Few-Shot Real Image Restoration via Distortion-Relation Guided Transfer Learning

Xin Li\(^1\) \* Xin Jin\(^1\) \* Jun Fu\(^1\) Xiaoyuan Yu\(^2\) Bei Tong\(^2\) Zhibo Chen\(^1\)†

\(^1\)University of Science and Technology of China \(^2\)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 restoration (IR) 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-IR) is more feasible to tackle this challenging real Image Restoration task. In this paper, we are the first to investigate the few-shot real image restoration 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 (i.e., synthetic distortions) and target tasks (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-IR. Extensive experiments on multiple benchmarks demonstrate the effectiveness of DRTL on few-shot real image restoration.

1. Introduction

Image restoration (IR) task aims to restore high-quality images from the degraded low-quality images. The real-world degradation process typically consists of a series of different distortions, such as motion blur [21, 33], noise [25, 44], compression artifacts, [24, 42], raining [6, 36], and etc. 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 contain various distortions have been proposed and deeply explored/studied, such as DIV2K [1], Gopro [26], and DID-MDN [43]. However, the degradations of these datasets are typically synthetic, or said, hand-craft, the distortions of which are different from the real-world ones.

Therefore, although some learning-based IR approaches that trained on the above-mentioned synthetic datasets have achieved a great success when handling the synthetic degra-
ations, those image restoration systems hardly perform well on the real-world scenarios due to the large gap between synthetic distortions and real-world distortions [23, 39,40]. Furthermore, re-collecting large clean-distorted image training pairs from real-world is also non-trivial [22].

The above-mentioned two disgusting weaknesses of the existing image restoration algorithms seriously limit their practical applications and hurt their industrial values. To alleviate the dependency for clean distortion/clean image training pairs, some studies [9,34] 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 IR scheme with Few-Shot distortion/clean image pairs image (FS-IR) is more feasible to tackle this challenging real Image Restoration.

Previous works have explored different transfer learning strategies [4,11,28,30] 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 restoration. For instance, IPT [3] introduces a large-scale pre-training dataset to improve the restoration performance w.r.t the target distortion. Soh et al. [31] 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 restoration problem.

In this paper, we are the first to focus on the challenging few-shot real image restoration problem, where the auxiliary tasks (i.e., synthetic distortions) and target task (i.e., real-world restorations) are different and exist a large gap. Therefore, we face two key issues needed be solved: 1) “how to select the proper auxiliary synthetic distortions?” and 2) “how to capture the proper prior knowledge 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 [19]. To better capture the relation 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 distortion, 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 [18].

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 inferred from distortion-graph into the optimization loop of pre-training/MAML [12] with gradient weighting. Extensive experiments on multiple benchmarks have demonstrated the effectiveness of our DRTL on few-shot real image restoration.

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 restoration (FS-IR) problem.
- We introduce a distortion graph to guide the few-shot real image restoration 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-IR, 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 multiple FS-IR 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-IR framework that is compatible with many existing IR networks. We will release our source code upon acceptance.

2. Related Work

2.1. Image restoration

Deep learning has actually accelerated the development of image restoration techniques. Compared with the traditional methods that are mostly based on handcrafted image priors, learning-based methods could automatically capture the statistic prior of original images from distorted images. Supervised Image Restoration Based on the characteristics of different distortions, numerous classic fully-supervised learning-based works have been proposed and applied into many practical applications, such as SRCNN [7]/FSRCNN [8] targets for image super-resolution, DeepDeblur [26]/DeblurGAN [20] targets for image deblurring,
DnCNN [44] targets for image denoising, and DerainNet [13] targets for image deraining. Besides, in order to further improve the generalization and effectiveness of image restoration, some works [5, 27, 32, 33, 37] have been devoted into modifying the network to be more suitable for specific distortion as much as possible. However, these works typically only focus on handling synthetic distortions in a fully-supervised manner, and also require a large-scale clean-distorted image pairs for training.

