Transitional Learning: Exploring the Transition States of Degradation for Blind Super-resolution

Yuanfei Huang, Jie Li, Yanting Hu, Xinbo Gao, Senior Member, IEEE, and Hua Huang, Senior Member, IEEE

Abstract—Being extremely dependent on iterative estimation of the degradation prior or optimization of the model from scratch, the existing blind super-resolution (SR) methods are generally time-consuming and less effective, as the estimation of degradation proceeds from a blind initialization and lacks interpretable degradation priors. To address it, this paper proposes a transitional learning method for blind SR using an end-to-end network without any additional iterations in inference, and explores an effective representation for unknown degradation. To begin with, we analyze and demonstrate the transitionality of degradations as interpretable prior information to indirectly infer the unknown degradation model, including the widely used additive and convolutive degradations. We then propose a novel Transitional Learning method for blind Super-Resolution (TLSR), by adaptively inferring a transitional transformation function to solve the unknown degradations without any iterative operations in inference. Specifically, the end-to-end TLSR network consists of a degree of transitionality (DoT) estimation network, a homogeneous feature extraction network, and a transitional learning module. Quantitative and qualitative evaluations on blind SR tasks demonstrate that the proposed TLSR achieves superior performances and costs fewer complexities against the state-of-the-art blind SR methods. The code is available at github.com/YuanfeiHuang/TLSR.

Index Terms—Blind super-resolution, multiple degradations, transition state, degradation representation.

1 INTRODUCTION

Single image super-resolution (SISR), aiming at reconstructing a high-resolution (HR) image from a degraded low-resolution (LR) image, is considered an ill-posed inverse problem. For decades, many studies have been proposed to solve this ill-posed problem, including interpolation based [1], reconstruction based [2, 3] and example learning based [4, 5, 6, 7] methods.

Recently, with the development of the high-profile deep convolutional neural networks (CNNs), and to learn the co-occurrence priors or non-linear representations between HR-LR pairs from a large training datasets, deep learning based SISR methods have received considerable attention due to their excellent performance and real-time processing. As the first attempt, Dong et al. [8, 9] proposed SRCNN by stacking a shallow CNN to learn the non-linear mappings of LR-to-HR pairs, which outperforms most existing example-based SR methods and leads the trend of using end-to-end networks for SISR [10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21]. Nevertheless, as the degradation is generally synthesized under an assumed degradation prior (e.g., bicubic down-sampling), these deep learning based SR models were trained by optimizing on numerous synthetic HR-LR pairs with single degradation, and unfortunately show less effectiveness on other degradations.

Thus, the integrated models with conditional information have been developed by training a single model on multiple/variational degradations, e.g., SRMD [22] and UDVD [23]. However, these integrated models need giving a degradation prior condition to guide the model to solve the corresponding degradation and show less effectiveness when the degradation prior is unknown. Then, blind super-resolution (BSR) has been raised to improve the generalization of the model on unknown degradations, and the corresponding methods can be divided into two branches: integrated learning based methods (e.g., DrCNN [24], VDN [25], IKC [26] and DASR [27]) and zero-shot learning based methods (e.g., ZSSR [28], DIP [29], KernelGAN [30] and MZSR [31]).

However, by iteratively estimating the prior information of degradations or optimizing the model from scratch, most of the existing integrated learning based and zero-shot learning based BSR methods are time-consuming and not controllable enough. Therefore, it is natural to raise two questions: (1) Is it possible to facilitate the inference/testing phase in a non-iterative way? (2) Is there a natural characteristic in degradations to improve the adaptability of models?

(1) For the first issue, an end-to-end network is necessary to simultaneously estimate the degradation priors and reconstruct the degraded LR images. Thus, we need a sufficiently accurate degradation estimation without any additional iterative corrections.

(2) For the second issue, we revisit and analyze the relations between various degradations, and find that most degradations are generally transition states of two given primary degradations. Therefore, it is feasible to super-
2 RELATED WORK

2.1 Non-Blind Super-Resolution

Non-blind super-resolution aims at reconstructing the degraded LR images using a single model trained with the given prior information of degradations and well-designed objective optimization functions, which is the same as SISR. Representatively, SRCNN [9], VDSR [32], EDSR [16], SRGAN [14] and ESRGAN [35] were presented to learn the non-linear representation of LR-to-HR mappings using $\ell_2/\ell_1$ loss functions or adversarial learning, where the degraded LR images are generated by a single bicubic down-sampling degradation.

