录用信息：NeurIPS 2023
代码链接：https://github.com/XPixelGroup/TGSR

Real-World Image Super-Resolution as Multi-Task Learning

Wenlong Zhang[1,2], Xiaohui Li[2,3], Guangyuan Shi[1], Xiangyu Chen[2,4,5]
Xiaoyun Zhang[3], Yu Qiao[2,5], Xiao-Ming Wu[1]†, Chao Dong[2,5]†
[1]The HongKong Polytechnic University, [2]Shanghai AI Laboratory
[3]Shanghai Jiao Tong University [4]University of Macau
[5]Shenzhen Institute of Advanced Technology, CAS
wenlong.zhang@connect.polyu.hk, xiao-ming.wu@polyu.edu.hk, chao.dong@siat.ac.cn

Abstract

Real-World Image Super-Resolution (Real-SR)：

包含多种未知的退化类型的图像超分辨率任务

新的角度：Multi-task Learning ---- 使用单个共享模型解决多个不同退化类型 (degradation) 任务

提出问题：Task Competition / Task Conflict

TGSR (Task Grouping)：识别Real-SR模型性能差对应的退化任务，将这些任务进行分组用于Fine-tune

实验结果：在各种退化场景下都能显著提高性能

Background

Image Super-Resolution：

合成数据超分 (Bicubic Downsampling / Single Image Super-Resolution，SISR)：

- Low-Resolution (LR) 图像从 High-Resolution (HR) 图像downsample获得 (数据集易获取)
- 泛化到真实场景存在着很大问题

盲超分 (Blind-Super-Resolution，Blind-SR)：

- 在Bicubic / Bilinear downsample的基础上进一步考虑了Blur、Noise等一系列因素
- 真实场景的退化类型远比盲超分目前建模的模型面对的退化类型多得多

真实超分 (Real-World Image Super-Resolution，Real-SR)：

- 数据量少，现存数据集少 (RealSR，DRealSR)
- 设计各种退化模型来模拟真实场景中的退化过程 (BSRGAN)
- 处理真实世界成像条件中存在的复杂和未知的退化和限制 (Blur、Noise、Compression Artifacts、Sensor Limitations)

Introduction

Prior Studies：使用复杂的退化模型生成的训练数据来训练单个网络

- 改进Backbone (SwinIR)
- 优化推理效率 (DASR)
- 增强泛化能力 (Reflash Dropout)
- 提高Modulation能力 (MM-RealSR)

Multi-task Learning的角度看Real-SR及存在的问题：

- 单个共享模型同时解决大量不同的退化任务
- 存在任务竞争或任务冲突
- 在很大一部分退化任务中无法产生令人满意的结果

整体的方法：

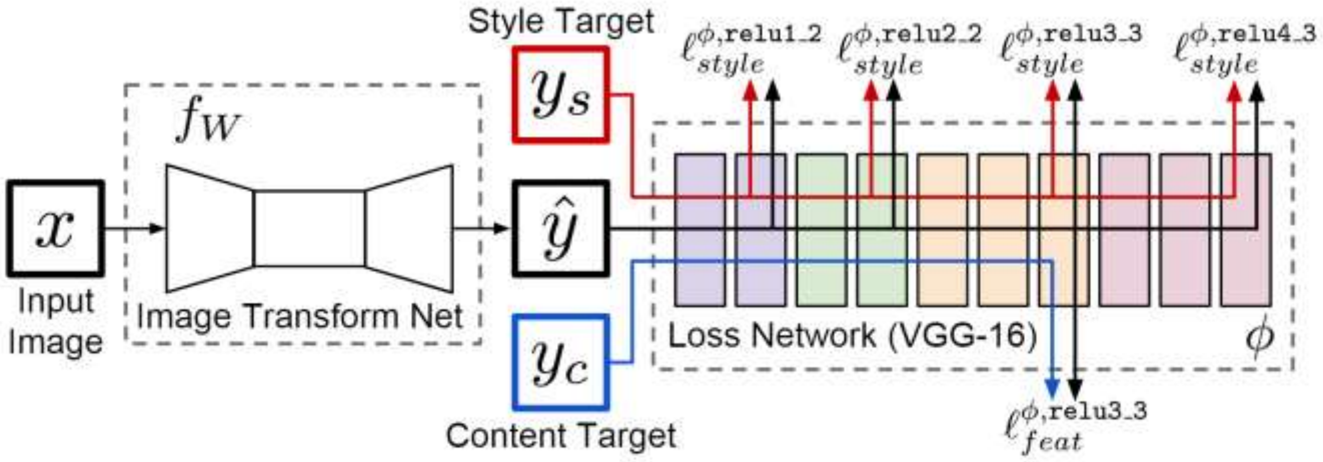
- Task Grouping ---- 减轻任务竞争带来的影响
- Performance Indicator ---- 区分退化任务是否Satisfactory
- Performance Improvement Score ---- 给Unsatisfactory的任务分组
- TGSR (Task Grouping-based Real-SR)

