Few-Shot Real Image Super-resolution via Distortion-Relation Guided Transfer Learning

Xin Li\textsuperscript{1\ast}, Xin Jin\textsuperscript{1\ast}, Jun Fu\textsuperscript{1}, Xiaoyuan Yu\textsuperscript{2}, Bei Tong\textsuperscript{2}, and Zhibo Chen\textsuperscript{1\dagger}

\textsuperscript{1} University of Science and Technology of China
\textsuperscript{2} Huawei Cloud

\{lixin666, jinxustc, fujun\}@mail.ustc.edu.cn
\{yuxiaoyuan, tongbei\}@huawei.com chenzhibo@ustc.edu.cn

Abstract. Collecting large clean-distorted training image pairs in real world is non-trivial, which seriously limits the practical applications of these supervised learning based image super-resolution (SR) methods. Previous works attempt to address this problem by leveraging unsupervised learning technologies to alleviate the dependency for paired training samples. However, these methods typically suffer from unsatisfactory textures synthesis due to the lack of clean image supervision. Compared with purely unsupervised solution, the under-explored scheme with Few-Shot clean images (FS-RSR) is more feasible to tackle this challenging Real Image Super-Resolution task. In this paper, we are the first to investigate the few-shot real image super-resolution and propose a Distortion-Relation guided Transfer Learning (termed as DRTL) framework. DRTL assigns a knowledge graph to capture the distortion relation between auxiliary tasks (\textit{i.e.}, synthetic distortions) and target tasks (\textit{i.e.}, real distortions with few images), and then adopt a gradient weighting strategy to guide the knowledge transfer from auxiliary task to target task. In this way, DRTL could quickly learn the most relevant knowledge from the prior distortions for target distortion. We instantiate DRTL integrated with pre-training and meta-learning pipelines as an embodiment to realize a distortion-relation aware FS-RSR. Extensive experiments on multiple benchmarks demonstrate the effectiveness of DRTL on few-shot real image super-resolution.

Keywords: Few-shot RealSR, Distortion Relation Graph, Transfer Learning.

1 Introduction

Image restoration (IR) task aims to restore high-quality images from the degraded low-quality counterparts. The real-world degradation process typically consists of a series of different distortions, such as motion blur \cite{38, 22}, noise

\textsuperscript{\ast}Equal Contribution
\textsuperscript{\dagger}Corresponding Author
Distortion-Relation guided Transfer-Learning (DRTL). We introduce a distortion-relation graph to capture the relationship (i.e., 0.44, 0.21 and 0.12 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.

Although the fast-developed deep learning has significantly promoted the advancement of image restoration (IR) techniques, the success of these learning-based IR methods usually rely on training with a large-scale dataset. As a result, numerous image restoration datasets that contains various distortions have been proposed and deeply explored/studied, such as DIV2K [1], Gopro [30], and DID-MDN [52]. However, the degradations of these datasets are typically synthetic, or said, hand-craft, the distortions of which are different from the real-world ones. Real image super-resolution (i.e., RSR) is the typical case of real image restoration, where low-resolution (LR) images suffer from severe hybrid distortions, such as blur, noise and compression artifacts.

Although some learning-based super-resolution (SR) approaches [47,55] trained on the above-mentioned synthetic datasets have achieved a great success when handling the synthetic degradations (i.e., bicubic degradation), those image SR systems hardly perform well on the real-world scenarios due to the large gap between synthetic distortions and real-world distortions [47,48,24]. Furthermore, re-collecting large clean-distorted image training pairs from real-world is also non-trivial [23]. The above-mentioned two disgusting weaknesses of the existing image SR algorithms seriously limit their practical applications and hurt their industrial values. To alleviate the dependency for clean distortion/clean image training pairs, some studies [39,12,44] introduce unsupervised learning to
real image restoration. However, these methods typically suffer from unsatisfactory/spurious textures synthesis due to the lack of clean image supervision. Compared with purely unsupervised solution, the under-explored SR scheme with Few-Shot distortion/clean image pairs image (FS-RSR) is more feasible to tackle this challenging Real Image Super-Resolution.

Previous works have explored different transfer learning strategies [6,33,14,35] for few-shot problems in classification task, extracting helpful knowledge from auxiliary tasks to target task. Among them, Pre-training and Meta-learning are two representative technologies, which have been also explored in image super-resolution. For instance, IPT [5] introduces a large-scale pre-training dataset to improve the restoration performance w.r.t the target distortion. Soh et al. [36] propose a meta-learning based method to implement the fast adaptation for zero-shot super-resolution task, achieving a SOTA performance. However, these transfer learning based methods still focus on handling the synthetic distortions, which is not suitable for a more challenging few-shot real-world image super-resolution problem.

In this paper, we are the first to focus on the challenging few-shot real-world image super-resolution (FS-RSR) problem, where the auxiliary tasks (i.e., synthetic distortions) and target task (i.e., real super-resolution) are different and exist a large gap. Therefore, we face the key issue needed be solved: “how to capture the proper prior knowledges for target distortion from auxiliary distortions?” Different from high-level classification task, of which the relation between different tasks can be modelling with a simple clustering of KNN [21]. To better capture the relationship between different distortions, we propose to learn a distortion-relation graph, which consists of two essential factors: Nodes (i.e., distortion embedding) and Edges (i.e., the similarity between distortions) as shown in Figure 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, we design a prior knowledge memory bank to storage the learnable distortion-relation from seen auxiliary tasks. Given an arbitrary real distorted sample, it can traverse the prior knowledge memory bank to acquire the needed distortion embeddings as in GCN [20].