To improve the generalization of models on various degradations, SRMD [22] and UDVD [23] were proposed by integrating numerous degradations in a single integrated network, which is a conditional method for solving various degradations under the non-blind degradation prior. Besides, for arbitrary blur kernels, plug-and-play [34], [35], [36], and deep unfolding network [37] were introduced to correct the model using non-blind priors.

2.2 Blind Super-Resolution

However, the degradations are generally unknown in real-world applications. Since the non-blind models are trained on a priori of single or multiple specific degradations, they would show less effectiveness when the prior of degradation is not given. Thus, blind super-resolution has been raised to solve this issue for a better generalization on real-world scenes, and could be divided into two branches according to the manner of optimization as follows.

2.2.1 Integrated learning based methods

For unknown degradations, based on a non-blind integrated model $F_{\Theta}$, which is trained with variational degradations and the corresponding given conditional information of degradation prior $D$, the pipeline of integrated learning based BSR methods (IL-BSR) is to predict $D$ by an estimator $F_D$ firstly and then utilize it as conditional guidance of $F_{\Theta}$ to estimate a desired result as

$$\begin{align*}
D_k &= F_D(y_{k-1}) \\
y_k &= F_{\Theta}(y_{k-1}, D_k)
\end{align*}$$

where $y_k$ denotes the current super-resolved result, and $y_0$ indicates the LR input $x$.

Representatively, IKC [26] was proposed to iteratively predict and correct the conditional information of degradation using the LR inputs or super-resolved results, and apply the estimated conditional information into an integrated non-blind model. Furthermore, DASR [27] utilized the contrastive learning [38] to estimate a discriminative degradation representation as conditional information. Moreover, toward blind image denoising, VDN [25] was proposed by exploiting a deep variational inference network for noise modeling and removal. However, these IL-BSR methods are not controllable enough, as they largely depend on the estimated $D$ which is high-dimension vectors and with high freedom for representation.

2.2.2 Zero-shot learning based methods

Except for modeling variational degradations by a single model, the zero-shot learning based BSR methods (ZSL-BSR) have been raised to generate a specific model using a single image for an unknown degradation. Without the ground truth, ZSL-BSR iteratively optimize the image-specific model $F_{\Theta_k}$ under the self-supervised guidelines, as

$$\begin{align*}
\Theta_k &= \arg\min_{\Theta_k} \mathcal{L}(y_{k-1}, x, F_{\Theta}) \\
y_k &= F_{\Theta_k}(y_{k-1})
\end{align*}$$

where $\mathcal{L}$ represents a predefined objective function.
Representatively, without any external dataset, ZSSR [28] was exploited to solve image-specific blind SR using internal patches of a single image. Similarly, without explicit prior information of degradation, DIP [29] utilized a deep network to learn the image priors for image restoration; Noise2Noise [39] applied statistical reasoning by corrupting the corrupted observations for image denoising; KernelGAN [30] was explored to estimate the blur kernels and use the estimated kernels to generate degraded images for further training or inference. Besides, meta-learning [40] was introduced to facilitate model adaptation, e.g., MZSR [31] and MLSR [41]. However, these methods are time-consuming as they require iterative optimization during inference.

### 3 Proposed Method

Inspired by transition state, if the given degradation is transitional, arbitrary unknown variants would be represented by two or more primary states. So, to begin with, we analyze the transitionality of the widely used additive and convolutive degradations. Then we describe the methodologies of transitional learning and its derived implementation on blind super-resolution.

#### 3.1 Transitionality of Degradations

In classic SR problem, there generally exists two types of degradation (convolutive degradation $\mathcal{B}$ for isotropic Gaussian blur and additive degradation $\mathcal{N}$ for AWGN) in the LR corrupted images, which is formulated as

$$x = (y \otimes B) \downarrow x + \mathcal{N}$$  \hspace{1cm} (3)

where $\otimes$ represents the convolution operator, $\downarrow x$ indicates $x$ downsampling (e.g., bicubic downsampling), $x$ and $y$ denote LR image and its HR counterpart respectively.