Real-SR as a Multi-task Learning Problem

Real-SR：从LR图像 y 中恢复HR图像 x ：

$$y = \mathcal{D}(x) = (f_n \circ \cdots \circ f_2 \circ f_1)(x)$$

f_i代表退化函数，$\mathcal{D}$代表了一个巨大的连续退化空间。



某种退化类型对应的SR任务 τ 可以定义为训练数据对：$(x, y = d(x))$，d是退化空间。

给定一组HR图像 $\mathcal{X}$ ，从 $\mathcal{D}$ 中采样 N $(N \gg |\mathcal{X}|)$ 种不同的退化，可以得到一组SR任务：

$$\mathcal{T} = \{\tau_i = (x_i, y_i)\}_{i=1}^{N}, x_i \in \mathcal{X}$$

一个Real-SR模型通常需要最小化该式来训练：$\mathcal{L}_{total}(\theta) = \sum_{i=1}^{N} \mathcal{L}_i(\tau_i; \theta)$

* TGSR遵循GAN-based SR Loss $\mathcal{L}$ (GAN Loss $\mathcal{L}_{GAN}$，L1 Loss $\mathcal{L}_1$，Perceptual Loss $\mathcal{L}_{per}$)训练Real-SR网络：

$$\mathcal{L} = w_1 \times \mathcal{L}_1 + w_2 \times \mathcal{L}_{per} + w_3 \times \mathcal{L}_{GAN}$$

Task Competition

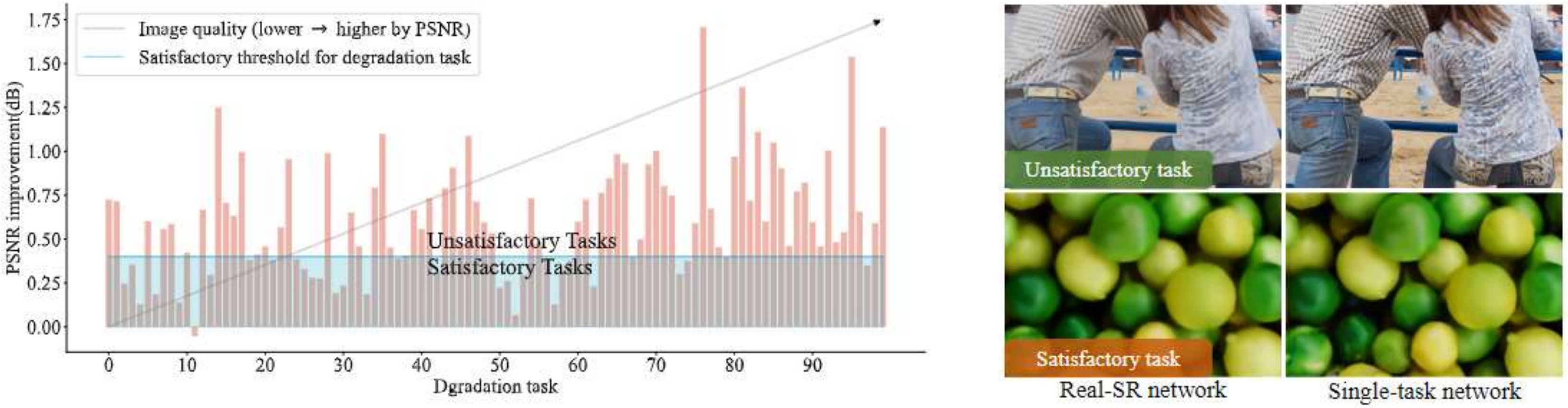
使用Real-ESRGAN中的退化模型随机抽样100个退化任务，训练一个Real-SR网络

