective is

$$\min_\theta \mathbb{E}_{(x,y)|p_{train}}\mathcal{L}_{bd}(y - \mathcal{S}_{m_\theta}(x))$$

Each subnet $\mathcal{S}_{m_\theta}$ with varying computational cost shares the weights of the supernet and is intended to handle image patches of different complexities.

Hierarchical Equalization Sampling

Sample-Level Sampling uniformly samples patches from the training dataset with equal probability, which ensures that the model learns stable initial weights early in training.

Class-Level Sampling aims to assign a higher sampling probability to rare difficult samples. The threshold for the k_th class and the sampling possibility P_k can be calculated as follows

$$t_k = t[\tfrac{k\cdot N}{K}], k\in[1,K]\ ,\ P_k = \frac{\sum_{j=1}^{K}\frac{1}{N_j}}{N_k\cdot\delta^k}$$

Balanced Diversity Loss

To balance the uncertainty of model diversity predictions and avoid excessive optimization, our BDLoss is defined as the likelihood function

$$\log p_{\text{train}}(y \mid x;\theta) = \log\frac{p_{\text{bal}}(y|x;\theta)\cdot p_{\text{train}}(y)}{\int_Y p_{\text{bal}}(y'|x;\theta)\cdot p_{\text{train}}(y')dy'}$$

$$= \log\mathcal{N}\left(y;\hat{y},\sigma^2_{\text{noise}}\,\text{I}\right) + \log p_{\text{train}}(y) - \log\int_Y \mathcal{N}\left(y';\hat{y},\sigma^2_{\text{noise}}\,\mathbf{I}\right)\cdot p_{\text{train}}(y')dy'$$

Gradient Projection Dynamic Inference

Figure 2 (b) illustrates the testing process of our WBSR. Gradient Projection calculates gradient vectors to measure the complexity of the patch contents and constructs a gradient projection map online to project the gradient vector of an image patch to the selection of each subset model. Dynamic Inference adopts the dynamic supernet to individually distribute image patches of k classes to M subnets to obtain better computational performance trade-offs.

Experiments

Qualitative Results

Scale	Method	#Pramas (M)	B100 [31] PSNR↑	B100 [31] #FLOPs (G)	Urban100 [16] PSNR↑	Urban100 [16] #FLOPs (G)	Test2K [9] PSNR↑	Test2K [9] #FLOPs (G)	Test4K [9] PSNR↑	Test4K [9] #FLOPs (G)
×2	SRResNet [24]	1.52	32.19	20.78 (100%)	32.11	20.78 (100%)	30.39	20.78 (100%)	31.90	20.78 (100%)
	+ClassSR [23]	3.12	31.68	14.75 (71%)	31.15	16.21 (78%)	30.24	14.13 (68%)	31.89	13.51 (65%)
	+ARM [4]	1.52	31.69	16.21 (78%)	31.16	16.83 (81%)	30.26	15.59 (75%)	31.90	13.71 (66%)
	+WBSR (Ours)	1.52	32.15	12.26 (59%)	31.98	13.30 (64%)	30.41	12.05 (58%)	32.02	12.68 (61%)
	RCAN [51]	15.59	32.40	130.40 (100%)	32.33	130.40 (100%)	30.86	130.4 (100%)	32.26	130.40 (100%)
	+ClassSR [23]	30.10	31.88	91.28 (70%)	31.72	103.02 (79%)	30.79	83.46 (64%)	32.24	83.46 (64%)
	+ARM [4]	15.59	31.89	99.10 (76%)	31.74	109.54 (84%)	30.80	105.62 (81%)	32.24	97.80 (75%)
	+WBSR (Ours)	15.59	32.34	88.67 (68%)	32.31	96.50 (74%)	30.91	75.63 (58%)	32.37	77.65 (60%)
×4	SRResNet [24]	1.52	27.34	5.19 (100%)	25.30	5.19 (100%)	26.19	5.19 (100%)	27.65	5.19 (100%)
	+ClassSR [23]	3.12	26.53	3.83 (74%)	24.53	4.23 (81%)	26.20	3.62 (70%)	27.66	3.30 (63%)
	+ARM [4]	1.52	26.53	4.34 (83%)	24.54	4.48 (86%)	26.21	3.76 (72%)	27.66	3.33 (64%)
	+WBSR (Ours)	1.52	27.36	3.99 (77%)	25.32	4.36 (84%)	26.26	3.37 (65%)	27.73	3.22 (62%)
	RCAN [51]	15.59	27.76	32.60 (100%)	25.82	32.60 (100%)	26.39	32.60 (100%)	27.89	32.60 (100%)
	+ClassSR [23]	30.10	26.70	22.82 (70%)	25.13	26.08 (80%)	26.39	21.22 (65%)	27.88	19.49 (60%)
	+ARM [4]	15.59	26.74	25.75 (79%)	25.14	28.36 (87%)	26.39	26.70 (82%)	27.88	25.10 (77%)
	+WBSR (Ours)	15.59	27.75	25.10 (77%)	25.81	27.01 (83%)	26.45	18.52 (57%)	27.94	19.40(59%)

