A close look at few-shot real image super-resolution from the distortion relation perspective

Xin Li[1] ✉, Xin Jin[2] ✉, Jun Fu[1], Xiaoyuan Yu[3], Bei Tong[3], and Zhibo Chen[1] ✉

[1]*Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China;*
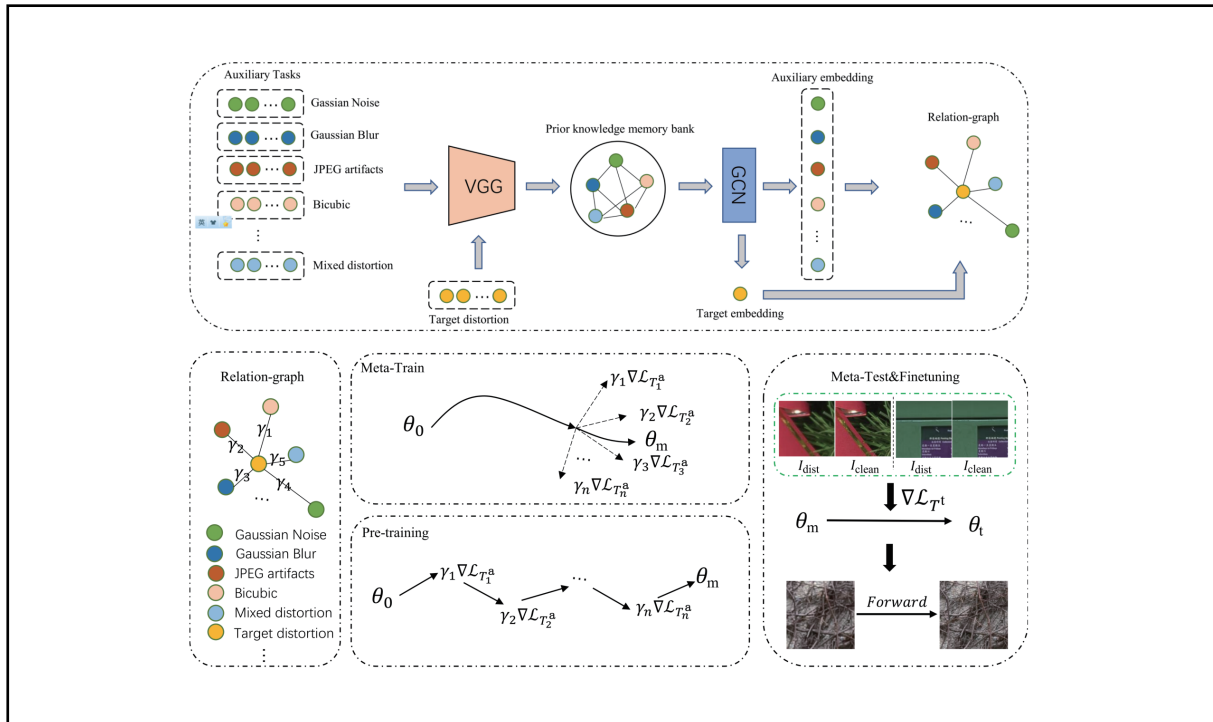[2]*School of Computer Science and Technology, Eastern Institute of Technology, Ningbo, Ningbo 315200, China;*
[3]*Huawei, Hangzhou, 310052, China*

✉Correspondence: Xin Li, E-mail: xin.li@ustc.edu.cn; Xin Jin, E-mail: jinxin@eias.ac.cn; Zhibo Chen, E-mail: chenzhibo@ustc.edu.cn

Graphical abstract

This is our framework that establishes the distortion graph for knowledge transfer in few-shot real image super-resolution.

Public summary

■ This work takes a close look at a brand-new RealSR task, few-shot real image super-resolution (RealSR), which aims to transfer the rich knowledge from auxiliary distortions (i.e., synthesized distortions) to the target RealSR task with few-shot real-world distorted/clean image pairs.

■ This work proposes the distortion relation graph to measure the transferability between distortions, where a prior knowledge memory bank is exploited to store the learned knowledge relation priors from seen auxiliary distortions.

■ This work proposes the novel distortion relation-guided transfer learning (DRTL) for few-shot RealSR, which exploits the distortion relation to reweight optimization direction in the knowledge transfer process, and obtains more reliable and rich restoration knowledge for target distortion from auxiliary distortions with few-shot real distorted/clean image pairs.

A close look at few-shot real image super-resolution from the distortion relation perspective

Xin Li[1] ✉, Xin Jin[2] ✉, Jun Fu[1], Xiaoyuan Yu[3], Bei Tong[3], and Zhibo Chen[1] ✉

[1]*Department of Electronic Engineering and Information Science, University of Science and Technology of China, Hefei 230027, China;*
[2]*School of Computer Science and Technology, Eastern Institute of Technology, Ningbo, Ningbo 315200, China;*
[3]*Huawei, Hangzhou, 310052, China*

✉Correspondence: Xin Li, E-mail: xin.li@ustc.edu.cn; Xin Jin, E-mail: jinxin@eias.ac.cn; Zhibo Chen, E-mail: chenzhibo@ustc.edu.cn

Cite This: *JUSTC*, **2025**, 55(7): 0701 (12pp)   Read Online

Abstract: Collecting amounts of distorted/clean image pairs in the real world is non-trivial, which severely limits the practical application of these supervised learning-based methods to real-world image super-resolution (RealSR). Previous works usually address this problem by leveraging unsupervised learning-based technologies to alleviate the dependency on paired training samples. However, these methods typically suffer from unsatisfactory texture synthesis due to the lack of supervision of clean images. To overcome this problem, we are the first to take a close look at the under-explored direction for RealSR, i.e., few-shot real-world image super-resolution, which aims to tackle the challenging RealSR problem with few-shot distorted/clean image pairs. Under this brand-new scenario, we propose distortion relation guided transfer learning (DRTL) for the few-shot RealSR by transferring the rich restoration knowledge from auxiliary distortions (i.e., synthetic distortions) to the target RealSR under the guidance of the distortion relation. Concretely, DRTL builds a knowledge graph to capture the distortion relation between auxiliary distortions and target distortion (i.e., real distortions in RealSR). Based on the distortion relation, DRTL adopts a gradient reweighting strategy to guide the knowledge transfer process between auxiliary distortions and target distortions. In this way, DRTL is able to quickly learn the most relevant knowledge from the synthetic distortions for the target distortion. We instantiate DRTL with two commonly-used transfer learning paradigms, including pretraining and meta-learning pipelines, to realize a distortion relation-aware few-shot RealSR. Extensive experiments on multiple benchmarks and thorough ablation studies demonstrate the effectiveness of our DRTL.

Keywords: few-shot RealSR; distortion relation graph; transfer learning
CLC number: TP181; TP37   **Document code:** A

1 Introduction

The image restoration (IR) tasks[1–11] aim to restore high-quality images from their degraded low-quality counterparts. The degradation process typically consists of a series of different distortions, such as low-resolution[12,13], motion blur[1,14], noise[2,15], compression artifacts[16–18], rain[7], etc. Although the fast-developed deep learning has significantly promoted the advancement of IR techniques, their success usually relies on training with a large-scale dataset. As a result, numerous image restoration datasets that contain various distortions, such as DIV2K[19], Flick2K[20], Gopro[21], and DID-MDN[22], have been proposed and explored in detail. However, the degradations of these datasets are typically synthetic and differ from those in the real world. To access real-world image restoration, the real-world image super-resolution (RealSR) task is proposed as the pioneer for real-world image restoration, where low-resolution (LR) images suffer from severe hybrid distortions, such as low-resolution, blur, noise, and compression artifacts.

Although some learning-based super-resolution (SR) approaches[23, 24] trained on the above-mentioned datasets have achieved great success in handling the synthetic degradations, such as bicubic degradation, those image SR systems rarely perform well in real-world scenarios since the large distribution gap between synthetic and real-world distortions[25,26]. Furthermore, recollecting a large scale of distorted/clean image training pairs from the real world is also non-trivial. The two above-mentioned weaknesses severely limit their practical application and negatively affect their industrial value. To alleviate the dependency for clean distortion/clean image training pairs, some studies[27–30] have introduced unsupervised learning to real image restoration. However, these methods typically suffer from unsatisfactory/spurious texture synthesis due to the lack of clean image supervision. Compared with a purely unsupervised solution, few-shot RealSR is a potentially under-explored scheme that only exploits few-shot distorted/clean image pairs and is more feasible for tackling this challenging RealSR problem.

Previous works have explored different transfer learning strategies[31–35] for few-shot problems in high-level tasks, extracting useful knowledge from auxiliary tasks to the target task. Among them, pretraining and meta-learning are two representative technologies in few-shot learning, which are also

preliminarily explored in image super-resolution tasks but with different purposes. For instance, IPT[23] introduced a large-scale pretraining dataset to improve the restoration performance w.r.t the target distortion. Soh et al.[36] proposed a meta-learning-based method to implement the fast adaptation for zero-shot super-resolution tasks, achieving a SOTA performance. However, these works still focused on handling the synthetic distortions and were not designed for few-shot learning of image restoration, which is not suitable for a more challenging few-shot RealSR problem.