Aiming at solving the SR task on this complex degradation, most existing non-blind and blind SR methods attempt to optimize a well-designed specific or integrated model $\mathcal{F}(x; \Theta, \uparrow x) : x \mapsto y$. For a specific degradation in problem space, the accurate degradation priors make the optimization of models more effective. However, these models would be less effective if a degradation is unseen in the problem space. Therefore, we expect to find an effective representation for arbitrary degradation in or beyond the problem space, including additive degradations and convolutive degradations.

#### 3.1.1 Additive degradation $\mathcal{N}$

Since the additive degradations are independent and linearly additive, given the degree of transitionality (DoT) $\tau$, the transition states $\{x^\tau\}$ are then formulated as

$$x^\tau = x + \mathcal{N}^\tau = \tau(x + \mathcal{N}^0) + (1 - \tau)(x + \mathcal{N}^1)$$  \hspace{1cm} (4)
where, $N^\tau = \tau N^0 + (1 - \tau) N^1$ represents that the transitional additive degradations can be represented by two primary degradations $N^0$ and $N^1$. Particularly, when the original state is noise-free ($N^0 = 0$), the estimation of transition state is simplified into a linear polynomial function as $N^\tau = (1 - \tau) N^1$.

### 3.1.2 Convolutive degradation $B$

Unlike the additive degradations, the convolutive degradations are generated by non-linear functions (e.g., Gaussian kernels) and not linear additive. Nevertheless, inspired by the Gaussian mixture model (GMM) \(^7\), we can infer any convolutive Gaussian distribution as a mixture of multiple Gaussian distributions as

$$B = \sum_{i=0}^{K-1} \tau_i B^i \sim \sum_{i=1}^{K-1} \tau_i \mathcal{N}(\mu_i, \sigma_i^2)$$  \hspace{1cm} (5)

particularly, when $K = 2$, assume that the primary convolutive degradations are independent and $B^0 \sim \mathcal{N}(0, \sigma_0^2)$, $B^1 \sim \mathcal{N}(0, \sigma_1^2)$. Specifically in Proposition 1, the transition state of convolutive degradation can be formulated as an algebra combination of the primary convolutive degradations, which demonstrates the transitionality of convolutive degradation.

**Proposition 1.** Given two Gaussian blur kernels $B^0 \sim \mathcal{N}(0, \sigma_0^2)$, $B^1 \sim \mathcal{N}(0, \sigma_1^2)$ ($\sigma_0 > \sigma_1$), and degree of translationality $\tau$, then in element-wise $(i,j)$, $B^\tau \sim \mathcal{N}(0, \sigma^2)$ can be represented as a transition state between $B^0$ and $B^1$:

$$B^\tau(i,j) \propto (B^0(i,j))^{\frac{\tau^2}{\sigma_0^2}} (B^1(i,j))^{\frac{\tau^2}{\sigma_1^2}}$$  \hspace{1cm} (6)

where $\sigma^2 = (1 - \tau) \sigma_0 + \tau \sigma_1$. See proofs in Appendix 1.

Specifically, when a primary degradation $B^0$ is blur-free ($\sigma_0 = 0^+$), the transition state $\{x^\tau\}$ are formulated as

$$x^\tau \propto x \otimes (B^1)^{\frac{1}{\sigma_1^2}}$$  \hspace{1cm} (7)

namely, the estimation of transition states is simplified as a variant of $B^1$.

Furthermore, if $\sigma_0 \neq 0$, the transitional degradation is a non-linear combination of $B^0$ and $B^1$, and the transition state is transformed into

$$\hat{y} \propto y \otimes \left((B^0)^{\frac{\sigma^2}{\sigma_0^2}}(B^1)^{\frac{\sigma^2}{\sigma_1^2}}\right)$$  \hspace{1cm} (8)

Therefore, as Fig. 2 shows, both additive and convolutive degradations are transitional. Thus, it is feasible to represent an unknown degradation by the transition state between two or more primary degradations.

### 3.2 Transitional Learning

To solve the SR problem of a LR image $x$, it allows learning the energy function by Maximum A Posteriori (MAP) as

$$\hat{y} = \operatorname{arg\ max}_y \log P(y|x)$$  \hspace{1cm} (9)

particularly, $P(x|y)$ indicates the degradation prior and is commonly unknown in blind tasks.