**Unsupervised 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. [35] 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. [10] 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 for 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. [39, 40] 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. [31] propose to leverage the Meta-transfer learning to deal with the challenging task of zero-shot super-resolution (ZSSR). Kim et al. [17] utilize the adaptive instance normalization to realize the knowledge transfer from synthetic noise to real noise.

Moreover, with the development of real image restoration, some real image restoration datasets with limited real-world clean-distorted image pairs have been collected and released [2, 38]. However, it is hard to directly train models on these small-scale datasets to get the satisfying restoration performance due to the data limitation. Thus, a method that could utilize few-shot samples to achieve a better real image restoration performance is highly desired.

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 [3, 15]. 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 $f_0$ 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,

$$\theta_i' = \theta - \alpha \nabla_{\theta} \mathcal{L}_{\mathcal{T}^a}(f_0), \quad \text{(1)}$$

where the $\mathcal{L}_{\mathcal{T}^a}(f_0)$ refers to the $L_1$ loss in $\mathcal{T}^a$. The optimization objective of pre-training is to minimize the loss function in all auxiliary tasks as Eq. 2.

$$\min_{\theta} \sum_{\mathcal{T}^a \sim p(\mathcal{T}^a)} \mathcal{L}_{\mathcal{T}^a}(f_0) \quad \text{(2)}$$
In the fine-tuning stage, the best parameter $\theta_m$ in pre-training stage is used as a initial parameters. Then model $f_{\theta_m}$ is updated with the optimization objective to acquire the best performance in target task as Equ: 

$$\min_{\theta} \mathcal{L}_{T_t}(f_{\theta})$$

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 [12]. 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 $\mathcal{L}_{T_a}$, 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: 

$$\min_{\theta} \sum_{T_a \sim p(T_a)} \mathcal{L}_{T_a}(f_{\theta-\alpha\nabla_\theta \mathcal{L}_{T_a}(f_{\theta}))}$$

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

$$\theta_m=\theta-\beta\nabla_\theta \sum_{T_a \sim p(T_a)} \mathcal{L}_{T_a}(f_{\theta'})$$

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, 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.

3.2.1 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 information between [0, 1]. We propose to utilize the feature extractor of VGG-11 [29] 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.

3.2.2 Knowledge Memory Bank

Generally, the graph learning aims to capture internal relationship between samples or structures, and then assist the current task [14, 18]. 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 distor-
tions 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 $H_m$. 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.

### 3.2.3 Graph Nodes and Edges

The distortion relation graph $\mathcal{R} = (\mathcal{C}_\mathcal{R}, \mathcal{A}_\mathcal{R})$ consists of two essential components, i.e., Nodes $\mathcal{C}_\mathcal{R} = \{c^i | \forall i \in [1, K]\} \in \mathbb{R}^{K \times d}$ and Edges $\mathcal{A}_\mathcal{R} = \{\mathcal{A}_\mathcal{R}(c^i, c^j) | \forall i, j \in [1, K]\} \in \mathbb{R}^{K \times K}$. In this paper, Nodes refer to feature embedding of each distorted sample projected in knowledge memory bank $\mathcal{M} = (H_M, A_M)$, 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

$$A_p^i = \sigma(||F_{dist}^i - H_M||_2).$$

Then we concatenate the relation adjacent matrix $A_p^i$ and the adjacent matrix of prior knowledge memory bank $A_M$,

$$A = [A_p^i, A_M],$$

where $A_M$ can be computed as Eq- 8

$$A_M = \sigma(||H_m - H_n||_2)|m, n \in \{1, Q\}. (8)$$

Finally, the nodes of distortion relation graph can be generated with graph convolution network (GCN) [18] as,

$$c_i = GCN ([F_{dist}, H_M], A).$$

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 samples, which acting as the edges $\mathcal{A}_\mathcal{R}$ 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 image restoration. 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 distortion restoration task, 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 $T_i^{ad}$ and target distortion $T_t^{ad}$ from our proposed distortion relation graph $\mathcal{R}$. 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.

$$\theta_m = \theta - \beta_t \nabla_{\theta} \sum_{T_i^{ad} \in p(T^{ad})} \gamma_i \mathcal{L}_{T_i^{ad}}(f_{\theta^i}), (10)$$

$$\theta'_i = \theta - \alpha_t \nabla_{\theta} \gamma_i \mathcal{L}_{T_i^{ad}}(f_{\theta}), (11)$$

In the process of meta-test or fine-tuning, the optimal transferable model parameters $\theta_m$ only need to be fine-tuned with the few-shot distorted/clean pairs of target distortion as Eq-3.