In this paper, we are the first to take a close look at the challenging few-shot RealSR problem, where we aim to transfer the rich restoration knowledge from auxiliary distortions (i.e., synthetic distortions) to the target distortion (i.e., real-world super-resolution) with few-shot distorted/clean real-world image pairs. However, a large domain gap exists between the auxiliary distortions and the target distortion. Therefore, we face the key issue that needs to be solved: "How to capture the proper prior knowledge for the target distortion from auxiliary distortions?" Different from high-level classification tasks, in which the relation between different tasks can be modeled with a simple clustering, such as KNN[37]. To capture the relation between different distortions adaptively, we propose to learn a distortion relation graph, which consists of two essential factors: nodes (i.e., distortion embeddings) and edges (i.e., the similarity between distortions) as shown in Fig. 1. We introduce a distortion relation network (DRN) to extract/learn the expected distortion relation embeddings. To make the DRN general for arbitrary real image super-resolution (RealSR), we design a prior knowledge memory bank to store the learnable distortion relation priors from seen auxiliary tasks. Given an arbitrary real distorted sample, it can traverse the prior knowledge memory bank to acquire the required distortion embeddings as in the
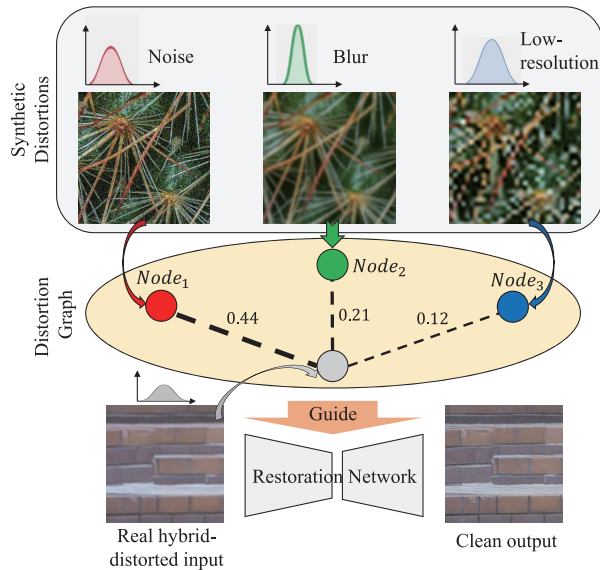
GCN[38].

After obtaining the required distortion embeddings, we compute the edges that represent the distortion similarity between different distortions with cosine similarity as the relation. This relation can measure the transferability of a specific auxiliary distortion to the target distortion. With the guidance of the distortion relation, we propose the novel distortion relation guided transfer learning (DRTL), which utilizes the distortion relation to revise the optimization direction via the gradient reweighting strategy in the transfer learning process and allows the learned knowledge from auxiliary distortions to be more available for the target distortion. We instantiate our DRTL with two commonly-used transfer learning paradigms in few-shot learning: pretraining and meta-learning paradigms (denoted as $DRTL_p$ and $DRTL_m$).

Extensive experiments on several typical backbones have demonstrated the effectiveness of our DRTL on few-shot RealSR. A more thorough analysis of the distortion relation and few-shot RealSR is provided in the ablation studies.

The main contributions of this paper can be summarized as follows:

● To our knowledge, we are the first to take a close look at a brand-new RealSR task, few-shot RealSR, which aims to transfer the rich knowledge from auxiliary distortions (i.e., synthesized distortions) to the target RealSR task with few-shot real-world distorted/clean image pairs.

● We propose the distortion relation graph to measure the transferability between distortions, where a prior knowledge memory bank is exploited to store the learned knowledge relation priors from seen auxiliary distortions.

● With the guidance of the distortion relation, we propose the novel DRTL for few-shot RealSR, which exploits the distortion relation to reweight the optimization direction in the knowledge transfer process, and obtains more reliable and rich restoration knowledge for target distortion from auxiliary distortions with few-shot real distorted/clean image pairs.

● We instantiate our DRTL with two popular transfer learning paradigms: pretraining and meta-learning. Extensive experiments on multiple typical restoration networks and these two transfer learning paradigms have revealed the effectiveness of our proposed DRTL on few-shot RealSR.

The remaining parts of this paper are organized as follows: In Section 2, we comprehensively review the related works for our few-shot RealSR, including image super-resolution and real image restoration. In Section 3, we first introduce background knowledge for two commonly used few-shot learning paradigms, pretraining and meta-learning, and then present our approach which is composed of our proposed distortion relation graph and distortion relation-guided transfer learning. The experimental results are reported in Section 4, where we perform a thorough investigation of our DRTL and few-shot RealSR. In Section 5, we provide a thorough conclusion for our proposed DRTL.

2 Related works

2.1 Image super-resolution

Deep learning has accelerated the development of image

Fig. 1. Illustration of the proposed distortion relation guided transferlearning (DRTL). We introduce a distortion relation graph to capture the relationship (i.e., 0.44, 0.21, and 0.12 as shown above) between auxiliary distortions and real hybrid distortion (i.e., RealSR), Then, we use these relations to adaptively guide the knowledge transfer process for real super-resolution.

super-resolution techniques since the pioneering works of the SRCNN[12] and FSRCNN[39]. Most works[4,23,24,40–42] are devoted merely to the synthetic degradation (i.e., bicubic down-sampling). However, in real-world scenarios, the degradation factors are composed of hybrid distortions, such as blur, noise, and compression artifacts. To tackle the challenge of real-world super-resolution, a series of works[43,44] have applied image translation technology and cycle consistency[45] to implement unsupervised image super-resolution, which has achieved high subjective quality. However, these methods typically suffer from unsatisfactory texture synthesis due to the lack of supervision of clean images. There are also some works[46,47] designed to solve the blind/unknown image super-resolution, of which the degradations are still far from the real-world distortions. Moreover, some real image super-resolution datasets with limited real-world clean/distorted image pairs have been collected and released (i.e., RealSR[48] and DRealSR[49]), which are costly from the perspective of time and manpower. Although some frameworks[26,48–51] have achieved great progress in RealSR, they ignore the fact that limited real-world data will prevent their further improvements. Unlike the above works, we are the first to take a close look at the brand-new direction for RealSR, few-shot RealSR, which is vital for the application of SR methods in the real world.

2.2　Real image restoration

Collecting a large-scale clean/distorted training dataset in the real world is nontrivial, severely preventing the successful application of fully supervised IR methods. To reduce the dependencies for real-world datasets, three categories of schemes have been proposed to improve the performance of the IR model in real-world scenarios: unsupervised learning-based methods[28,52], distortion simulation-based methods[29], and transfer learning-based methods[36,53]. Unsupervised learning-based methods usually adopt unpaired distorted/clean images or only distorted images for optimization with generative models. For instance, Ulyanov et al.[27] proposed to utilize the CNN to capture the deep image statistics through an iterative self-supervised optimization. Yuan et al.[52] introduced a cycle-in-cycle structure to exploit unpaired distorted/clean images for more general image super-resolution. Du et al.[28] introduced a discrete disentangled representation learning method to capture the invariant clean representations from unpaired distorted/clean image pairs. Distortion simulation-based methods[29] aim to simulate real-world distortions and utilize the synthesized datasets to optimize the IR network, which achieves excellent performance in real-world scenarios. The final category of IR methods intends to leverage transfer learning techniques to achieve image restoration for real distorted images. For example, Wei et al.[25,54] proposed to capture the distortion priors of rain streaks from the auxiliary synthetic/fake rain streaks to achieve the clean image restoration via semi-supervised learning. However, the above methods are designed based on the assumption that rain streaks can be modeled with a Gaussian distribution, which is not always satisfactory for other distortions. With the advancement of transfer learning, Soh et al.[36] and Park et al.[55] proposed to leverage the meta-transfer

learning to deal with the challenging task of zero-shot super-resolution (ZSSR). Kim et al.[53] utilized the adaptive instance normalization to realize the knowledge transfer from synthetic noise to real noise. Unlike the above works, which involve knowledge transfer between the homogeneous distortions (e.g., noise to noise, rain to rain), our DRTL focuses on a more challenging setting, where the auxiliary tasks can be heterogeneous synthetic distortions (e.g., the JPEG artifacts to RealSR, mixed distortions to RealSR).

2.3　Transfer learning

Transfer learning aims to excavate the prior knowledge from the source domain/task to improve the performance on the target domain/task, which can be roughly divided into four categories: (ⅰ) feature-based transfer learning[56,57] which transforms the pre-trained features into the target space (i.e., asymmetric feature-based transfer learning) or the common feature space of the source and target domains (i.e., symmetric feature-based transfer learning); (ⅱ) instance-based transfer learning[58] which is achieved by reweighting/selecting source instances to prioritize knowledge relevant to the target domain; (ⅲ) parameter/model-based transfer learning[32,59–60] which tends to transfer the knowledge existing in partial parameters (e.g., feature extractor) or whole model; (ⅳ) relational-based transfer learning[61] aims to excavate and transfer the relation between objects and classes, through graph/prototype modeling. In contrast, few-shot transfer learning seeks to transfer pretrained knowledge to target tasks with few-shot target samples, which is more data efficient but challenging. Early works on few-shot transfer learning[31] were typically based on fine-tuning or model-agnostic meta-learning (MAML)[32] as stated in Section 3.1. However, the above methods usually fail to capture the knowledge structure of downstream tasks with few-shot instances. To overcome this problem, some works establish the class/attribute-wise graphs/prototypes to enhance the structural knowledge perception of the target tasks. Recent advancements in large foundation models, such as vision language models (VLMs)[62] have demonstrated exceptional transferability on downstream tasks, which has spurred two typical efficient transfer learning methods, prompt learning and adapter-style tuning. Among them, prompt learning-based methods[63–65] intend to design the textual/visual prompts to adjust the input features of modules for downstream tasks, whereas adapter-style tuning[66–68] introduces few extra parameters to adjust the output/middle features of modules. The above progress on few-shot transfer learning has inspired us to develop our DRTL for the real image super-resolution task.