After obtaining the needed distortion embeddings, we compute the edges (that represent the distortion similarity) between different distortions with cosine similarity. Since the different auxiliary distorted samples contain different degrees of useful priors for target real distortion, we propose a Distortion-Relation guided Transfer Learning (DRTL). We instantiate our DRTL integrated with pre-training and meta-learning framework. Specifically, we integrate the distortion similarity that is inferenced out from distortion-graph into the optimization loop of pre-training/MAML [15] with gradient weighting. Extensive experiments on multiple benchmarks have demonstrated the effectiveness of our DRTL on few-shot real image super-resolution.

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

- To our knowledge, we are the first to focus on the challenging few-shot real-world image super-resolution (FS-RSR) problem.
— We introduce a distortion graph to guide the few-shot real image super-resolution from the perspective of exploring the relationship between synthetic distortions and real-world distortion.
— Based on the learned distortion knowledge graph, we propose a knowledge transfer/training strategy embedded with distortion-relation prior guidance for FS-RSR, i.e., Distortion-Relation guided Transfer Learning (DRTL).

We conduct extensive experiments based on different baseline schemes, e.g., pre-training based scheme and meta-learning based scheme. The experimental results on FS-RSR datasets with few-shot settings have demonstrated the effectiveness of our DRTL. DRTL is simple yet effective and can be used as a general FS-RSR framework that is compatible with many existing SR networks. We will release our source code upon acceptance.

2 Related Work

2.1 Image Super-Resolution

Deep learning has accelerated the development of image super-resolution techniques since the pioneer works SRCNN [9] and FSRCNN [10]. Most works [27,32,7,18,54,4,26] are only devoted into the synthetic degradation (i.e., bicubic downsampling). However, in real-world scenario, the degradation factors are composed of hybrid distortions, such as blur, noise and compression artifacts, etc. To tackle the challenge of real-world super-resolution, a series of works [28,49,3,41] apply image translation technology and cycle consistence [56] to implement unsupervised image super-resolution, which have achieved great subjective quality. However, these methods typically suffer from unsatisfactory textures synthesis due to the lack of clean image supervision. Meanwhile, some real image super-resolution (RSR) datasets with limited real-world clean/distorted image pairs have been collected and released (i.e., RealSR [2] and DRealSR [46]), which are costly from the perspective of time and manpower. Despite some frameworks [2,46,11,45,43] have achieved great progressive in RealSR, they ignore the fact that limited data in real world prevent their further improvements. Different from the above works, we are the first to investigate the few-shot real image super-resolution (FS-RSR), which is vital for the application of SR methods in real world.

2.2 Real Image Restoration

Unsupervised Real Image Restoration. Collecting a large-scale clean/distorted training dataset in the real world is non-trivial, which also makes these fully-supervised IR methods failed. To reduce the real-world data dependencies of image restoration, Ulyanov et al. [40] propose to utilize the structure of CNN to capture the deep image statistics (named as deep image prior, DIP) through an iterative self-supervised optimization. However, when to stop DIP optimization is hard to decided for the different distorted images. Du et al. [13] introduce a
discrete disentangling representation learning method to capture the invariant clean representations from unpaired clean-distorted image pairs. However, this method only focuses on handling the noise distortions and ignores other complex distortions that are widely-existed in the real-world scenarios.

**Transfer Learning based Real Image Restoration.** Since the real-world distortion datasets are difficult to collect, some studies intend to leverage the transfer learning techniques to achieve image restoration for real distorted images. For example, Wei et al. [47, 48] propose to capture the distortion priors of rain streaks from the auxiliary synthetic/fake rain streaks to achieve the clean image restoration in a semi-supervised learning manner. However, above methods are designed based on an assumption that rain streaks can be modeled with Gaussian distribution, which is not always satisfied for other distortions. With the advancement of transfer learning, Soh et al. [36] and Park et al. [31] propose to leverage the Meta-transfer learning to deal with the challenging task of zero-shot super-resolution (ZSSR). Kim et al. [19] utilize the adaptive instance normalization to realize the knowledge transfer from synthetic noise to real noise.

Different from the above works, investigating the knowledge transfer between the homogeneous distortions (*e.g.*, noise to noise, rain to rain, etc.), our DRTL focus on more challenge setting, where the auxiliary tasks can be heterogeneous synthetic distortions (*e.g.*, the jpeg artifacts to RealSR, mixed distortions to RealSR, etc.).

### 3 Distortion-Relation Guided Transfer Learning (DRTL) Framework

#### 3.1 Knowledge Preliminary

**Pre-training based Transfer Learning.** As the basic technology of transfer learning, pre-training has been widely applied to different vision tasks [5, 17]. Pre-training based transfer learning can be divided into two processes, respectively as pre-training and fin-tuning. Given the auxiliary tasks $\mathcal{T}^a$, target task $\mathcal{T}^t$ and a learning model $f_\theta$ with parameters $\theta$, pre-training typically first update the model parameters on auxiliary tasks $\mathcal{T}^a$ as Eq. 1 to obtain the task-relevant knowledge,

$$\theta' = \theta - \alpha \nabla_\theta L_{\mathcal{T}^a_i}(f_\theta),$$  \hspace{1cm} (1)

where the $L_{\mathcal{T}^a_i}(f_\theta)$ refers to the $L_1$ loss in $\mathcal{T}^a_i$. The optimization objective of pre-training is to minimize the loss function in all auxiliary tasks as Eq. 2,

$$\min_\theta \sum_{\mathcal{T}^a_i \sim p(\mathcal{T}^a)} L_{\mathcal{T}^a_i}(f_\theta)$$  \hspace{1cm} (2)
In the fine-tuning stage, the best parameter $\theta_m$ in pre-training stage is used as an initial parameters. Then model $f_{\theta_m}$ is updated with the optimization objective to acquire the best performance in target task as Eq. 3:

$$\min_{\theta} \mathcal{L}_{T^t}(f_{\theta_m})$$  \hspace{1cm} (3)