在每个退化任务上独立地对Real-SR网络进行Fine-tune，获得100个微调的single-task网络

在每个退化任务上计算Real-SR网络和single-task网络之间的PSNR距离

Negative Transfer

Task Grouping



(a) PSNR distance between the real-SR and single-task SR networks across 100 randomly selected degradation tasks.

(b) Visual comparison between satisfactory and unsatisfactory tasks.

Figure 1: **Illustration of task competition.** The jointly-trained multi-task real-SR network falls short in producing satisfactory results for almost half of the degradation tasks, on which the fine-tuned single-task networks obtain more than 0.4dB performance gain, indicating these tasks are dominated by other tasks during the learning process.

Real-SR via Task Grouping

Performance indicator：

$$z_t^i = \frac{\mathcal{L}_i(\tau_i, \theta_i^t)}{\mathcal{L}_i(\tau_i, \theta)}$$

- 在退化任务τ_i上对预训练的Real-SR网络进行少次微调
- 比较退化任务τ_i在更新θ前后的Loss

Average Performance indicator：

$$\hat{z}_i = \frac{1}{t_e - t_s} \sum_{t=t_s}^{t_e} z_i^t$$

Unsatisfactory的退化任务仍很多且潜在差异大 ➞

- 在每个退化任务上Fine-tune预训练的Real-SR网络 ➞ 耗时
- Average Performance indicator ➞ 不够准确

Algorithm 1: Degradation Task Grouping for Real-SR

Input: A set of degradation tasks $\mathcal{T} = \{\tau_1, \tau_2, ..., \tau_n\}$, the pre-trained real-SR model θ, the number of groups c, and the threshold values $t_0, t_1, \cdots, t_c$.

for *any* $\tau_i \in \mathcal{T}$ **do**

  Compute the average performance indicator $\hat{z}_i$ with Eq. 4;

end

```
// Select unsatisfactory tasks
```

Let $\hat{\mathcal{T}} = \{\tau_i | \hat{z}_i > t_0\}$;

```
// Group unsatisfactory tasks
```

for $i = 1, \ldots, c$ **do**

  Fine-tune the pre-trained real-SR network with all unsatisfactory tasks;

  for *any* $\tau_j \in \hat{\mathcal{T}}$ **do**

    Compute the performance improvement score s_j with Eq. 5;

  end

  Let $G_i = \{\tau_j | s_j > t_i\}$, where t_i is the threshold for group i;

  Let $\hat{\mathcal{T}} = \hat{\mathcal{T}} \backslash G_i$;

end

Output: Degradation groups $G = \{G_1, G_2, ..., G_c\}$.