3　Approach

3.1　Recap of transfer learning

In this section, we clarify two commonly-used transfer learning methods, pretraining and meta-learning, as the basis of our proposed distortion relation-guided transfer learning (DRTL).

3.1.1　Pretraining-based transfer learning

As the basic transfer learning technology, pretraining has

been widely applied to different vision tasks[23,69]. Pretraining-based transfer learning can be divided into two processes, including pretraining and fine-tuning. Given the auxiliary tasks $\mathcal{T}^{\mathrm{a}}$, target task $\mathcal{T}^{\mathrm{t}}$, and a learning model f_θ with parameters θ, the pretraining process aims to learn the optimal model parameters on auxiliary tasks $\mathcal{T}^{\mathrm{a}}$ as Eq. (1) to obtain the task-relevant knowledge,

$$\min_\theta \sum_{\mathcal{T}_i^{\mathrm{a}} \sim p(\mathcal{T}^{\mathrm{a}})} \mathcal{L}_{\mathcal{T}_i^{\mathrm{a}}}(f_\theta(x_i^{\mathrm{a}}), y_i^{\mathrm{a}}), \tag{1}$$

where the $\mathcal{L}_{\mathcal{T}_i^{\mathrm{a}}}$ refers to the optimization loss in the ith auxiliary task $\mathcal{T}_i^{\mathrm{a}}$. x_i^{a} and y_i^{a} denote the samples and their corresponding labels of the task $\mathcal{T}_i^{\mathrm{a}}$. The optimization objective of pretraining is to minimize the loss function in all auxiliary tasks.

In the fine-tuning stage, the best parameters θ_{m} in the pretraining stage are used as the initial parameters. Then, the model $f_{\theta_{\mathrm{m}}}$ is subsequently updated with the optimization objective to achieve the best performance in the target task as expressed in Eq. (2):

$$\min_\theta \mathcal{L}_{\mathcal{T}^{\mathrm{t}}}(f_{\theta_{\mathrm{m}}}(x^{\mathrm{t}}), y^{\mathrm{t}}), \tag{2}$$

where x^{t} and y^{t} are the training samples and their corresponding labels of the target task.

3.1.2 Meta-learning based transfer learning

Unlike pretraining-based transfer learning, which directly learns the task-relevant knowledge from auxiliary tasks, meta-learning-based transfer learning aims to learn a fast adaptation capability to all auxiliary tasks[32]. The typical work is MAML[32] (model-agnostic meta-learning) which can be divided into two processes: meta-train and meta-test. For the meta-train process, the model is first optimized with multiple tasks $\{\mathcal{T}_i^{\mathrm{a}} \in \mathcal{T}^{\mathrm{a}}, 1 \leqslant i \leqslant N\}$ as Eq. (3).

$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i^{\mathrm{a}}}(f_\theta(x_i^{\mathrm{a}}), y_i^{\mathrm{a}}). \tag{3}$$

To enable the model to have a fast adaptation capability with few-shot samples, the model is optimized through the second-order gradient from the multiple tasks of $\mathcal{T}^{\mathrm{a}}$ based on

the look-ahead gradient θ_i' in Eq. (3) as:

$$\theta_{\mathrm{m}} = \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i^{\mathrm{a}} \sim p(\mathcal{T}^{\mathrm{a}})} \mathcal{L}_{\mathcal{T}_i^{\mathrm{a}}}(f_{\theta_i'}(x_i^{\mathrm{a}}), y_i^{\mathrm{a}}). \tag{4}$$

The meta-test stage is the same as the fine-tuning stage, which can be represented as Eq. (2).

3.2 Building distortion relation graph

The distortion relation graph aims to explore and measure the transferability between multiple synthetic distortions and target real-world distortions. Once such a relation graph is obtained, the most transferable auxiliary distortions can be selected to assist in the optimization of the target RealSR with few-shot samples. Moreover, we could leverage the distortion relation to guide the knowledge transfer process from auxiliary distortions to the target RealSR. In this paper, we propose a distortion relation graph, where the nodes in the graph denote the relation embeddings between each sample and the prior knowledge memory bank. The edges are created based on the similarity between different nodes corresponding to different distortions (i.e., the transferability between different distortions). The architecture of this distortion relation graph is described in Fig. 2 and the graph is generated with our proposed distortion relation network (DRN), which consists of four key components: distortion-aware feature extractor, prior knowledge memory bank, relation-aware nodes and edges. We will introduce each component in detail in the following subsections.

3.2.1 Distortion-aware feature extractor

To reduce interference from the texture and structure of the distorted image to extracting the distortion features, we compute the residual I_{r} between the clean image I and the distorted image I_{d} as $I_{\mathrm{r}} = I - I_{\mathrm{d}}$, which are the inputs of the distortion-aware feature extractor. Notably, previous works[70–71] have demonstrated the effectiveness of the VGG[72] network for the perception of distortions. Inspired by this, we propose to utilize VGG-11[72] as our distortion-aware feature extractor to extract the distortion representation for each distortion as $F_{\mathrm{d}} = \mathrm{VGG}(I_{\mathrm{r}})$, and the feature extractor is optimized with the distortion classification.
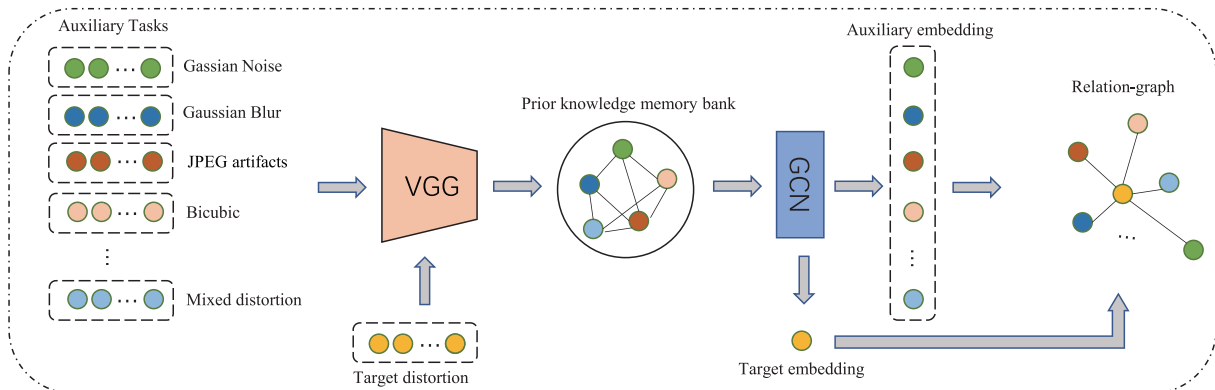

Fig. 2. Architecture of our proposed distortion relation network (DRN), which consists of a feature extractor (VGG11), prior knowledge memory bank, and Graph convolution network. The representation of each distortion sample can be tapped into the prior knowledge memory bank to capture the distortion embedding with the GCN[38] as the node of the distortion relation graph. Then the edges of the distortion relation graph can be computed with the cosine similarity of different nodes.

3.2.2 Knowledge memory bank

Generally, graph learning aims to capture the internal relations between samples or structures, and then assist with the current task[38,73]. To capture the distortion relation with the graph, both the target distortion and auxiliary synthetic distortions must be taken into the same graph as nodes and infer their relation weights (i.e., the edge of the graph). However, directly predicting relation weights between synthetic and real distortions will inevitably face two essential issues. First, the target distortion cannot be achieved before the establishment of a distortion relation graph. Second, directly establishing the relation graph with the current real target distortion may make the graph hard to generalize well to other unseen real distortions. To obtain a universal and general distortion relation graph, we assign a long-term distortion memory bank to record the knowledge from previously seen distortion types, where the distortion knowledge is stored with a memory graph $\mathcal{M} = (\mathcal{H}_{\mathcal{M}}, \mathcal{A}_{\mathcal{M}})$. Here, the memory nodes $\mathcal{H}_{\mathcal{M}}$ and edges $\mathcal{A}_{\mathcal{M}}$ are trainable in the optimization process of our distortion relation network (DRN). In this way, each distortion can be tapped into the distortion memory bank to obtain its corresponding representation/node in the same space, which is suitable for measuring the distortion relation. Note that, with the distortion memory bank, the target distortion and auxiliary distortions can extract their corresponding nodes independently.