**MAML-based Transfer Learning.** Different from pre-training based transfer-learning, which directly learn the task-relevant knowledge from auxiliary tasks. MAML aims to learn ability of fast adaptation to all auxiliary tasks [15]. Therefore, MAML can be divided into two processes, respectively as Meta-train and Meta-test. For Meta-train, the model first update in one task $\mathcal{L}_{T^a_i}$, which is randomly sampled from auxiliary tasks $p(T^a)$ as Eq. 1. And then the model are optimized across tasks sampled from $T^a$ with the meta-objective as follows:

$$\min_{\theta} \sum_{T^a_i \sim p(T^a)} \mathcal{L}_{T^a_i}(f_{\theta - \alpha \nabla_{\theta} \mathcal{L}_{T^a_i}(f_{\theta})})$$ \hspace{1cm} (4)

After getting the gradient from above optimization, the parameter $\theta$ can be updated with Eq. 5

$$\theta_m = \theta - \beta \nabla_{\theta} \sum_{T^a_i \sim p(T^a)} \mathcal{L}_{T^a_i}(f_{\theta_i}).$$ \hspace{1cm} (5)

As described in Eq. 5, Meta-Train aims to learn the general $\theta$, which can be transferred to all tasks well. The Meta-Test stage is the same as fine-tuning, which can be represented as Eq. 3

### 3.2 Building Distortion-Relation Graph

The distortion-relation graph aims to explore and store the relationship between multiple synthetic distortions and target real-world distortion. Once getting such relation graph, the most relevant auxiliary distortions can be selected to assist target real SR. Moreover, we could leverage it to re-weight/guide the knowledge transfer process from auxiliary distortions to target distortion. Specifically, we propose a prototype-based distortion relation graph, where the nodes in the graph denote the relationship embeddings between each sample and prior knowledge memory bank. The edges are created based on similarity between prototypes. The architecture of this prototype-based distortion relation graph is described in Figure 2, which consists of four key components: feature extractor, prior knowledge memory bank, relation-ware nodes and edges. We will introduce each component in detail in the following sub-sections.

**Feature Extractor** To reduce the interference caused by the raw image texture and structure of distorted image when extracting the distortion features, we first compute the residual $I_{res}$ between clean image $I$ and distorted image $I_{dist}$ as $I_{res.} = (I - I_{dist})/2 + 0.5$. Such operation aims to normalize the residual
Fig. 2. Architecture of our proposed distortion relation network (DRN), which consists of feature extractor (VGG11), prior knowledge memory bank and Graph convolution network. The representation of each distortion sample can be tapped into prior knowledge memory bank to capture the distortion embedding with GCN as the node of distortion relation graph. Then the edges of distortion relation graph can be computed with cosine similarity of different nodes.

information between $[0, 1]$. We propose to utilize the feature extractor of VGG11 [34] to further extract the distortion representation for each distortion as $F_{dist} = VGG(I_{res.})$, and the feature extractor is trained with our distortion-relation graph.

Knowledge Memory Bank Generally, the graph learning aims to capture internal relationship between samples or structures, and then assist the current task [20,16]. In order to achieve our purpose that leveraging the relation graph to re-weight/guide the knowledge transfer, both of the target distortion and auxiliary synthetic distortions must be taken into the same graph as nodes and infer out their relation weights. 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 distortion relation graph. Second, directly establishing the relation graph with the current faced real target distortion may make the graph hard to generalize well to other unseen real distortions. To obtain an universal and general distortion relation graph, we assign a long-term distortion memory bank to record the previous seen distortion types, which consists of several memory nodes $\mathcal{H}_m$. Based on such distortion memory bank, the relationship of auxiliary synthetic distortions can be automatically established and stored into distortion memory graph as the training going on. And, arbitrary distortions can be tapped into the same distortion memory bank to acquire the corresponding representations, which property can be used to establish the intra-relationship among different distortions. Note that, the target distortions and auxiliary distortions can extract the relation embedding independently.

Graph Nodes and Edges. The distortion relation graph $\mathcal{R} = (\mathcal{C}_R, \mathcal{A}_R)$ consists of two essential components, i.e., Nodes $\mathcal{C}_R = \{c^i | \forall i \in [1, K]\} \in \mathbb{R}^{K \times d}$ and
Edges $\mathcal{A}_R = \{ |\mathcal{A}_R(c^i, c^j)| \forall i, j \in [1, K] \} \in \mathbb{R}^{K \times K}$. In this paper, Nodes refer to feature embedding of each distorted sample projected in knowledge memory bank $\mathcal{M} = (\mathcal{H}_M, \mathcal{A}_M)$, which saves the corresponding distortion knowledge in a memory graph format. To generate the relation-aware nodes (i.e., feature embedding), we first compute the relation between the $i$th distortion feature $F_{dist}^i$ with each nodes of memory bank as Eq. 6

$$A_p^i = \sigma(||F_{dist}^i - \mathcal{H}_M||_2^2).$$  \hspace{1cm} (6)$$

Then we concatenate the relation adjacent matrix $A_p^i$ and the adjacent matrix of prior knowledge memory bank $\mathcal{A}_M$,

$$\mathcal{A} = [A_p^i, \mathcal{A}_M],$$ \hspace{1cm} (7)$$

where $\mathcal{A}_M$ can be computed as Eq. 8

$$\mathcal{A}_M = \{ \sigma(||\mathcal{H}_m - \mathcal{H}_n||_2^2) | m, n \in \{1, Q\} \}. \hspace{1cm} (8)$$