权衡方案：同时在所有Unsatisfactory任务上Fine-tune预训练的Real-SR网络 θ，得到新网络 $\hat{\theta}$，计算

***Performance Improvement Score (PIS)*：**

$$s_i = I(\mathcal{D}_i^{val}; \hat{\theta}) - I(\mathcal{D}_i^{val}; \theta)$$

其中$\mathcal{D}_i^{val}$是Unsatisfactory任务τ_i的验证集，I是IQA。

选择PIS大于某个阈值的任务来组成一个任务组，重复这个过程寻找剩余的任务组。

TGSR: A Task Grouping Based Real-SR Strategy

Fine-tune the Pre-trained Real-SR network by sampling the degradation task from the found groups

TGSR: A Task Grouping Based Real-SR Network

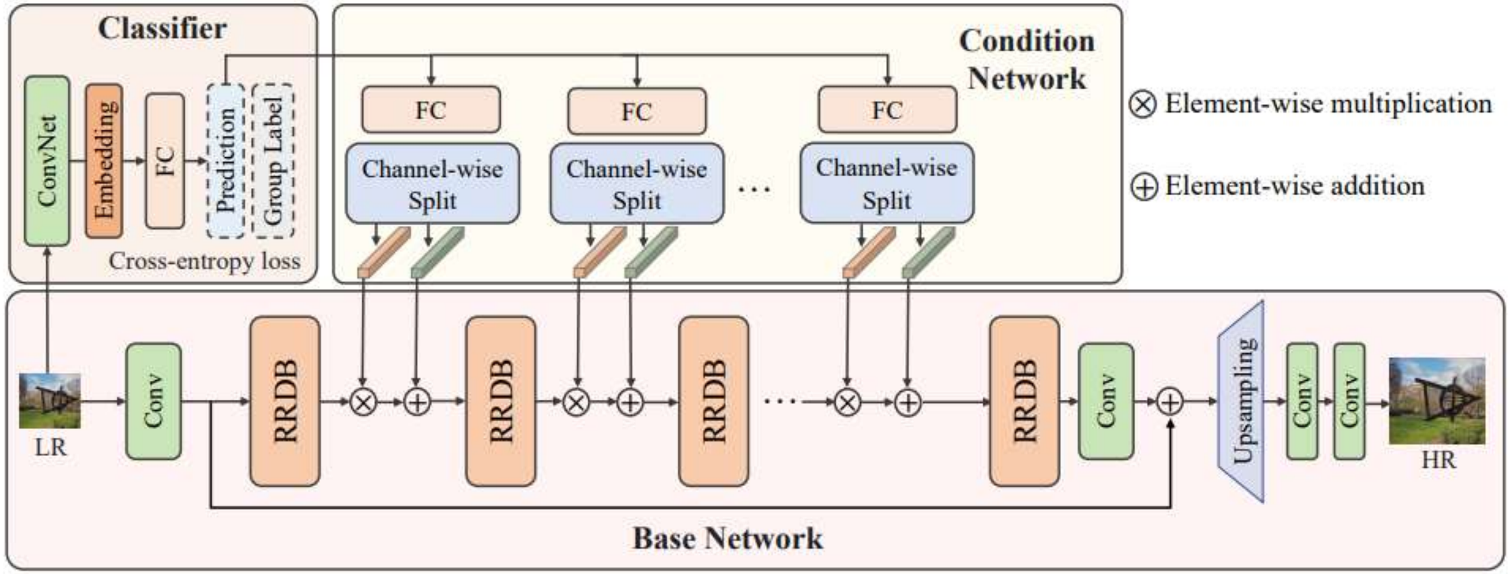


Figure 3: The proposed TGSR consists of a base SR network, a degradation task group classifier and a condition network. The prediction of the classifier is fed to the condition network, generating condition vectors to modulate the intermediate features of the base network for the restoration process.

Experiments

Training Datasets: DIV2K、Flickr2K、OutdoorSceneTraining
Evaluation Datasets:

- DIV2K ➞ DIV2K5G
- AIM2019 DIV2K_random
- RealSR set

All evaluations are conducted on $\times$ 4 SR

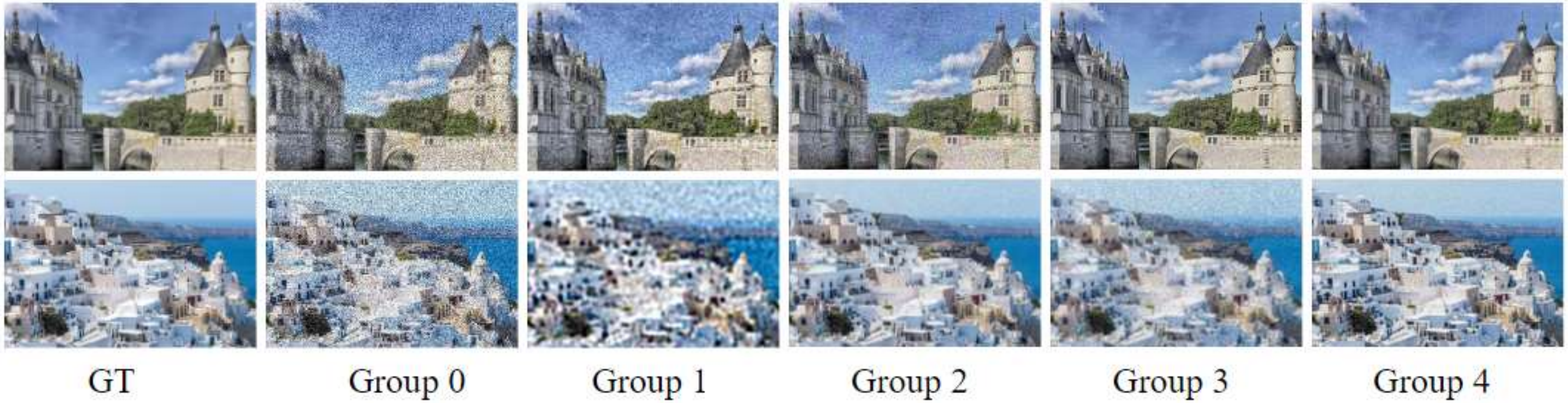


Figure 3: Sample images from different degradation task groups in our DIV2K5G datasets.

Experiments

- Real-SR退化模型：RealESRGAN
- 从整个退化空间采样$N(4 \times 10^3)$个退化任务
- Single-task模型通过预训练模型进行100次Fine-tune后计算后10次迭代的Average performance indicator
- API Top 40%的退化任务选为Unsatisfactory任务
- PIS阈值[0.8,0.6,0.4,0.2]，每组任务数[14,29,84,200]，对应Group1-4，其他记为Group0

Table 1: Quantitative results of different methods on DIV2K5G. Group0 denotes the validation set with satisfactory degradation tasks, and Group1-4 represent the validation sets with unsatisfactory degradation tasks. The ESRGAN trained on the non-blind setting and RDSR trained on the MSE-based setting are marked in gray.

	Group0		Group1		Group2		Group3		Group4	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
ESRGAN	21.49	0.6166	20.29	0.6697	20.95	0.6530	22.42	0.5940	22.02	0.5837
RDSR	25.00	0.5196	21.18	0.6167	23.06	0.5586	25.00	0.4993	25.58	0.4814
MM-RealSR	23.49	0.4549	19.74	0.5326	21.76	0.4731	23.37	0.4205	23.97	0.3989
DASR	23.87	0.4683	19.97	0.5752	22.31	0.5093	24.00	0.4474	24.85	0.4225
BSRGAN	23.99	0.4549	20.07	0.5799	22.45	0.4961	24.34	0.4388	24.80	0.4210
RealSRGAN	23.88	**0.4599**	20.10	0.5586	22.12	0.5019	23.85	0.4499	24.48	0.4270
RealSRGAN-TG	**23.95**	0.4617	**21.14**	**0.5323**	**23.06**	**0.4802**	**24.69**	**0.4248**	**25.05**	**0.4112**
RealESRGAN	23.85	0.4325	20.10	0.5355	22.07	0.4701	24.30	0.4147	24.58	0.3970
RealESRGAN-TG	**23.99**	**0.4286**	**21.10**	**0.5056**	**23.15**	**0.4494**	**24.62**	**0.3975**	**25.03**	**0.3851**
RealSwinIR	23.35	0.4468	19.60	0.5624	21.65	0.4905	23.66	0.4265	24.16	0.4077
RealSwinIR-TG	**23.90**	**0.4168**	**20.62**	**0.4925**	**22.70**	**0.4377**	**24.56**	**0.3815**	**24.98**	**0.3670**
RealHAT	24.26	0.4084	20.64	0.5022	22.44	0.4468	24.42	0.3918	25.02	0.3734
RealHAT-TG	**24.31**	**0.4110**	**21.41**	**0.4932**	**23.20**	**0.4395**	**24.87**	**0.3841**	**25.32**	**0.3674**