3.2.3 Graph nodes and edges

Our distortion relation graph $\mathcal{R} = (\mathcal{C}_{\mathcal{R}}, \mathcal{A}_{\mathcal{R}})$ consists of two essential components, nodes $\mathcal{C}_{\mathcal{R}} = \{c^i | \forall i \in [1, K]\} \in \mathbb{R}^{K \times d}$ and edges $\mathcal{A}_{\mathcal{R}} = \{|\mathcal{A}_{\mathcal{R}}(c^i, c^j)| \forall i, j \in [1, K]\} \in \mathbb{R}^{K \times K}$, where k and d are the number and dimension of nodes. As stated in the above section, the nodes of the distortion relation graph denote the feature representation of each distortion, which is obtained by projecting distorted samples to the knowledge memory bank $\mathcal{M} = (\mathcal{H}_{\mathcal{M}}, \mathcal{A}_{\mathcal{M}})$. To generate the nodes of distortions, we have to warp the features of different distortions into the same space as the memory bank. Concretely, we first compute the distance between the ith distortion feature F_d^i and each node of the memory bank as Eq. (5):

$$\mathcal{A}_p^i = \sigma(\|F_d^i - \mathcal{H}_{\mathcal{M}}\|_2^2), \tag{5}$$

where σ is a linear transform. We can then warp the distortion feature into the space of the memory bank with a graph convolution network (GCN)[38]. In the computing process, the

---

Algorithm 1: Distortion relation graph

1: **Inputs:** Auxiliary distortions $\mathcal{T}_i^a$, where $1 \leqslant i \leqslant N$ and N represents the number of the auxiliary distortions; target distortion $\mathcal{T}^t$ (i.e., RealSR); our proposed distortion relation network (DRN).

2: Get the distortion embedding nodes c_a^i for each auxiliary distortion with $c_a^i = \text{DRN}(\mathcal{T}_i^a)$;

3: Get the distortion embedding nodes c_t for target distortion with $c_t = \text{DRN}(\mathcal{T}^t)$;

4: Compute the relation of the ith auxiliary distortion and target distortion with $\gamma_i = \text{cosine\_simlarity}(c_a^i, c_t)$;

5: **Output:** The relation matrix $\{\gamma_i | 1 \leqslant i \leqslant N\}$.

Algorithm 2: The pretraining stage of DRTL$_p$

1: **Inputs:** Auxiliary distortions $\mathcal{T}_i^a$, where $1 \leqslant i \leqslant N$; Target distortion $\mathcal{T}^t$; Relation matrix $S = \{\gamma_i | 1 \leqslant i \leqslant N\}$; A restoration model f_θ; learning rate: α;

2: **for** iteration = 0 to T **do**

3:     **for** $i = 0$ to N **do**

4:         Update θ with $\theta_i' = \theta - \alpha \nabla_\theta \gamma_i \mathcal{L}_{\mathcal{T}_i^a}(f_\theta)$;

5:     **end for**

6: **end for**

7: **Output:** The optimal initial restoration network f_{θ_m} on target distortion.

---

adjacent matrix contains two parts, including the edges of the memory bank $\mathcal{A}_{\mathcal{M}}$ and the distance matrix of each distortion with memory nodes $\mathcal{A}_p^i$ are as:

$$\mathcal{A} = [\mathcal{A}_p^i, \mathcal{A}_{\mathcal{M}}], \tag{6}$$

where $\mathcal{A}_{\mathcal{M}}$ can be computed by measuring the distances between different memory nodes as:

$$\mathcal{A}_{\mathcal{M}} = \{\sigma(\|\mathcal{H}_m - \mathcal{H}_n\|_2^2) | m, n \in \{1, Q\}\}. \tag{7}$$

In this way, we can obtain the nodes of each distortion in the memory space with GCN[38] as:

$$c_i = \text{GCN}([F_d^i, \mathcal{H}_{\mathcal{M}}], \mathcal{A}). \tag{8}$$

After obtaining the nodes $\mathcal{C}_{\mathcal{R}} = \{c^i | \forall i \in [1, K]\} \in \mathbb{R}^{K \times d}$ for distortion relation graph, we utilize cosine similarity to measure the distance of different distortion nodes, which act as the edges $\mathcal{A}_{\mathcal{R}}$ of the distortion relation graph and are used to measure the transferability between different distortions. Based on the above methods, we can obtain the relation γ between each auxiliary distortion and target distortion as Algorithm 1.

3.3 Distortion relation-guided transfer-learning

To transfer the knowledge from auxiliary distortions to the target RealSR with few-shot distorted/clean image pairs, one naive method is to utilize the classic transfer learning technology to achieve few-shot RealSR. However, directly applying transfer learning ignores the relation between auxiliary distortions and the target distortion, and thus easily results in a suboptimal solution. To extract reliable and optimal distortion knowledge from auxiliary synthetic distortions, in this paper, we propose distortion relation-guided transfer-learning (DRTL). The workflow of DRTL is described in Fig. 3, which is based on two intuitions: (ⅰ) multiple auxiliary distortions that are similar to the target RealSR can provide more transferable knowledge for RealSR than a single distortion; (ⅱ) when the auxiliary distortion is more pretraining/meta-train process.

To achieve this, our proposed DRTL aims to utilize the distortion relation to guide the optimization process of each auxiliary distortion in the transfer learning. The DRTL can be divided into three steps: distortion relation computing, pretraining/meta-train with auxiliary distortions, and fine-tuning/
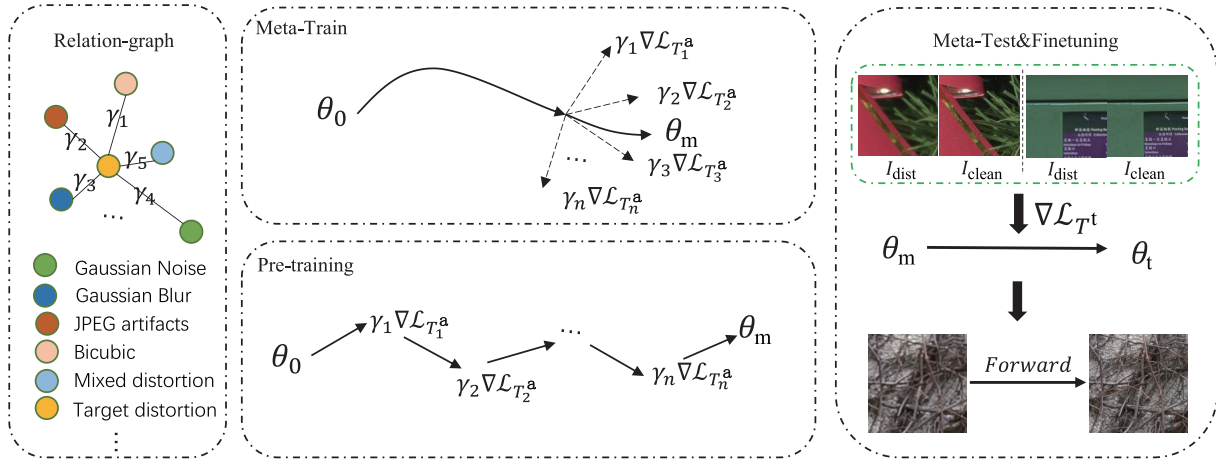
Fig. 3. Overall workflow of distortion relation-guided transfer-learning (DRTL). I_{dist} and I_{clean} denote the distorted and high-quality image pairs in RealSR.

meta-test with the target RealSR.

As shown in Fig. 3, in the first step, we can obtain the distortion relation γ by computing the edges of our pretrained distortion relation graph. Here, γ will increase when the distortions are more relevant. In the second step, we utilize the relation coefficient γ to guide the transfer learning process in a gradient modulation/reweighting manner. In this paper, we instantiate our DRTL with two commonly-used transfer learning techniques, pretraining and MAML, as DRTL_{p} and DRTL_{m}, respectively. For DRTL_{p}, we revise the optimization process of pretraining in Eq. (9) as:

$$\theta_{\text{m}} = \theta - \alpha \nabla_{\theta} \sum_{\mathcal{T}_i^{\text{a}} \sim p(\mathcal{T}^{\text{a}})} \gamma_i \mathcal{L}_{\mathcal{T}_i^{\text{a}}}(f_\theta(I_{d_i}^{\text{a}}), I_i^{\text{a}}), \qquad (9)$$

where $I_{d_i}^{\text{a}}$ and I_i^{a} represent the distorted and clean images in the ith auxiliary distortion task. For DRTL_{m}, we revise the optimization process of meta-train in Eq. (10) as:

$$\theta_{\text{m}} = \theta - \beta \nabla_{\theta} \sum_{\mathcal{T}_i^{\text{a}} \sim p(\mathcal{T}^{\text{a}})} \gamma_i \mathcal{L}_{\mathcal{T}_i^{\text{a}}}(f_{\theta_i'}(I_{d_i}^{\text{a}}), I_i^{\text{a}}). \qquad (10)$$

In this way, the optimization direction of MAML and pre-training can be closer to target distortion, which can capture most of the task-relevant knowledge. In the process of meta-test or fine-tuning, the optimal transferable model parameters θ_{m} need to be fine-tuned only with few-shot distorted/clean pairs of target distortions. The algorithms of the pretraining stage in DRTL_{p} and the meta-train stage of DRTL_{m} are shown in Algorithms 2 and 3.

4 Experiments

In this section, we first introduce the auxiliary distortion datasets and target RealSR datasets in Section 4.1. We present the implementation details of our DRTL in Section 4.2. We then present a comprehensive comparison and analysis of the components in DRTL to demonstrate their effectiveness and superiority in Sections 4.3, 4.4, 4.5, and 4.6.