### 4. Experiments

In this section, we first introduce the auxiliary distortions datasets and target distortion 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 7 common synthetic distortions as auxiliary tasks based on the distortion relation, including Bicubic downsampling [1], Bicubic downsampling with An-isotropic kernels [31], Gaussian noise, Gaussian blur, Mixed mild distortion, Mixed moderate distortion, Mixed severe distortion [32]. 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. **Target dataset** Since the real-world distortions are usually hybrid distortions, we select the typical RealSR [2] dataset.
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.

| Models     | Baseline  | Pre-training | MAML    | DRTL(Pre-training) | DRTL(MAML) |
|------------|-----------|--------------|---------|---------------------|------------|
|            | PSNR      | SSIM         | PSNR    | SSIM                | PSNR       | SSIM       | PSNR    | SSIM                | PSNR       | SSIM       |
| DnCNN [44] | 29.2967   | 0.8738       | 30.8581 | 0.8861              | 23.4663    | 0.8060     | 30.9772 | 0.8892              | 31.1007    | 0.8908     |
| VDSR [16]  | 31.0673   | 0.8849       | 31.2903 | 0.8901              | 31.2185    | 0.8877     | 31.3580 | 0.8908              | 31.3670    | 0.8907     |
| RCAN [45]  | 31.2985   | 0.8933       | 31.6978 | 0.8977              | 31.5464    | 0.8946     | 31.8111 | 0.8988              | 31.5969    | 0.8956     |
| RDN [46]   | 31.2354   | 0.8914       | 31.5349 | 0.8948              | 31.3919    | 0.8926     | 31.6554 | 0.8965              | 31.4925    | 0.8940     |

Figure 4. Visualization of nodes in distortion relation graph.

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 IR model VDSR [16], that contains a series of non-linear Resblocks, as our baseline model. The implementation of DRTL is based on PyTorch 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 64x64. $L_1$ loss have been proved effective to optimize the model especially in image restoration [2,45]. 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, 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 learned relationship between different auxiliary distortions and target distortion, we further visualize the adjacent similarity matrix of different distortion nodes in Figure 5. 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 distor-
Figure 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.

4.4. Effectiveness of Leveraging Graph Relation

To prove the effectiveness of introducing a distortion-relation graph for real-world FS-IR, 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 distorted image restoration. 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 transfer-ability/large PSNR could be achieved.

Table 2. Quantitative comparison of our DRTL with DIP.

| Methods            | PSNR       | SSIM       | LPIPS      |
|--------------------|------------|------------|------------|
| Baseline           | 31.2985    | 0.8933     | 0.1502     |
| DIP [35]           | 30.3202    | 0.8704     | 0.1971     |
| DRTL (Pre-training)| 31.8111    | 0.8988     | 0.1428     |

4.5. Comparison with state-of-the-art methods

Comparison with Transfer Learning based IR Method. In this section, we compare our proposed DRTL with the SOTA transfer learning based IR methods in Table 1, including the basic pre-training [15] and MAML [12] 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 IR models, including (DnCNN [44], VDSR [16], RCAN [45], and RDN [46]) 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) 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 restoration. 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 29.2967dB. In contrast, when applying our DRTL into DnCNN, this scheme could achieve a large performance improvements (31.1007dB 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 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.

Comparison with Unsupervised IR Method. Since DIP [35] is an 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. In terms of qualitative/subjective comparison, as shown in Figure 7, 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 real distortion.

Table 3. Quantitative comparisons of our proposed DRTL with/without memory bank, which is based on VDSR framework and DRTL(MAML).

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

4.6. Ablation Study

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 8, with the number of samples going down, the performance of 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.