Table 1. Quantitative comparison results of different methods on four testing datasets.

Qualitative Results

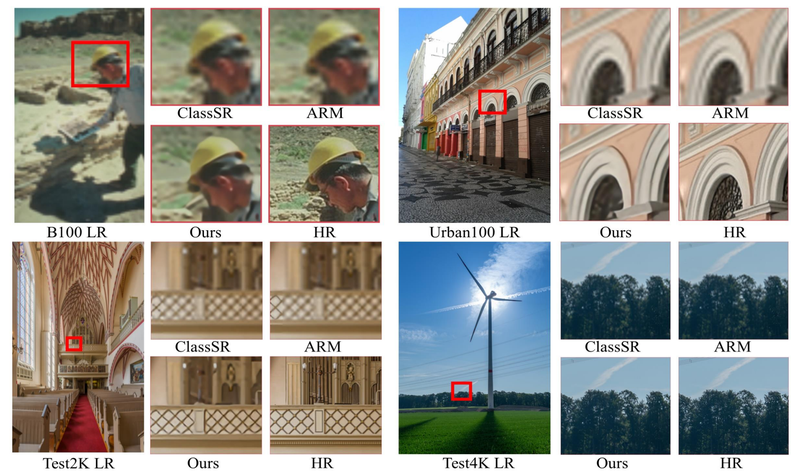


Figure 3. Qualitative comparison results of different methods on four testing datasets.

Ablation Studies

Method (×4)	Test2K [9] PSNR ↑	Test2K [9] SSIM↑	Test2K [9] #FLOPs (G)	Test4K [9] PSNR↑	Test4K [9] SSIM↑	Test4K [9] #FLOPs (G)
SRResNet [24]	26.19	0.7624	5.19 (100%)	27.65	0.7966	5.19 (100%)
+HES	26.24	0.7665	3.58 (69%)	27.71	0.7986	3.43 (66%)
+$\mathcal{L}_{bd}$	26.21	0.7658	3.58 (69%)	27.70	0.7984	3.43 (66%)
+WBSR	26.26	0.7673	3.37 (65%)	27.73	0.7993	3.22 (62%)
+WBSR†	26.38	0.7684	5.19 (100%)	27.80	0.8026	5.19 (100%)
RCAN [51]	26.39	0.7706	32.60 (100%)	27.89	0.8058	32.60 (100%)
+HES	26.43	0.7748	20.86 (64%)	27.92	0.8086	19.89 (61%)
+$\mathcal{L}_{bd}$	26.42	0.7746	20.86 (64%)	27.91	0.8077	19.89 (61%)
+WBSR	26.45	0.7755	18.52 (57%)	27.94	0.8106	19.40 (59%)
+WBSR†	26.51	0.7756	32.60 (100%)	28.10	0.8138	32.60 (100%)

Table 2. Ablation studies of our WBSR. † indicates using the supernet for inference.