Experiments

Table 2: Quantitative results of different methods on the real-world test set. The ESRGAN trained on the non-blind setting and RDSR trained on the MSE-based setting are marked in gray.

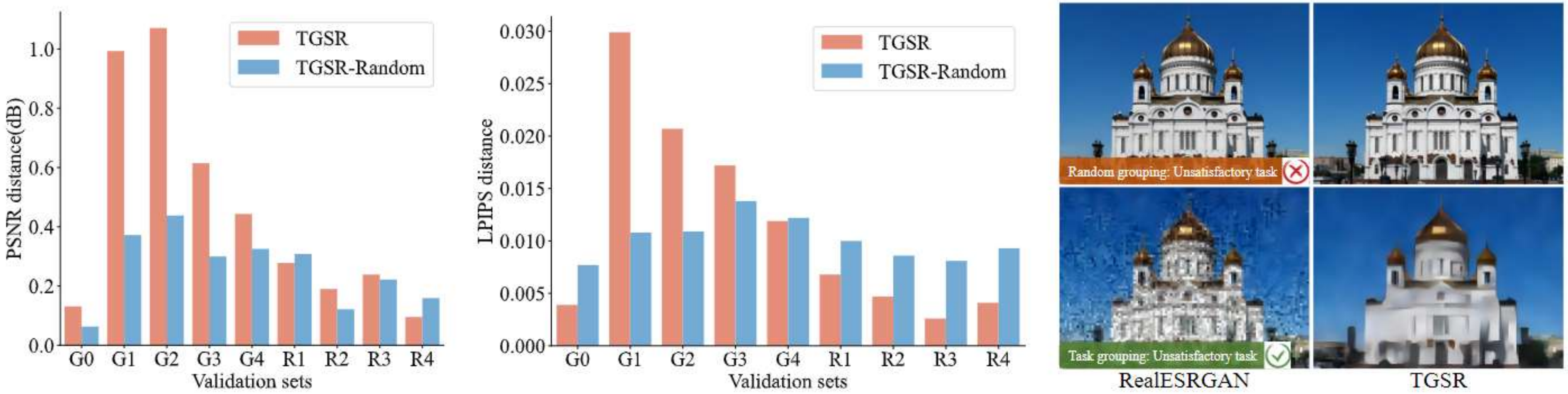
	DIV2K_random		AIM2019		RealSRset-Nikon		RealSRset-Cano	
	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS	PSNR	LPIPS
ESRGAN	20.63	0.6345	23.16	0.5500	27.40	0.4132	27.73	0.4054
RDSR	24.61	0.5268	24.44	0.4803	26.39	0.4053	26.93	0.3795
MM-RealSR	23.17	0.4394	23.48	0.3917	23.78	0.3841	24.42	0.3664
DASR	23.52	0.4832	23.76	0.4210	26.68	0.3972	27.68	0.3792
BSRGAN	23.76	0.4622	24.20	0.4000	26.11	0.3900	26.90	0.3648
SwinIR	23.13	0.4432	23.89	0.3870	26.20	0.3616	26.68	0.3469
RealESRGAN	23.54	0.4423	23.89	0.3960	25.62	0.3820	26.06	0.3629
RealSRGAN	23.58	0.4710	23.72	0.4247	24.71	0.4159	25.42	0.3902
RealSRGAN-TG	**23.79**	**0.4705**	**23.97**	**0.4174**	**25.18**	**0.4018**	**25.93**	**0.3759**
RealESRGAN	23.54	0.4423	23.89	0.3960	25.62	0.3820	26.06	**0.3629**
RealESRGAN-TG	**23.84**	**0.4368**	**24.27**	**0.3899**	**26.01**	**0.3819**	**26.33**	0.3637
RealSwinIR	23.67	0.4216	23.98	0.3804	25.70	**0.3700**	26.43	**0.3506**
RealSwinIR-TG	**23.74**	**0.4190**	**24.10**	**0.3766**	**26.36**	0.3751	**27.18**	0.3585
RealHAT	24.04	**0.4156**	24.19	0.3742	25.88	**0.3532**	26.56	**0.3339**
RealHAT-TG	**24.21**	0.4189	**24.41**	**0.3723**	**26.12**	0.3635	**26.69**	0.3451