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 w/o Memory Bank degrades since the correlation between different distortions cannot be captured effectively/accurately.

Moreover, 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.

5. Conclusion

In this paper, we are the first to focus on solving the challenging few-shot real-world image restoration (FS-IR) 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 distorted image restoration. 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. And we propose a distortion-relation guided transfer learning (DRTL) framework with gradient re-weighting for real FS-IR. Extensive experiments on the real few-shot image restoration have validated the effectiveness of
References

[1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pages 126–135, 2017. 1, 5, 11
[2] Jianrui Cai, Hui Zeng, Hongwei Yong, Zisheng Cao, and Lei Zhang. Toward real-world single image super-resolution: A new benchmark and a new model. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 3086–3095, 2019. 3, 5, 6
[3] Hanting Chen, Yunhe Wang, Tianyu Guo, Chang Xu, Yiping Deng, Zhenhua Liu, Siwei Ma, Chunjing Xu, Chao Xu, and Wen Gao. Pre-trained image processing transformer. arXiv preprint arXiv:2012.00364, 2020. 2, 3
[4] Yinbo Chen, Zhuang Liu, Huijuan Xu, Trevor Darrell, and Xiaolong Wang. Meta-baseline: Exploring simple meta-learning for few-shot learning. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 9062–9071, 2021. 2
[5] Tao Dai, Jianrui Cai, Yongbing Zhang, Shu-Tao Xia, and Lei Zhang. Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 11065–11074, 2019. 3
[6] Sen Deng, Mingqiang Wei, Jun Wang, Yidan Feng, Luming Liang, Haoran Xie, Fu Lee Wang, and Meng Wang. Detail-recovery image deraining via context aggregation networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14560–14569, 2020. 1
[7] Chao Dong, Chen Change Loy, Kaiming He, and Xiaou Tang. Image super-resolution using deep convolutional networks. IEEE transactions on pattern analysis and machine intelligence, 38(2):295–307, 2015. 2
[8] Chao Dong, Chen Change Loy, and Xiaou Tang. Accelerating the super-resolution convolutional neural network. In European conference on computer vision, pages 391–407. Springer, 2016. 2
[9] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14483–14492, 2020. 2
[10] Wenchao Du, Hu Chen, and Hongyu Yang. Learning invariant representation for unsupervised image restoration. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 14483–14492, 2020. 3
[11] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In International Conference on Machine Learning, pages 1126–1135. PMLR, 2017. 2, 4, 7
[12] Chelsea Finn, Pieter Abbeel, and Sergey Levine. Model-agnostic meta-learning for fast adaptation of deep networks. In International Conference on Machine Learning, pages 1126–1135. PMLR, 2017. 2, 4, 7
[13] Xueyang Fu, Jiabin Huang, Xinghao Ding, Yinghao Liao, and John Paisley. Clearing the skies: A deep network architecture for single-image rain removal. IEEE Transactions on Image Processing, 26(6):2944–2956, 2017. 3
[14] Will Hamilton, Zhitao Ying, and Jure Leskovec. Inductive representation learning on large graphs. In Advances in neural information processing systems, pages 1024–1034, 2017. 4
[15] Kaiming He, Ross Girshick, and Piotr Dollár. Rethinking imagenet pre-training. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4918–4927, 2019. 3, 7
[16] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 1646–1654, 2016. 6, 7
[17] Yoonsik Kim, Jae Woong Soh, Gu Yong Park, and Nam Ik Cho. 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, pages 3482–3492, 2020. 3
[18] Thomas N Kipf and Max Welling. Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907, 2016. 2, 4, 5
[19] K Krishna and M Narasimha Murty. Genetic k-means algorithm. IEEE Transactions on Systems, Man, and Cybernetics, Part B (Cybernetics), 29(3):433–439, 1999. 2
[20] Orest Kupyn, Volodymyr Budzan, Mykola Mykhailych, Dmytro Mishkin, and Jiří Matas. Deblurgan: Blind motion deblurring using conditional adversarial networks. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 8183–8192, 2018. 2
[21] Orest Kupyn, Tetiana Martyniuk, Junru Wu, and Zhangyang Wang. Deblurgan-v2: Deblurring (orders-of-magnitude) faster and better. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 8878–8887, 2019. 1
[22] Xin Li, Xin Jin, Jianxin Lin, Sen Liu, Yaojun Wu, Tao Yu, Wei Zhou, and Zhibo Chen. Learning disentangled feature representation for hybrid-distorted image restoration. In ECCV, pages 313–329. Springer, 2020. 2
[23] Xin Li, Xin Jin, Tao Yu, Yingxue Pang, Simeng Sun, Zhizheng Zhang, and Zhibo Chen. Learning omni-frequency region-adaptive representations for real image super-resolution. AAAI, 2021. 2
[24] Xin Li, Simeng Sun, Zhizheng Zhang, and Zhibo Chen. Multi-scale grouped dense network for vvc intra coding. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, pages 158–159, 2020. 1
[25] Nick Moran, Dan Schmidt, Yu Zhong, and Patrick Coady. Noisier2noise: Learning to denoise from unpaired noisy data. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 12064–12072, 2020. 1
[26] Seungjun Nah, Tae Hyun Kim, and Kyoung Mu Lee. Deep multi-scale convolutional neural network for dynamic scene
deblurring. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3883–3891, 2017. 1, 2