Finally, the nodes of distortion relation graph can be generated with graph convolution network (GCN) [20] as,

$$c_i = GCN([F_{dist}^i, \mathcal{H}_M], \mathcal{A}).$$ \hspace{1cm} (9)$$

After obtaining the nodes $C_R = \{ c^i | \forall i \in [1, K] \} \in \mathbb{R}^{K \times d}$ for distortion relation graph, we utilize cosine similarity to measure the distance of different distortion samples, which acting as the edges $\mathcal{A}_R$ of distortion relation graph.

### 3.3 Distortion-Relation guided Transfer-Learning

In order to extract the task-relevant distortion knowledge from auxiliary synthetic distortions, we utilize transfer-learning to assist few-shot real SR. However,
directly applying transfer-learning does not consider the relation between auxiliary distortion and target distortion, and thus cannot efficiently pick out most task-relevant knowledge. Therefore, in this paper, we propose distortion-relation guided transfer-learning (DRTL). The workflow of DRTL is described as Figure 3. Since sufficient and comprehensive relevant auxiliary tasks could further bring the performance improvements for the target real super-resolution, based on our proposed distortion relation graph, we utilize relation coefficient $\gamma$ to guide the transfer learning process in a gradient modulation/re-weighting manner. Here, we utilize two general transfer-learning methods (including MAML and Pre-training) to clarify our DRTL methods. As shown in Figure 3, we obtain the relevance $\gamma_i$ between auxiliary distortion $T_{ad}^i$ and target distortion $T_{td}$ from our proposed distortion relation graph $R$. Then we respectively modify the meta-training of MAML and pre-training process as Eq. 10 and Eq. 11.

In this way, the optimization direction of MAML and pre-training can be closer to target distortion, which can capture most task-relevant knowledge.

$$\theta_m = \theta - \beta \nabla_{\theta} \sum_{T_{ad}^i \sim p(T_{ad})} \gamma_i L_{T_{ad}^i(f_{\theta}^i)}.$$ 

$$\theta'_{i} = \theta - \alpha \nabla_{\theta} \gamma_i L_{T_{ad}^i(f_{\theta})}.$$  

| Models | Baseline | Pre-training | MAML | DRTL(PT) | DRTL(MAML) |
|--------|----------|--------------|-------|----------|------------|
|        | PSNR | SSIM | PSNR | SSIM | PSNR | SSIM |
| DnCNN  | 29.297 | 0.8738 | 30.858 | 0.8861 | 23.466 | 0.8060 |
| VDSR   | 31.067 | 0.8849 | 31.290 | 0.8901 | 31.219 | 0.8877 |
| RCAN   | 31.299 | 0.8933 | 31.698 | 0.8977 | 31.546 | 0.8946 |
| RDN    | 31.235 | 0.8914 | 31.535 | 0.8948 | 31.392 | 0.8926 |

Table 1. Quantitative comparisons of our proposed DRTL, Pre-training, MAML and Baseline on testing dataset of target distortion. Baseline methods refer to directly training with few-shot clean-distorted image pairs without transfer-learning. DRTL(PT) denotes the DRTL with pretraining.

4 Experiments

In this section, we first introduce the auxiliary distortions datasets and target RealSR datasets in Sec. 4.1. Then, we present the implementation details of our DRTL in Sec. 4.2. And, we show a comprehensive comparison and analysis about the component designs of DRTL to demonstrate their effectiveness and superiority in Sec. 4.3, Sec. 4.4, Sec. 4.5, Sec. 4.6.
4.1 Datasets

**Auxiliary datasets.** To demonstrate the effectiveness and robustness of our DRTL, we select most relevant 7 common synthetic distortions as auxiliary tasks based on the our proposed distortion relation graph, including Bicubic downsampling [1], Bicubic downsampling with An-isotropic kernels [36], Gaussian noise, Gaussian blur, Mixed mild distortion, Mixed moderate distortion, Mixed severe distortion [37]. Then, we take 800 clean images of DIV2K dataset [1] as original/basic images, and add above-mentioned distortions into original images to generate 7 auxiliary datasets. The detailed synthesis method of auxiliary distortions can be found in the Supplementary.

![Fig. 4. (a) Visualization of nodes in distortion relation graph and (b) Visualization of adjacent matrix in distortion relation graph.](image)

**Target dataset** We select the typical RealSR [2] dataset as our target task. To get a few-shot real-world dataset for training/fine-tune, we randomly select 30 clean-distorted image pairs from the train set of RealSR [2] as our training dataset with target distortion. For the final evaluation, we further use 30 clean/distorted images in the test set of RealSR [2] 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 [2].

4.2 Implementation Details

Since the proposed DRTL is a model-agnostic optimization strategy, we adapt the general model VDSR [18], that contains a series of non-linear Res-blocks, as our baseline model. The implementation of DRTL is based on Py-
Torch framework. The whole training process can be divided into two stages: 1) Pre-training/Meta-Train. 2) Fin-tuning/Meta-Test. For the first step, we utilize Adam optimizer with an initial learning rate of 0.0001 for optimization. For the Fin-tuning/Meta-Test step, the learning rate decay by a factor of 0.8 each 3000 iterations. We set the batch-size as 32 and leverage random flip, rotation, cropping to achieve data augmentation. The size of cropped image is $64 \times 64$. $L_1$ loss have been proved effective to optimize the model especially in image restoration [54,2]. Therefore, we only use $L_1$ loss to optimize the DRTL in this paper.