4.1 Datasets

Auxiliary datasets. To demonstrate the effectiveness and robustness of our DRTL, we select the most relevant 7

Algorithm 3: The meta-train stage of DRTL_{m}

1: **Inputs:** Auxiliary distortions $\mathcal{T}_i^{\text{a}}$, where $1 \leqslant i \leqslant N$; target distortion $\mathcal{T}^{\text{t}}$; relation matrix $S = \{\gamma_i | 1 \leqslant i \leqslant N\}$; learning rate: α and β;

2: **for** iteration = 0 to T **do**

3:     **for** i = 0 to N **do**

4:         Sample batch from auxiliary distortions $\mathcal{T}_i^{\text{a}}$;

5:         Evaluate gradient $\nabla_{\theta}\mathcal{L}_{\mathcal{T}_i^{\text{a}}}(f_\theta)$ with respect to the sampled batch of distortion $\mathcal{T}_i^{\text{a}}$;

6:         Update parameters: $\theta_i' = \theta - \alpha \nabla_{\theta}\mathcal{L}_{\mathcal{T}_i^{\text{a}}}(f_\theta)$;

7:     **end for**

8:     Update parameters: $\theta \leftarrow \theta - \beta \nabla_{\theta} \sum_{i=1}^{K} \gamma_i \mathcal{L}_{\mathcal{T}_i^{\text{a}}}(f_{\theta_i'})$;

9: **end for**

10: **Output:** The optimal initial model $f_{\theta_{\text{m}}}$ on target distortion.

---

common synthetic distortions as auxiliary tasks based on our proposed distortion relation graph, including bicubic downsampling[19], bicubic downsampling with anisotropic kernels[36], Gaussian noise, Gaussian blur, mixed mild distortion, mixed moderate distortion, mixed severe distortion[74]. Then, we take 800 clean images of the DIV2K dataset[19] as original clean images and add the above-mentioned distortions into the clean images to generate 7 auxiliary datasets. The method of distortion generation is shown in Table 1. In particular, the mixed distortions [10, 76] are composed of Gaussian noise, Gaussian blur, and compression artifacts, which are divided into 10 levels with $\sigma \sim [0, 50]$, $\sigma \sim [0, 5]$ and compression quality $q \sim [10, 100]$, respectively. Following Ref. [75], the mixed distortion can be divided into three levels, i.e., mild, moderate, and severe. The mixed mild distortion denotes the total distortion level of Gaussian noise, Gaussian blur, and JPEG artifacts in the range of [9, 11]. Mixed moderate distortion and mixed severe distortion are in the ranges of [12, 17] and [18, 20], respectively.

Target dataset. We select the typical RealSR[48] dataset as our target task. To obtain a few-shot real-world dataset for

training/fine-tuning, we randomly select 30 clean/distorted image pairs from the training set of RealSR[48] as our training dataset with the target distortion. For the final evaluation, we further use 30 clean/distorted images in the test set of RealSR[48] as our final testing dataset. Note that, the target clean-distorted image pairs have the same resolution. The distorted images contain complicated real-world distortions, which are captured by Canon and Nikon cameras[48].

4.2 Implementation details

The implementation of DRTL is based on the PyTorch framework. The whole training process can be divided into two stages: pretraining/meta-train and fin-tuning/meta-test. For the first step, we utilize the Adam optimizer with an initial learning rate of 0.0001 for optimization. For the fin-tuning/meta-test step, the learning rate decays by a factor of 0.8 every 3000 iterations. We set the batch size to 32 and leverage random flipping, rotation, and cropping to achieve data augmentation. The size of the cropped image is 64×64. $\mathcal{L}_1$ loss has been proven effective in optimizing the model, especially in image restoration[42,48]. Therefore, we use only $\mathcal{L}_1$ loss to optimize the DRTL in this paper. For the distortion

relation graph, we optimize the distortion relation network to have the capability to identify distortions and capture their relations with the distortion classification constraint.

4.3 Graph visualization and explanation

In this section, we elaborately discuss how our distortion relation graph works. As shown in Fig. 4a, we visualize the feature embedding for each sample with the auxiliary and target distortions. With the distortion relation network (DRN), the features of all samples with the same distortion are clustered in the same region/cluster, which reveals that the DRN can successfully capture/model the characteristics of each distortion. Notably, despite the target distortion not being observed during the DRN training, it can still be clustered/categorized well in the feature space.

To better understand the learned relations between different auxiliary distortions and target distortion (i.e., RealSR), we further visualize the adjacent similarity matrix of different distortion nodes learned by our distortion relation graph in Fig. 4b. According to this adjacent matrix, we can see that the target distortion is more similar (with a similarity of 0.68) to the mixed mild distortion. Moreover, we observe that the

Table 1. Methods of generating auxiliary distortions.

Distortion types	Generation
Bicubic	Bicubic with scale 8
Ani_bic	Bicubic with scale 4 + anistropic blur
Noise	Gassian noise with σ from the range of [0, 50]
Blur	Gaussin blur with σ from the range of [0, 5]
Mild	Gaussian noise + Gaussian blur + JPEG artifacts; Distortion level lies in the range of [9, 11]
Moderate	Gaussian noise + Gaussian blur + JPEG artifacts; Distortion level lies in the range of [12, 17]
Severe	Gaussian noise + Gaussian blur + JPEG artifacts; Distortion level lies in the range of [18, 20]

To save space, the distortion types are abbreviated as follows: bicubic downsampling (bicubic), bicubic downsampling with anisotropic blurring (ani_bic), Gaussian noise (noise), Gaussian blur (blur), mixed mild (mild), mixed moderate (moderate), mixed severe (severe).
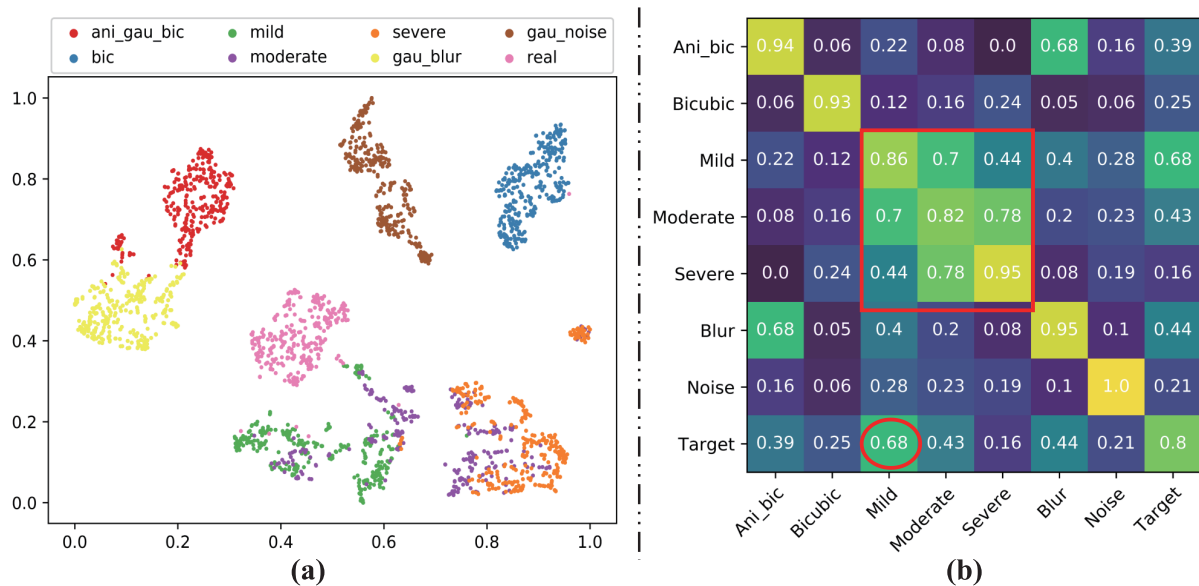

Fig. 4. (a) Visualization of nodes in the distortion relation graph. (b) Visualization of the adjacent matrix in the distortion relation graph. The Target refers to the distortion in the RealSR task.

mixed mild, moderate, and severe distortions are similar/closed since these distortions contain the same three distortion types with different levels, which reveals that our proposed distortion relation graph could effectively capture/model the relations of different kinds of distortions.

4.4 Effectiveness of leveraging the distortion relation

To validate the correlation between the distortion relation and the transferable knowledge in the auxiliary distortions, we measure the transferability of each auxiliary distortion by individually utilizing each auxiliary distortion as the pretraining task. Then, we can measure its corresponding transferability with the performance of the target RealSR (i.e., PSNR). The experimental results are shown in Fig. 5a, where we can conclude that with the increasing distortion relation, the auxiliary distortion contains more powerful transferable knowledge for the target RealSR. This also reveals the effectiveness of our distortion relation graph for the measurement of transferability.