[27] Yajun Qiu, Ruxin Wang, Dapeng Tao, and Jun Cheng. Embedded block residual network: A recursive restoration model for single-image super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4180–4189, 2019. 3

[28] Andrei A Rusu, Dushyant Rao, Jakub Synkowski, Oriol Vinyals, Razvan Pascanu, Simon Osindero, and Raia Hadsell. Meta-learning with latent embedding optimization. arXiv preprint arXiv:1807.05960, 2018. 2

[29] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556, 2014. 4

[30] Jake Snell, Kevin Swersky, and Richard S Zemel. Prototypical networks for few-shot learning. arXiv preprint arXiv:1703.05175, 2017. 2

[31] Jae Woong Soh, Sunwoo Cho, and Nam Ik Cho. Meta-transfer learning for zero-shot super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3516–3525, 2020. 2, 3, 5, 11

[32] Masanori Suganuma, Xing Liu, and Takayuki Okatani. 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, pages 9039–9048, 2019. 3, 5

[33] Xin Tao, Hongyun Gao, Xiaoyong Shen, Jue Wang, and Jiayu Jia. Scale-recurrent network for deep image deblurring. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pages 8174–8182, 2018. 1, 3

[34] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 9446–9454, 2018. 2

[35] Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 9446–9454, 2018. 3, 7

[36] Tianyu Wang, Xin Yang, Ke Xu, Shaozhe Chen, Qiang Zhang, and Rynson WH Lau. 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, pages 12270–12279, 2019. 1

[37] Wei Wang, Ruiming Guo, Yapeng Tian, and Wenming Yang. Cfsnet: Toward a controllable feature space for image restoration. In Proceedings of the IEEE/CVF International Conference on Computer Vision, pages 4140–4149, 2019. 3

[38] Pengxu Wei, Ziwei Xie, Hannan Lu, Zongyuan Zhan, Qixiang Ye, Wangmeng Zuo, and Liang Lin. Component divide-and-conquer for real-world image super-resolution. In European Conference on Computer Vision, pages 101–117. Springer, 2020. 3

[39] Wei Wei, Deyu Meng, Qian Zhao, Zongben Xu, and Ying Wu. Semi-supervised transfer learning for image rain removal. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 3877–3886, 2019. 2, 3

[40] Rajeev Yasarla, Vishwanath A Sindagi, and Vishal M Patel. Syn2real transfer learning for image deraining using gaussian processes. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 2726–2736, 2020. 2, 3

[41] Ke Yu, Chao Dong, Liang Lin, and Chen Change Loy. Crafting a toolchain for image restoration by deep reinforcement learning. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2443–2452, 2018. 11