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

Ablation

Effectiveness of task grouping algorithm

Compare RealESRGAN-TG with task grouping and random grouping



(a) Comparison on PSNR. (b) Comparison on LPIPS. (c) Visual comparison

Figure 7: Performance comparison of RealESRGAN and our TGSR/ TGSR with random grouping on (a) PSNR and (b) LPIPS. The results indicate random grouping achieves limited improvement compared with our proposed task grouping. (c) Visual results demonstrate that random grouping chooses a satisfactory degradation task that is not required for further training, while our task grouping method finds an unsatisfactory degradation task that needs to be further improved.

Ablation

Study of the performance upper bound

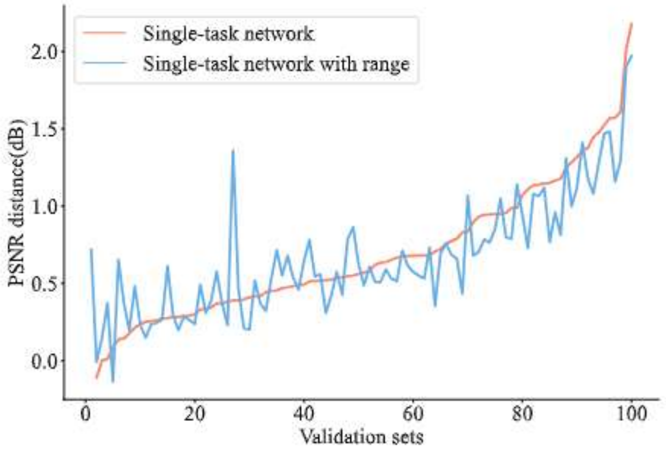


Figure 8: Performance comparison of SR network with a single task and a *range* of tasks.

Table 3: Ablation experiments on the performance upper bound. We fine-tune RealESRGAN directly on each task group to get the corresponding performance upper bound, denoted as RealESRGAN-SG.

Metrics	Model	Group0	Group1	Group2	Group3	Group4
PSNR (↑)	RealESRGAN	23.85	24.58	24.00	22.07	20.10
	RealESRGAN-SG	23.63	24.58	24.39	22.67	20.46
	RealESRGAN-TG (ours)	23.88	25.08	24.62	23.11	21.06
LPIPS (↓)	RealESRGAN	0.4325	0.3970	0.4147	0.4701	0.5355
	RealESRGAN-SG	0.4242	0.3792	0.3929	0.4478	0.4990
	RealESRGAN-TG (ours)	0.4291	0.3845	0.3981	0.4524	0.5087