4.5 Comparison with state-of-the-art methods

Comparison with the transfer learning-based method. In this section, we compare our proposed DRTL with the two transfer learning-based methods in Table 2, including the pretraining[69] and MAML[32] schemes. The baseline denotes that the model is trained directly on the few-shot target distortion dataset. Since our proposed DRTL is a model-agnostic optimization strategy, we select five general models, including, DnCNN[15], VDSR[41], RCAN[42], RDN[76], and SwinIR[4], as the backbones for evaluation. Specifically, we compare the training schemes of the baseline, pretraining, MAML, pretraining-based DRTL (DRTL$_p$), and MAML-based DRTL (DRTL$_m$) methods on these backbones. As shown in Table 2, for all five backbone models, the proposed DRTL-related schemes (DRTL$_p$ and DRTL$_m$) achieve the best performance on the real-world distorted test set, which indicates the effectiveness and generalizability of our DRTL. Compared with the baseline scheme, the DRTL-enabled schemes can stably achieve gains of nearly 0.3−0.5 dB, which further reveals the effectiveness of our DRTL for few-shot image super-resolution. Moreover, for the simple-structured DnCNN, we find that it hardly converges well on the few-shot training clean-distorted real image pairs, and only achieves 29.297 dB. In contrast, when applying our DRTL to DnCNN, this scheme can achieve a large performance improvement (31.1 dB in PSNR). We also observed a significant performance drop of DnCNN with MAML, which might be caused by the instability of second-order parameter updating in the MAML
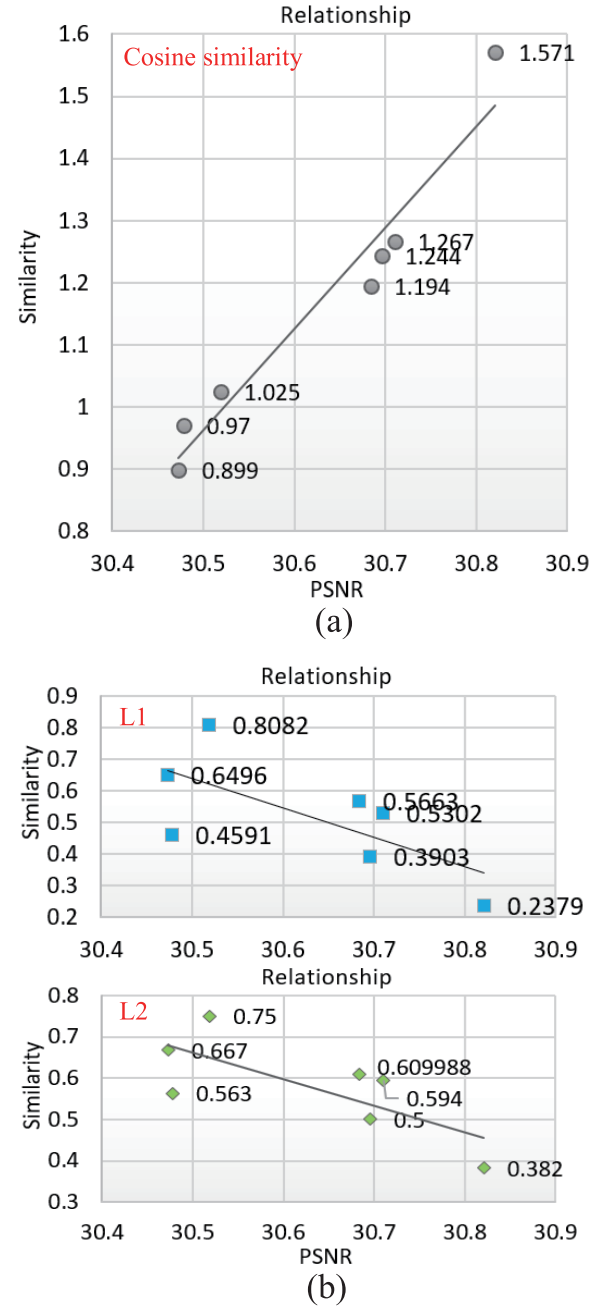

(a)

(b)

Fig. 5. Visualization of the relationship between the similarity of distortions and transferability. Here, the transferability is measured with PSNR. (a) Modeling the similarity with cosine similarity, which is adopted in our paper. (b) Modeling the similarity with L1 or L2 loss.

Table 2. Quantitative comparisons of our proposed DRTL, pretraining, MAML, and baseline on the testing dataset of target distortion.

Models	Baseline		Pretraining		MAML		DRTL$_p$		DRTL$_m$	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
DnCNN[15]	29.297	0.8738	30.858	0.8861	23.466	0.8060	30.977	0.8892	**31.101**	**0.8908**
VDSR[41]	31.067	0.8849	31.290	0.8901	31.219	0.8877	31.358	**0.8908**	**31.367**	0.8907
RCAN[42]	31.299	0.8933	31.698	0.8977	31.546	0.8946	**31.811**	**0.8988**	31.597	0.8956
RDN[76]	31.235	0.8914	31.535	0.8948	31.392	0.8926	**31.655**	**0.8965**	31.493	0.8940
SwinIR[4]	31.336	0.8897	31.526	0.8956	31.402	0.8923	**31.785**	**0.8986**	31.447	0.8924

Baseline methods refer to direct training with few-shot clean-distorted image pairs without transfer learning. The best PSNR and SSIM are bolded.

algorithm.

Moreover, as shown in Table 2, with our selected seven distortions based on our distortion relation, the pretraining and MAML schemes both achieve obvious gains in comparison with the baseline. However, they both ignore exploring how to better utilize the distortion relation for knowledge transfer. Owing to simple gradient modulation with distortion relation guidance, our DRTL-enabled schemes could achieve extra obvious gains compared with the pretraining/MAML schemes when different backbones are used.

In terms of subjective comparison, as shown in Fig. 6, our

method can restore more texture details for real-world distorted images, e.g., the buildings in the first row and the windows in the second row. We analyze that because our DRTL optimization strategy could transfer more valuable knowledge with distortion relation guidance from the auxiliary distortions to the target RealSR.

Comparison with other methods. We also select the unsupervised method DIP[27], and zero-shot method Meta-ZSSR[36] for comparison. For the distortion-augmented method, we compare our method with the famous Real-ESRGAN[29]. As shown in Table 3, DIP cannot work very
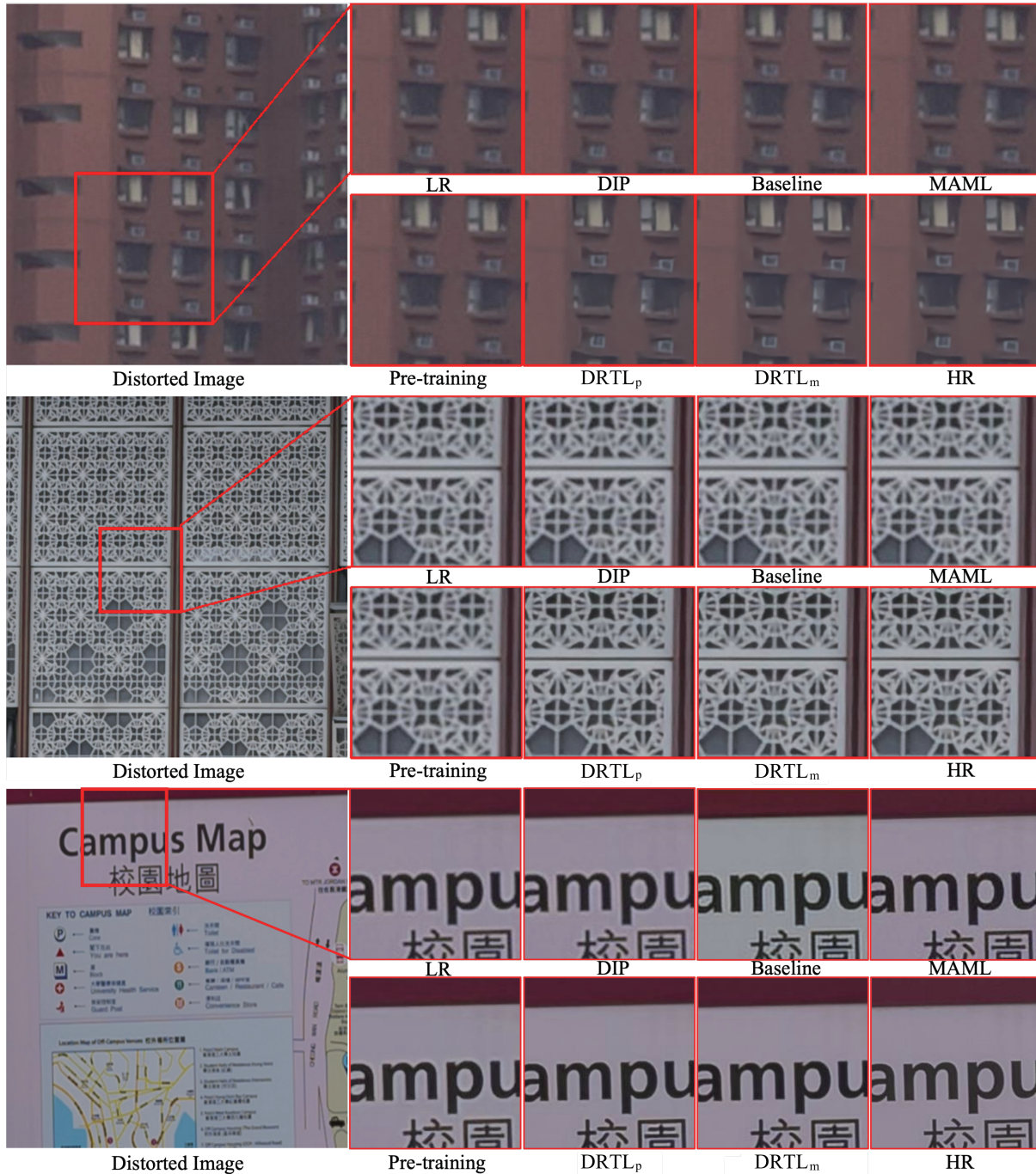

Fig. 6. Subjective comparison of our DRTL with the state-of-the-art methods. Here, DRTL$_p$ and DRTL$_m$ denote our DRTL is integrated into Pretraining and MAML.