[42] Ke Yu, Chao Dong, Chen Change Loy, and Xiaou Tang. Deep convolution networks for compression artifacts reduction. arXiv preprint arXiv:1608.02778, 2016. 1

[43] He Zhang and Vishal M Patel. Density-aware single image de-raining using a multi-stream dense network. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 695–704, 2018. 1

[44] Kai Zhang, Wangmeng Zuo, Yunjin Chen, Deyu Meng, and Lei Zhang. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. IEEE transactions on image processing, 26(7):3142–3155, 2017. 1, 3, 6, 7

[45] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European conference on computer vision (ECCV), pages 286–301, 2018. 6, 7

[46] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 2472–2481, 2018. 6, 7
Appendix

6. Details about the Algorithm of DRTL

Algorithm 1 Distortion Graph Learning

Require: Auxiliary distortions $T_{ad}$, where $1 \leq i \leq K$ and $K$ represents the number of the auxiliary distortions; Target distortion $T_d$: Our proposed distortion relation network (DRN) in this paper;
1: Get the distortion embedding nodes $c^i_a$ for each auxiliary distortion with $c^i_a = DRN(T_{ad})$.
2: Get the distortion embedding nodes $c_t$ for target distortion with $c_t = DRN(T_d)$.
3: Compute the relation of the $i$-th auxiliary distortion and target distortion with
   $\gamma_i = cosine_similarity(c^i_a, c_t)$
Ensure: The relation matrix $S = \{\gamma_i | 1 \leq i \leq K\}$

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

Require: Auxiliary distortions $T_{ad}$, where $1 \leq i \leq K$ and $K$ represents the number of the auxiliary distortions; Target distortion $T_d$: Relation matrix $S = \{\gamma_i | 1 \leq i \leq K\}$; A learning model $f_\theta$; Step size hyperparameters $\alpha$.
1: for iteration = 0 to MAX iteration do
2:   for $i = 0$ to $K$ do
3:       Update $\theta$ with $\theta_i = \theta - \alpha \nabla_\theta \gamma_i L_{T_{ad}}(f_\theta)$;
4:   end for
5: end for
6: 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 $T_d$ as $\theta_t = \theta_m - \alpha \nabla_\theta L_{T_d}(f_{\theta_m})$
Ensure: The optimal model on target distortion $f_{\theta_t}$.

Algorithm 3 Distortion-Relation guide Transfer-Learning (DRTL) for MAML

Require: Auxiliary distortions $T_{ad}$, where $1 \leq i \leq K$ and $K$ represents the number of the auxiliary distortions; Target distortion $T_d$: Relation matrix $S = \{\gamma_i | 1 \leq i \leq K\}$; Step size hyperparameters $\alpha$ and $\beta$.
1: for iteration = 0 to MAX iteration do
2:   for $i = 0$ to $K$ do
3:       Sample batch from auxiliary distortions $T_{ad}^i$;
4:       Evaluate $\nabla_\theta L_{T_{ad}^i}(f_\theta)$ with respect to sampled batch of distortion $T_{ad}^i$;
5:       Compute parameters with gradient descent: $\theta$ with $\theta_i = \theta - \alpha \nabla_\theta L_{T_{ad}^i}(f_\theta)$;
6:   end for
7:   Update $\theta$ with $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{i=1}^{K} \gamma_i L_{T_{ad}^i}(f_{\theta_i})$
8: end for
9: 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 $T_d$ as $\theta_t = \theta_m - \alpha \nabla_\theta L_{T_d}(f_{\theta_m})$
Ensure: 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 randomly select 7 commonly-used synthesis distortions, respectively as Bicubic downsampling [1], Bicubic downsampling with Anisotropic kernels [31], Gaussian noise, Gaussian blur, mixed mild distortion, mixed moderate distortion and mixed severe distortion [41]. The way of distortion generation is shown in Table 4. 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 [41], 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 [41].

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
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).

| 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] |

DRTL and the state-of-the-art methods in Figure. 9 and 10.
Note that the baseline method refers to directly training the model without pre-training/MAML.
Figure 10. Subjective comparison of our DRTL with the state-of-the-art methods.