4.3 Graph Visualization and Explanation

In this section, we elaborately discuss how our distortion-relation graph works. As shown in Figure 4(a), we visualize the feature embedding for each sample in all auxiliary and target distortions data. With the distortion-relation graph network (DGN), the extracted features of all samples with the same distortion have been clustered in the same region/cluster, which reveals that DGN could successfully capture/model the characteristics of each distortion. Furthermore, despite the target distortion has not been seen during the DGN training, it still can be clustered/categorized well in the feature space. To better understand the

Fig. 5. Subjective comparison of our DRTL with the state-of-the-art methods.
learned relationship 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 Figure 4(b). According to this adjacent matrix, we can see that the target distortion is more similar (with similarity of 0.68) to the mixed mild distortion. Moreover, we observe that the mixed mild, moderate and severe distortions are similar/closed since these distortions have nearly the same characteristics with different levels, which in turn reveals the proposed distortion-relation graph indeed could effectively capture/model the different kinds of distortions.

4.4 Effectiveness of Leveraging Graph Relation

To prove the effectiveness of introducing a distortion-relation graph for real-world FS-RSR, we analyze the transferable distortion knowledge from auxiliary distortions to target distortion. As we all know, more transferable distortion knowledge (larger weight) can bring more positive influences for target distortion. Thus, we measure the task transfer-ability with the objective quality (i.e., PSNR) on target ReslSR. As shown in Figure 6(a), the edge coefficients of distortion-relation graph are positively associated with the task transfer-ability, which means that more similar of auxiliary and target distortions, more transferability/large PSNR could be achieved.

4.5 Comparison with State-of-the-arts

Comparison with Transfer Learning based Method. In this section, we compare our proposed DRTL with the SOTA transfer learning based methods in Table 1, including the basic pre-training [17] and MAML [15] schemes. Our baseline scheme is only directly trained on the few-shot target distortion dataset. Since our proposed DRTL is a model-agnostic optimization strategy, we select 4 general models, including (DnCNN [53], VDSR [18], RCAN [54], and RDN [55]) as backbones for evaluation. Specifically, we compare the training schemes of directly training (baseline), basic Pre-training, basic MAML, Pre-training with DRTL, and MAML with DRTL on these backbone models. As shown in Table 1, for all 4 backbone models, the proposed DRTL-related schemes (DRTL (Pre-training) (i.e., DRTL(PT)) and DRTL (MAML)) achieve the best performance on the real-world distorted test set, which indicates the effectiveness and generalization of our DRTL framework. Compared with the baseline scheme, the DRTL-related schemes could stably achieve nearly 0.3dB~0.5dB gains, which totally-fair results further reveal the effectiveness of our DRTL for few-shot image super-resolution. Moreover, for the simple-structured DnCNN, we find that it hardly converge well on the few-shot training clean-distorted real image pairs, and only achieves 29.297dB. In contrast, when applying our DRTL into DnCNN, this scheme could achieve a large performance improvements (31.1dB in PSNR).

Moreover, as shown in Table 1 we also observe that the basic Pre-training and MAML schemes both achieve obvious gains in comparison with baseline. But, they both ignore to explore how to better utilize the distortion relation
Fig. 6. Visualization of the relationship between similarity of distortions and transfer-ability. Here, the transfer-ability 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.

for knowledge transfer. Thanks to a simple gradient modulation with distortion relation guidance, our DRTL-related schemes could achieve extra obvious gains compared with the basic Pre-training/MAML schemes when using different backbones.

| Methods | Baseline | DIP [40] | Real-ESRGAN [44] | DRTL (PT) |
|---------|----------|----------|------------------|-----------|
| PSNR    | 31.2985  | 30.3202  | 28.2500          | 31.8111   |
| SSIM    | 0.8933   | 0.8704   | 0.8595           | 0.8988    |
| LPIPS   | 0.1502   | 0.1971   | 0.1412           | 0.1428    |

Table 2. Quantitative comparison of our DRTL with unsupervised methods.

Comparison with Unsupervised Method. Since DIP [40] and Real-ESRGAN [44] are general and famous unsupervised method for unknown distortions, we also compared it with our proposed DRTL. As shown in Table 2, we see that DIP cannot works very well when processing real-world distortions due to their complicated characteristics. Real-ESRGAN [44] achieves the best performance in LPIPS, while obtaining the worst performance in terms of PSNR and SSIM since the fake texture generation. Our DRTL on FS-RSR can effectively avoid this problem. In terms of qualitative/subjective comparison, as shown in Figure 5, our method has the capability to restore more texture details for real-world distorted image, e.g., the buildings in first row and the windows in the second row. We analyse that is because our DRTL optimization strategy could transfer more valuable knowledge with distortion-relation guidance from the auxiliary distortions to the target realSR.
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 relationships. To validate its effectiveness and necessity, we remove such design from our distortion-relation graph network, and evaluate the performance of this scheme on the MAML framework. As shown in Table 3, the performance of scheme \textit{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 L1 or L2 distance to compute the edges of distortion-relation graph. However, as shown in Figure 6, the relation coefficients do not satisfy a well linear trend as cosine similarity did.