Table 3. Quantitative comparison of our DRTL with unsupervised, zero-shot, and distortion-augmented methods.

Methods	DIP[27]	Meta-ZSSR[36]	Real-ESRGAN[29]	DRTL
PSNR	30.32	29.68	28.25	**31.81**
SSIM	0.870	0.856	0.860	**0.899**
LPIPS	0.197	0.168	**0.141**	0.143

The best PSNR, SSIM, and LPIPS are bolded.

well when processing real-world distortions. Real-ESRGAN[29] achieves the best performance in the LPIPS, while it achieves the worst performance in terms of the PSNR and SSIM since fake texture generation. Our DRTL on few-shot RealSR can effectively avoid this problem.

4.6 Ablation study

Study on the knowledge memory bank and similarity metric choice. In our DRTL, we introduce a knowledge memory bank to better store the distortion priors. To validate its effectiveness and necessity, we remove the design from our distortion relation network and evaluate the performance of this scheme on the MAML framework. As shown in Table 4, the performance of the scheme w/o memory bank degrades since the correlation between different distortions cannot be captured effectively/accurately. We also attempt to replace the cosine similarity with the L1 or L2 distance to compute the edges of the distortion relation graph. However, as shown in Fig. 5, the relation coefficients do not satisfy a linear trend as the cosine similarity does.
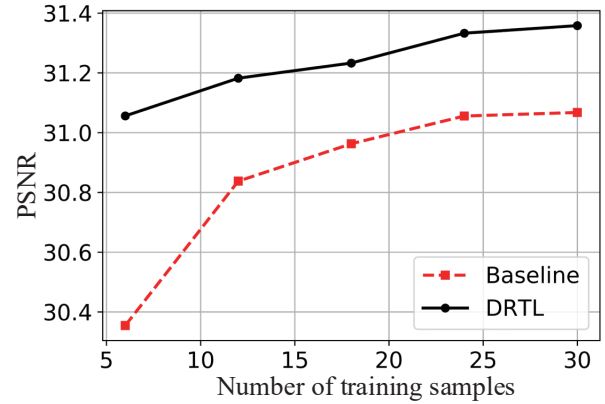
Influence of the number of real training data. To study the influence of the number of real distorted/clean image training pairs w.r.t to the final performance, we set several cases with 5, 10, 15, 20, 25, and 30 few-shot real image pairs for comparison. As shown in Fig. 7, as the number of samples decreases, the performance of the Baseline scheme quickly degrades. In contrast, our DRTL scheme decreases only slightly. Moreover, our DRTL method can achieve more gains when there are fewer samples, which further reveals the superiority of our DRTL method under the more challenging few-shot settings.

5 Conclusions

In this paper, we are the first to take a close look at the challenging few-shot RealSR problem. Since the real clean-distorted image pairs are difficult to collect, we propose to transfer the task-relevant distortion knowledge from auxiliary synthetic distortions to real-world SR. However, synthetic distortions and real distortions have a large distribution gap. Naive transfer learning cannot be adaptively optimized with

Table 4. Quantitative comparison of our proposed DRTL with/without memory bank, which is based on DRTL$_m$.

Methods	PSNR	SSIM	LPIPS
w memory bank	31.3670	0.8907	0.1476
w/o memory bank	31.2396	0.8879	0.1511

Fig. 7. Influence of the number of real training data.

the distortion relations. Therefore, we propose the distortion relation graph with a prior knowledge memory bank to model the dependencies of different synthetic distortions and real-world degradation in RealSR. Based on the distortion relation graph, we can select the most relevant auxiliary distortions for the target RealSR. Moreover, we propose a distortion relation-guided transfer learning (DRTL) framework with gradient reweighting for few-shot RealSR. Extensive experiments on few-shot RealSR have validated the effectiveness of our DRTL.

Acknowledgements

Conflict of interest

The authors declare that they have no conflict of interest.

Biographies

Xin Li is a postdoctoral researcher at the University of Science and Technology of China (USTC). He received his B.S. and Ph.D. degrees from USTC in 2019 and 2024, respectively. His research mainly focuses on image/video compression, restoration, quality assessment, AIGC, and AI agents.

Xin Jin is a tenure-track Assistant Professor at the Eastern Institute of Technology (EIT). He received his Ph.D. degree from the University of Science and Technology of China. His research mainly focuses on computer vision, intelligent media computing, and deep learning.

Zhibo Chen is a Professor at the University of Science and Technology of China. He received the B.S. and Ph.D. degrees from Tsinghua University in 1998 and 2003, respectively. His research mainly focuses on investigating artificial intelligence techniques for advanced visual signal generation, representation, processing and coding, as well as in other interdisciplinary research fields.

References

[1] Kupyn O, Martyniuk T, Wu J, et al. DeblurGAN-v2: Deblurring

(orders-of-magnitude) faster and better. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, **2019**: 8877–8886.

[2] Moran N, Schmidt D, Zhong Y, et al. Noisier2Noise: Learning to denoise from unpaired noisy data. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, **2020**: 12061–12069.

[3] Xia B, Zhang Y, Wang S, et al. DiffIR: Efficient diffusion model for image restoration. arXiv: 2303.09472, **2023**.

[4] Liang J, Cao J, Sun G, et al. SwinIR: Image restoration using Swin Transformer. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada: IEEE, **2021**: 1833–1844.

[5] Conde M V, Choi U J, Burchi M, et al. Swin2SR: SwinV2 transformer for compressed image super-resolution and restoration. In: Computer vision–ECCV 2022 workshops. Cham, Switzerland: Springer, **2023**: 669–687.

[6] Wang Z, Cun X, Bao J, et al. Uformer: A general U-shaped transformer for image restoration. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, **2022**: 17662–17672.

[7] Deng S, Wei M, Wang J, et al. Detail-recovery image deraining via context aggregation networks. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, **2020**: 14548–14557.

[8] Wang T, Yang X, Xu K, et al. Spatial attentive single-image deraining with a high quality real rain dataset. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, **2019**: 12262–12271.

[9] Li X, Li B, Jin X, et al. Learning distortion invariant representation for image restoration from a causality perspective. In: 2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). IEEE, **2023:** 1714–1724.

[10] Li X, Jin X, Lin J, et al. Learning disentangled feature representation for hybrid-distorted image restoration. In: Vedaldi A, Bischof H, Brox T, et al., editors. Computer Vision–ECCV 2020. Cham, Switzerland: Springer, **2020**: 313–329.

[11] Liu J, Lin J, Li X, et al. Lira: Lifelong image restoration from unknown blended distortions. In: Vedaldi A, Bischof H, Brox T, et al., editors. Computer Vision – ECCV 2020. Cham, Switzerland: Springer, **2020**: 616–632.

[12] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2016**, *38*: 295–307.

[13] Ren Y, Li X, Guo M, et al. MambaCSR: Dual-interleaved scanning for compressed image super-resolution with SSMs. arXiv: 2408.11758, **2024**.

[14] Tao X, Gao H, Shen X, et al. Scale-recurrent network for deep image deblurring. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, **2018**: 8174–8182.

[15] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising. *IEEE Transactions on Image Processing*, **2017**, *26*: 3142–3155.

[16] Yu K, Dong C, Loy C C, et al. Deep convolution networks for compression artifacts reduction. arXiv: 1608.02778, **2016**.

[17] Li X, Sun S, Zhang Z, et al. Multi-scale grouped dense network for VVC intra coding. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, **2020**: 615-618.

[18] Li B, Xin Li, Lu Y, et al. HST: Hierarchical swin transformer for compressed image super-resolution. In: Computer Vision–ECCV 2022 Workshops. New York: ACM, **2023**: 651–668.

[19] Agustsson E, Timofte R. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA: IEEE, **2017**: 1122–1131.

[20] Timofte R, Agustsson E, Van Gool L, et al. NTIRE 2017 challenge on single image super-resolution: methods and results. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA: IEEE, **2017**: 1110–1121.

[21] Nah S, Kim T H, Lee K M. Deep multi-scale convolutional neural network for dynamic scene deblurring. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Honolulu, HI, USA: IEEE, **2017**: 257–265.

[22] Zhang H, Patel V M. Density-aware single image de-raining using a multi-stream dense network. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, **2018**: 695–704.

[23] Chen H, Wang Y, Guo T, et al. Pre-trained image processing transformer. arXiv: 2012.00364, **2020**.

[24] Dai T, Cai J, Zhang Y, et al. Second-order attention network for single image super-resolution. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, **2019**: 11057–11066.

[25] Yasarla R, Sindagi V A, Patel V M. Syn2Real transfer learning for image deraining using Gaussian processes. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, **2020**: 2723–2733.