Ablation

Impact of the Increased Task Volume

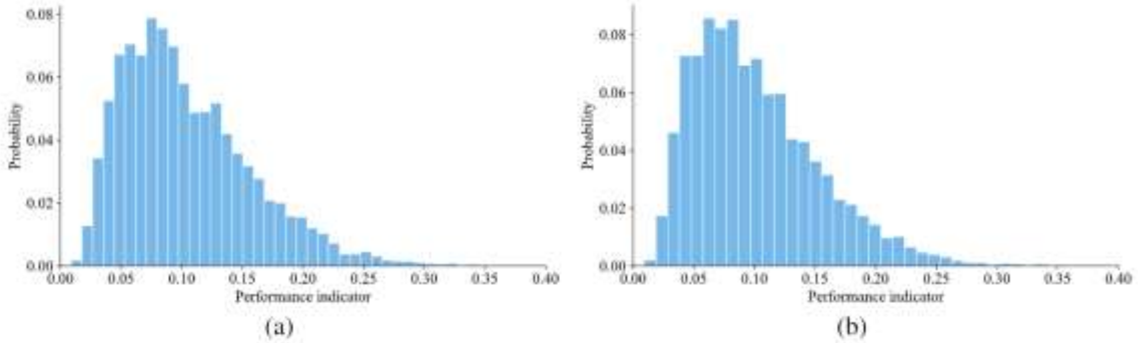


Figure 9: The histograms of the performance indicators with 4k (a) and 10k (b) degradation tasks respectively.

Table 5: Quantitative results of TGSR trained with the 4k and 10k degradation tasks, respectively. Group0 denotes the validation set with satisfactory degradation tasks, and Group1-4 represents the validation sets with unsatisfactory degradation tasks. The suffix 4k and 10k represent groups generated with the 4k and 10k degradation tasks, respectively.

	Metrics	Group0_4k	Group1_4k	Group2_4k	Group3_4k	Group4_4k
RealESRGAN	PSNR (↑)	23.85	20.11	22.08	24.00	24.59
	LPIPS (↓)	0.4325	0.5355	0.4701	0.4147	0.3970
RealESRGAN-TG_4k	PSNR (↑)	23.98	21.10	23.15	24.62	25.03
	LPIPS (↓)	0.4286	0.5056	0.4494	0.3975	0.3851
RealESRGAN-TG_10k	PSNR (↑)	23.91	21.03	22.98	24.50	24.97
	LPIPS (↓)	0.4299	0.5064	0.4525	0.4002	0.3859
	Metrics	Group0_10k	Group1_10k	Group2_10k	Group3_10k	Group4_10k
RealESRGAN	PSNR (↑)	23.81	19.85	20.95	23.77	24.38
	LPIPS (↓)	0.4324	0.5495	0.5073	0.4237	0.4019
RealESRGAN-TG_4k	PSNR (↑)	24.12	20.90	21.85	24.38	25.03
	LPIPS (↓)	0.4272	0.5187	0.4890	0.4046	0.3875
RealESRGAN-TG_10k	PSNR (↑)	24.05	20.94	21.85	24.31	24.93
	LPIPS (↓)	0.4285	0.5171	0.4907	0.4052	0.3886

Ablation

Necessity of Iterative Task Grouping

Table 6: Performance improvement score of some degradation tasks across four training loops. NA means that the task is not available, since it has been grouped in the previous loop.

Training loop	Deg.1	Deg.2	Deg.3	Deg.4	Deg.5	Deg.6	Deg.7	Deg.8	Deg.9	Deg.10
Train1	0.45	1.50	0.60	0.66	0.57	0.75	0.55	0.60	0.61	0.46
Train2	0.47	NA	0.63	0.67	0.41	0.76	0.49	0.59	0.64	0.48
Train3	0.48	NA	NA	NA	0.72	NA	0.65	0.63	NA	0.48
Train4	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Degradation	Deg.11	Deg.12	Deg.13	Deg.14	Deg.15	Deg.16	Deg.17	Deg.18	Deg.19	Deg.20
Train1	0.43	0.60	0.47	0.27	0.66	0.53	0.22	0.31	0.63	0.20
Train2	0.42	0.59	0.47	0.33	0.51	0.52	0.18	0.31	0.67	0.21
Train3	0.48	0.63	0.50	0.31	0.73	0.54	0.18	0.37	NA	0.25
Train4	NA	NA	NA	0.35	NA	NA	0.23	0.46	NA	0.26