Influence of the Number of Real Training Data. To study the influence of the number of real clean-distorted image training pairs w.r.t the final performance, we set several cases with the number of few-shot real image pairs respectively as 5, 10, 15, 20, 25 and 30 for comparison. As shown in Figure 7 with the number of samples going down, the performance of \textit{Baseline} scheme degrades quickly. In contrast, our DRTL scheme just decreases slightly. Moreover, our DRTL could achieve more gains when there are fewer samples, which further reveals the superiority of our DRTL method under the more challenging few-shot settings.

| Methods              | PSNR   | SSIM   | LPIPS  |
|----------------------|--------|--------|--------|
| w Memory Bank        | 31.3670| 0.8907 | 0.1476 |
| \textit{w/o Memory Bank} | 31.2396| 0.8879 | 0.1511 |

\textbf{Table 3.} Quantitative comparisons of our proposed DRTL with/without memory bank, which is based on DRTL(MAML).
5 Conclusion

In this paper, we are the first to focus on solving the challenging few-shot real-world image super-resolution (FS-RSR) 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 for real-world SR. However, the synthetic distortions and real distortion exist a large gap and the naive transfer learning cannot be adaptively optimized with the distortion relations. Therefore, we propose an distortion-relation graph with a prior knowledge memory bank to model the dependencies of different synthetic distortions and real-world distortion. Based on distortion-relation graph, we could select the most relevant auxiliary distortions for target task. Moreover, we propose a distortion-relation guided transfer learning (DRTL) framework with gradient re-weighting for real FS-RSR. Extensive experiments on the FS-RSR have validated the effectiveness of DRTL.
References

1. Agustsson, E., Timofte, R.: 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)
2. Cai, J., Zeng, H., Yong, H., Cao, Z., Zhang, L.: 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)
3. Castillo, A., Escobar, M., Pérez, J.C., Romero, A., Timofte, R., Van Gool, L., Arbelaez, P.: Generalized real-world super-resolution through adversarial robustness. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1855–1865 (2021)
4. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. arXiv preprint arXiv:2012.00364 (2020)
5. Chen, H., Wang, Y., Guo, T., Xu, C., Deng, Y., Liu, Z., Ma, S., Xu, C., Xu, C., Gao, W.: Pre-trained image processing transformer. arXiv preprint arXiv:2012.00364 (2020)
6. Chen, Y., Liu, Z., Xu, H., Darrell, T., Wang, X.: Meta-baseline: Exploring simple meta-learning for few-shot learning. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 9062–9071 (2021)
7. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11065–11074 (2019)
8. Deng, S., Wei, M., Wang, J., Feng, Y., Liang, L., Xie, H., Wang, F.L., Wang, M.: Detail-recovery image deraining via context aggregation networks. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14560–14569 (2020)
9. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence 38(2), 295–307 (2015)
10. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. pp. 391–407. Springer (2016)
11. Dong, Z., Cao, H., Shen, W., Gan, Y., Ling, Y., Zhai, G., Su, Y.: Frequency-aware physics-inspired degradation model for real-world image super-resolution. arXiv preprint arXiv:2111.03301 (2021)
12. Du, W., Chen, H., Yang, H.: Learning invariant representation for unsupervised image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14483–14492 (2020)
13. Du, W., Chen, H., Yang, H.: Learning invariant representation for unsupervised image restoration. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 14483–14492 (2020)
14. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning. pp. 1126–1135. PMLR (2017)
15. Finn, C., Abbeel, P., Levine, S.: Model-agnostic meta-learning for fast adaptation of deep networks. In: International Conference on Machine Learning. pp. 1126–1135. PMLR (2017)
16. Hamilton, W., Ying, Z., Leskovec, J.: Inductive representation learning on large graphs. In: Advances in neural information processing systems. pp. 1024–1034 (2017)
17. He, K., Girshick, R., Dollár, P.: Rethinking imagenet pre-training. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4918–4927 (2019)
18. Kim, J., Kwon Lee, J., Mu Lee, K.: 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)
19. Kim, Y., Soh, J.W., Park, G.Y., Cho, N.I.: Transfer learning from synthetic to real-noise denoising with adaptive instance normalization. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3482–3492 (2020)
20. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
21. Krishna, K., Murty, M.N.: Genetic k-means algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics) 29(3), 433–439 (1999)
22. Kupyn, O., Martyniuk, T., Wu, J., Wang, Z.: Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 8878–8887 (2019)
23. Li, X., Jin, X., Lin, J., Liu, S., Wu, Y., Yu, T., Zhou, W., Chen, Z.: Learning disentangled feature representation for hybrid-distorted image restoration. In: ECCV. pp. 313–329. Springer (2020)
24. Li, X., Jin, X., Yu, T., Pang, Y., Sun, S., Zhang, Z., Chen, Z.: Learning omnifrequency region-adaptive representations for real image super-resolution. AAAI (2021)
25. Li, X., Sun, S., Zhang, Z., Chen, Z.: Multi-scale grouped dense network for vvc intra coding. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 158–159 (2020)
26. Liang, J., Cao, J., Sun, G., Zhang, K., Van Gool, L., Timofte, R.: Swinir: Image restoration using swin transformer. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1833–1844 (2021)
27. Lim, B., Son, S., Kim, H., Nah, S., Mu Lee, K.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition workshops. pp. 136–144 (2017)
28. Lugmayr, A., Danelljan, M., Timofte, R.: Unsupervised learning for real-world super-resolution. In: 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW). pp. 3408–3416. IEEE (2019)
29. Moran, N., Schmidt, D., Zhong, Y., Coady, P.: Noisier2noise: Learning to denoise from unpaired noisy data. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12064–12072 (2020)
30. Nah, S., Hyun Kim, T., Mu Lee, K.: Deep multi-scale convolutional neural network for dynamic scene deblurring. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3883–3891 (2017)
31. Park, S., Yoo, J., Cho, D., Kim, J., Kim, T.H.: Fast adaptation to super-resolution networks via meta-learning. In: European Conference on Computer Vision. pp. 754–769. Springer (2020)
32. Qiu, Y., Wang, R., Tao, D., Cheng, J.: Embedded block residual network: A recursive restoration model for single-image super-resolution. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4180–4189 (2019)
33. Rusu, A.A., Rao, D., Sygnowski, J., Vinyals, O., Pascanu, R., Osindero, S., Hadsell, R.: Meta-learning with latent embedding optimization. arXiv preprint arXiv:1807.05960 (2018)

34. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

35. Snell, J., Swersky, K., Zemel, R.S.: Prototypical networks for few-shot learning. arXiv preprint arXiv:1703.05175 (2017)

36. Soh, J.W., Cho, S., Cho, N.I.: Meta-transfer learning for zero-shot super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3516–3525 (2020)

37. Suganuma, M., Liu, X., Okatani, T.: Attention-based adaptive selection of operations for image restoration in the presence of unknown combined distortions. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 9039–9048 (2019)

38. Tao, X., Gao, H., Shen, X., Wang, J., Jia, J.: Scale-recurrent network for deep image deblurring. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 8174–8182 (2018)

39. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9446–9454 (2018)

40. Ulyanov, D., Vedaldi, A., Lempitsky, V.: Deep image prior. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 9446–9454 (2018)

41. Umer, R.M., Foresti, G.L., Micheloni, C.: Deep generative adversarial residual convolutional networks for real-world super-resolution. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops. pp. 438–439 (2020)

42. Wang, T., Yang, X., Xu, K., Chen, S., Zhang, Q., Lau, R.W.: Spatial attentive single-image deraining with a high quality real rain dataset. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 12270–12279 (2019)

43. Wang, W., Zhang, H., Yuan, Z., Wang, C.: Unsupervised real-world super-resolution: A domain adaptation perspective. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 4318–4327 (2021)

44. Wang, X., Xie, L., Dong, C., Shan, Y.: Real-ergan: Training real-world blind super-resolution with pure synthetic data. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1905–1914 (2021)

45. Wei, P., Lu, H., Timofte, R., Lin, L., Zuo, W., Pan, Z., Li, B., Xi, T., Fan, Y., Zhang, G., et al.: Aim 2020 challenge on real image super-resolution: Methods and results. In: European Conference on Computer Vision. pp. 392–422. Springer (2020)

46. Wei, P., Xie, Z., Lu, H., Zhan, Z., Ye, Q., Zuo, W., Lin, L.: Component divide-and-conquer for real-world image super-resolution. In: European Conference on Computer Vision. pp. 101–117. Springer (2020)

47. Wei, W., Meng, D., Zhao, Q., Xu, Z., Wu, Y.: Semi-supervised transfer learning for image rain removal. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 3877–3886 (2019)

48. Yasarla, R., Sindagi, V.A., Patel, V.M.: Syn2real transfer learning for image deraining using gaussian processes. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 2726–2736 (2020)

49. Yoon, K.: Simple and efficient unpaired real-world super-resolution using image statistics. In: Proceedings of the IEEE/CVF International Conference on Computer Vision. pp. 1983–1990 (2021)
50. Yu, K., Dong, C., Lin, L., Loy, C.C.: Crafting a toolchain for image restoration by deep reinforcement learning. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2443–2452 (2018)

51. Yu, K., Dong, C., Loy, C.C., Tang, X.: Deep convolution networks for compression artifacts reduction. arXiv preprint arXiv:1608.02778 (2016)

52. Zhang, H., Patel, V.M.: Density-aware single image de-raining using a multi-stream dense network. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 695–704 (2018)

53. Zhang, K., Zuo, W., Chen, Y., Meng, D., Zhang, L.: Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing 26(7), 3142–3155 (2017)

54. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European conference on computer vision (ECCV). pp. 286–301 (2018)

55. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 2472–2481 (2018)

56. Zhu, J.Y., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of the IEEE international conference on computer vision. pp. 2223–2232 (2017)
Appendix

6 Details about the Algorithm of DRTL

To introduce the workflow of our Distortion-Relation guide Transfer-Learning (DRTL), we provide the detailed algorithm of DRTL. The DRTL can be divided into two components, respectively as the distortion graph learning and Relation guided Transfer Learning. The distortion graph learning is shown in Algorithm 1. The distortion embedding nodes of auxiliary distortions and target distortion of RealSR are first computed with our proposed distortion relation network (DRN). And then we compute the edges (i.e., distortion similarity) with the cosine similarity. The relation guided transfer learning is shown in Algorithm 2 and Algorithm 3. We integrate the distortion similarity that is inferenced out from distortion-graph into the optimization loop of Pre-training and MAML with gradient weighting.