[26] Li X, Jin X, Yu T, et al. Learning omni-frequency region-adaptive representations for real image super-resolution. *Proceedings of the AAAI Conference on Artificial Intelligence*, **2021**, *35*: 1975–1983.

[27] Lempitsky V, Vedaldi A, Ulyanov D. Deep image prior. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, **2018**: 9446–9454.

[28] Du W, Chen H, Yang H. Learning invariant representation for unsupervised image restoration. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, **2020**: 14471–14480.

[29] Wang X, Xie L, Dong C, et al. Real-ESRGAN: Training real-world blind super-resolution with pure synthetic data. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada: IEEE, **2021**: 1905–1914.

[30] Fei B, Lyu Z, Pan L, et al. Generative diffusion prior for unified image restoration and enhancement, arXiv: 2304.01247, **2023**.

[31] Chen Y, Liu Z, Xu H, et al. Meta-baseline: Exploring simple meta-learning for few-shot learning. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, **2021**: 9042–9051.

[32] Finn C, Abbeel P, Levine S. Model-agnostic meta-learning for fast adaptation of deep networks. In: Proceedings of the 34th International Conference on Machine Learning–Volume 70. New York: ACM, **2017**: 1126–1135.

[33] Snell J, Swersky K, Zemel R S. Prototypical networks for few-shot learning. arXiv: 1703.05175, **2017**.

[34] Zhou K, Yang J, Loy C C, et al. Conditional prompt learning for vision-language models. In: 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). New Orleans, LA, USA: IEEE, **2022**: 16795–16804.

[35] Lu Z, He S, Zhu X, et al. Simpler is better: Few-shot semantic segmentation with classifier weight transformer. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). IEEE, **2021**: 8721–8730.

[36] Soh J W, Cho S, Cho N I. Meta-transfer learning for zero-shot super-resolution. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, **2020**: 3513–3522.

[37] Krishna K, Narasimha Murty M. Genetic K-means algorithm. *IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics)*, **1999**, *29*: 433–439.

[38] Kipf T N, Welling M. Semi-supervised classification with graph convolutional networks. arXiv: 1609.02907, **2016**.

[39] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network. In: Computer Vision–ECCV 2016. Cham, Switzerland: Springer, **2016**: 391–407.

[40] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Honolulu, HI, USA: IEEE, **2017**: 1132–1140.

[41] Kim J, Lee J K, Lee K M. Accurate image super-resolution using very deep convolutional networks. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). Las Vegas, NV, USA: IEEE, **2016**: 1646–1654.

[42] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks. In: Ferrari V, Hebert M, Sminchisescu C, et al. editors. Computer Vision–ECCV 2018. Cham, Switzerland: Springer, **2018**: 294–310.

[43] Castillo A, Escobar M, Pérez J C, et al. Generalized real-world super-resolution through adversarial robustness. In: 2021 IEEE/CVF International Conference on Computer Vision Workshops (ICCVW). Montreal, BC, Canada: IEEE, **2021**: 1855–1865.

[44] Muhammad Umer R, Luca Foresti G, Micheloni C. Deep generative adversarial residual convolutional networks for real-world super-resolution. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Seattle, WA, USA: IEEE, **2020**: 1769–1777.

[45] Zhu J Y, Park T, Isola P, et al. Unpaired image-to-image translation using cycle-consistent adversarial networks. In: 2017 IEEE International Conference on Computer Vision (ICCV). Venice, Italy: IEEE, **2017**: 2242–2251.

[46] Tao G, Ji X, Wang W, et al. Spectrum-to-kernel translation for accurate blind image super-resolution. In: Proceedings of the 35th International Conference on Neural Information Processing Systems. New York: ACM, **2021**: 22643–22654.

[47] Zhang K, Liang J, Van Gool L, et al. Designing a practical degradation model for deep blind image super-resolution. In: 2021 IEEE/CVF International Conference on Computer Vision (ICCV). Montreal, QC, Canada: IEEE, **2021**: 4771–4780.

[48] Cai J, Zeng H, Yong H, et al. Toward real-world single image super-resolution: A new benchmark and a new model. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, **2019**: 3086–3095.

[49] Wei P, Xie Z, Lu H, et al. Component divide-and-conquer for real-world image super-resolution. In: Vedaldi A, Bischof H, Brox T, et al., editors. Computer Vision–ECCV 2020. Cham, Switzerland: Springer, **2020**: 101–117.

[50] Pang Y, Xin Li, Jin X, et al. FAN: Frequency aggregation network for real image super-resolution. In: Bartoli A, Fusiello A, editors. Computer Vision–ECCV 2020 Workshops. Cham, Switzerland: Springer, **2020**: 468–483.

[51] Wei P, Lu H, Timofte R, et al. AIM 2020 challenge on real image super-resolution: Methods and results. In: Bartoli A, Fusiello A, editors. Computer Vision–ECCV 2020 Workshops. Cham, Switzerland: Springer, **2020**: 392–422.

[52] Yuan Y, Liu S, Zhang J, et al. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW). Salt Lake City, UT, USA: IEEE, **2018**: 814–81409.

[53] Kim Y, Soh J W, Park G Y, et al. Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In: 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Seattle, WA, USA: IEEE, **2020**: 3479–3489.

[54] Wei W, Meng D, Zhao Q, et al. Semi-supervised transfer learning for image rain removal. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, **2019**: 3872–3881.

[55] Park S, Yoo J, Cho D, et al. Fast adaptation to super-resolution networks via meta-learning. In: Vedaldi A, Bischof H, Brox T, et al. editors. Computer Vision–ECCV 2020. Cham, Switzerland: Springer, **2020**: 754–769.

[56] Yang S, Wang Y, Wang K, et al. Local prediction aggregation: A frustratingly easy source-free domain adaptation method. arXiv: 2205.04183, **2022**.

[57] Ling X, Dai W, Xue G R, et al. Spectral domain-transfer learning. In: Proceedings of the 14th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining. New York: ACM, **2008**: 488–496.

[58] Long H, Geng R, Sheng W, et al. Small-sample solar power interval prediction based on instance-based transfer learning. *IEEE Transactions on Industry Applications*, **2023**, *59*: 5283–5292.

[59] Ganin Y, Ustinova E, Ajakan H, et al. Domain-adversarial training of neural networks. *Journal of Machine Learning Research*, **2016**, *17* (1): 2096–2030.

[60] Sun J, Qin J, Lin Z, et al. Prompt tuning based adapter for vision-language model adaption. arXiv: 2303.15234, **2023**.

[61] Fu M, Wang X, Wang J, et al. Prototype Bayesian meta-learning for few-shot image classification. *IEEE Transactions on Neural Networks and Learning Systems*, **2024**, *36*: 7010–7024.

[62] Radford A, Kim J W, Hallacy C, et al. Learning transferable visual models from natural language supervision. In: Proceedings of the 38th International Conference on Machine Learning. PMLR, **2021**: 8748–8763.

[63] Yang X, Cheng W, Zhao X, et al. Dynamic prompting: A unified framework for prompt tuning. arXiv: 2303.02909, **2023**.

[64] Khattak M U, Syed Talal Wasim S T, Naseer M, et al. Self-regulating prompts: Foundational model adaptation without forgetting. In: 2023 IEEE/CVF International Conference on Computer Vision (ICCV). Paris, France: IEEE, **2023**: 15144–15154.

[65] Li X, Li B, Jin Y, et al. UCIP: A universal framework for compressed image super-resolution using dynamic prompt. In: Leonardis A, Ricci E, Roth S, et al. editors. Computer Vision–ECCV 2024. Cham, Switzerland: Springer, **2024**: 107–125.

[66] Gao P, Geng S, Zhang R, et al. Clip-adapter: Better vision-language models with feature adapters. arXiv: 2110.04544, **2021**.

[67] Zhang R, Zhang W, Fang R, et al. Tip-adapter: Training-free adaption of clip for few-shot classification. In: Avidan S, Brostow G, Cissé M, et al. editors. Computer Vision–ECCV 2022. Cham, Switzerland: Springer, **2022**: 493–510.

[68] Li X, Lian D, Lu Z, et al. Graphadapter: Tuning vision-language models with dual knowledge graph. arXiv: 2309.13625, **2023**.

[69] He K, Girshick R, Dollar P. Rethinking ImageNet pre-training. In: 2019 IEEE/CVF International Conference on Computer Vision (ICCV). Seoul, Korea: IEEE, 2019: 4917–4926.

[70] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, 2018: 586–595.

[71] Ding K, Ma K, Wang S, et al. Image quality assessment: Unifying structure and texture similarity. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, **2022**, *44*: 2567–2581.

[72] Simonyan K, Zisserman A. Very deep convolutional networks for large-scale image recognition. arXiv: 1409.1556, **2014**.

[73] Hamilton W, Ying Z, Leskovec J. Inductive representation learning on large graphs. In: NIPS'17: Proceedings of the 31st International Conference on Neural Information Processing Systems. New York: ACM, **2017**: 1025–1035.

[74] Suganuma M, Liu X, Okatani T. Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In: 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR). Long Beach, CA, USA: IEEE, **2019**: 9031–9040.

[75] Yu K, Dong C, Lin L, et al. Crafting a toolchain for image restoration by deep reinforcement learning. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, **2018**: 2443–2452.

[76] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution. In: 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition. Salt Lake City, UT, USA: IEEE, **2018**: 2472–2481.