Algorithm 1 Distortion Graph Learning

1: **Inputs:** Auxiliary distortions $T_{ad}^i$, where $1 \leq i \leq K$ and $K$ represents the number of the auxiliary distortions; Target distortion $T_{td}$ (i.e., RealSR); Our proposed distortion relation network (DRN) in this paper;
2: Get the distortion embedding nodes $c_a^i$ for each auxiliary distortion with $c_a^i = DRN(T_{ad}^i)$.
3: Get the distortion embedding nodes $c_t$ for target distortion with $c_t = DRN(T_{td})$.
4: Compute the relation of the $i$-th auxiliary distortion and target distortion with $\gamma_i = \text{cosine_similarity}(c_a^i, c_t)$
5: **Output:** The relation matrix $S = \{\gamma_i | 1 \leq i \leq K\}$

Algorithm 2 Distortion-Relation guide Transfer-Learning (DRTL) for Pre-training

1: **Inputs:** Auxiliary distortions $T_{ad}^i$, where $1 \leq i \leq K$ and $K$ represents the number of the auxiliary distortions; Target distortion $T_{td}$; Relation matrix $S = \{\gamma_i | 1 \leq i \leq K\}$; A learning model $f_\theta$; Step size hyperparameters $\alpha$.
2: for iteration = 0 to MAX_iteration do
3: for $i = 0$ to $K$ do
4: Update $\theta$ with $\theta' = \theta - \alpha \nabla_{\theta} \gamma_i \mathcal{L}_{T_{ad}^i}(f_\theta)$;
5: end for
6: end for
7: After getting the optimal $\theta_m$ in above Pre-training process and the model $f_{\theta_m}$ will be optimized with few-shot target distortion $T_{td}$ as $\theta_t = \theta_m - \alpha \nabla_{\theta} \mathcal{L}_{T_{td}}(f_{\theta_m})$;
8: **Output:** The optimal model on target distortion $f_{\theta_t}$. 
Algorithm 3 Distortion-Relation guide Transfer-Learning (DRTL) for MAML

1: Inputs: Auxiliary distortions $T_{ad}^i$, where $1 \leq i \leq K$ and $K$ represents the number of the auxiliary distortions; Target distortion $T_{td}$; Relation matrix $S = \{ \gamma_i | 1 \leq i \leq K \}$; Step size hyperparameters: $\alpha$ and $\beta$.
2: for iteration = 0 to MAX_iteration do
3:   for $i = 0$ to $K$ do
4:     Sample batch from auxiliary distortions $T_{ad}^i$;
5:     Evaluate $\nabla_\theta L_{T_{ad}^i}(f_\theta)$ with respect to sampled batch of distortion $T_{ad}^i$;
6:     Compute parameters with gradient descent: $\theta$ with $\theta_i' = \theta - \alpha \nabla_\theta L_{T_{ad}^i}(f_\theta)$;
7:   end for
8: Update $\theta$ with $\theta \leftarrow \theta - \beta \gamma \sum_{i=1}^{K} \gamma_i L_{T_{ad}^i}(f_{\theta_i'})$
9: end for
10: After getting the optimal $\theta_m$ in above Meta-train process and the model $f_{\theta_m}$ will be optimized with few-shot target distortion $T_{td}$ as $\theta_i = \theta_m - \alpha \nabla_\theta L_{T_{td}}(f_{\theta_m})$
11: Output: The optimal model on target distortion $f_{\theta_t}$.

7 Details of auxiliary distortions

In this section, we clarify the auxiliary distortions in detail. Specifically, we select most relevant 7 commonly-used synthetic distortions as auxiliary tasks based on the our proposed distortion relation graph, respectively as Bicubic downsampling [1], Bicubic downsampling with Anisotropic kernels [36], Gaussian noise, Gaussian blur, mixed mild distortion, mixed moderate distortion and mixed severe distortion [50]. The way of distortion generation is shown in Table 7. Particularly, the mixed distortions are composed of Gaussian noise, Gaussian blur and compression artifacts. The Gaussian noise, Gaussian blur and compression artifacts in mixed distortion are respectively divided into 10 levels with $\sigma \sim [0, 50]$, $\sigma \sim [0, 5]$ and compression quality $\sim [10, 100]$. Following the [50], the mixed distortion can be divided into three levels, respectively as mild, moderate and severe. Mixed mild distortion refers to that the distortion level of Gaussian noise, Gaussian blur and Jpeg artifacts are in the range of [9,11]. Mixed moderate distortion refers to that the distortion level of Gaussian noise, Gaussian blur and Jpeg artifacts are in the range of [12,17], and Mixed severe distortion refers to that the distortion level of Gaussian noise, Gaussian blur and Jpeg artifacts are in the range of [18,20]. More detail about the mixed distortion can be seen in [50].

8 More subjective comparisons between our DRTL and the state-of-the-art methods

To demonstrate the effectiveness of our proposed DRTL, we provide more subjective comparisons between our DRTL and the state-of-the-art methods in Figure. 8 and 9. Note that the baseline method refers to directly training the model without Pre-training/MAML.
| Distortion types | Generation                                                                                          |
|------------------|------------------------------------------------------------------------------------------------------|
| Bicubic          | Bicubic with scale 8                                                                              |
| Ani_bic          | Bicubic with scale 4 + anistropic blur                                                              |
| Noise            | Gaussian noise with $\sigma$ from the range of $[0, 50]$                                            |
| Blur             | Gaussian blur with $\sigma$ 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 $[18, 20]$                                                  |

Table 4. The way of auxiliary distortions generation. To save space, the distortion type have been abbreviated as follows: Bicubic downsampling (i.e., Bicubic), Bicubic downsampling with anistropic blurring (i.e., Ani_bic), Gassian noise (i.e., Noise), Gaussian Blur (i.e., Blur), Mixed mild (i.e., Mild), Mixed moderate (i.e., Moderate), Mixed severe (i.e., Severe).

Fig. 8. Subjective comparison of our DRTL with the state-of-the-art methods.
Fig. 9. Subjective comparison of our DRTL with the state-of-the-art methods.