To solve this issue, if we assume that the combination of primary states $\{x^i\}_{i=0}^{K-1}$ is a complete set for representing an unknown state $x$, then as $\{x^i\}_{i=0}^{K-1}$ are conditionally independent, we would formulate the transitional degradation as

$$P(x|y) = P(x_0, x_1, ..., x^{K-1}|y) = \prod_{i=0}^{K-1} P(x^i|y)$$  \hspace{1cm} (10)

and

$$\hat{y} = \operatorname{arg\ max}_y \sum_{i=0}^{K-1} \log P(x^i|y) + \log P(y)$$  \hspace{1cm} (11)

so we can infer the optimal $\hat{y}$ under the prior of $K$ given primary degradation priors $\{P(x^i|y)\}_{i=0}^{K-1}$.

Therefore, when the degradation fulfills transitionality as illustrated in Section 3.1, although without precise information of the unknown degradation $B$ and $N$, it is feasible to infer a desired super-resolved result by finite independent primary degradations $\{B^i\}_{i=0}^{K-1}$ and the corresponding DoT $\{\tau_i\}_{i=0}^{K-1}$.
results $y$ can not be estimated by explicitly linear interpolation of $y^0 = \mathcal{F}(x^0; \Theta^0)$ and $y^1 = \mathcal{F}(x^1; \Theta^1)$. Therefore, in conventional deep-learning-based SR methods, a well-designed CNN model is generally applied to super-resolve $x^\tau$ as $\mathcal{F}(x^\tau; \Theta^\tau) : x^\tau \to y$, but it is inefficient to train numerous models for all $\tau \sim U(0, 1)$ in practical applications.

In particular, inspired by the network interpolation [53], which uses the linear interpolation of convolutional layers to rebuild a new network, we revisit it and exploit its feasibility to the non-linear representation of transitional state. Thus, given two convolutional layers $\mathcal{F}^0(\cdot; \Theta^0)$ and $\mathcal{F}^1(\cdot; \Theta^1)$, we introduce a transitional learning module to derive a transitional model $\mathcal{F}^\tau(\cdot; \Theta^\tau)$ using a transitional transformation function $\Upsilon$. In short, by utilizing the network interpolation in Fig. 4 we have $\Theta^\tau = \tau \Theta^0 + (1 - \tau) \Theta^1$.

Specifically in convolutional layers (with no bias), assume that the transitional LR image $x^\tau$ can be accurately estimated by the interpolation of $x^0$ and $x^1$. We then infer the transitional results by feeding $x^\tau$ into such transitional network as

$$\hat{y}_j(p_o) = \sum_{i=1}^{c_{in}} \sum_{p \in \mathcal{R}} x^0_i(p_o + p) \cdot w^0_{j,i}(p)$$

$$= \sum_{i=1}^{c_{in}} \sum_{p \in \mathcal{R}} (\tau x^0_i(p_o + p) + (1 - \tau) x^1_i(p_o + p)) \cdot (\tau w^0_{j,i}(p) + (1 - \tau) w^1_{j,i}(p))$$

(15)

where $w$ represents the kernel weights of convolution, moreover, $c_{in}$ is the number of channels in input images (or features), $p_o$ is the initial position in images/features, and $p$ is the position of convolution in receptive filed $\mathcal{R}$ respectively.

More generally, from Eq.(14), the transition state $x^\tau$ can be accurately estimated by the interpolation of $x^0$ and $x^1$, we then infer the results by feeding $x^\tau$ into such transitional model as

$$\hat{y} = \mathcal{F}^\tau(x^\tau; \Theta^\tau)$$

$$= \tau^2 \mathcal{F}^0(x^0; \Theta^0) + (1 - \tau) \mathcal{F}^1(x^1; \Theta^1) +$$

$$\tau (1 - \tau) x^0_i(p_o + p) \cdot w^0_{j,i}(p) +$$

$$\tau (1 - \tau) x^1_i(p_o + p) \cdot w^1_{j,i}(p) +$$

$$\tau (1 - \tau)^2 x^1_i(p_o + p) \cdot w^1_{j,i}(p) +$$

(16)

this transitional model shows a substantial capacity of feature representation with more than 4 SSISR effects.

### 3.2.2 Convolutional degradation $B$

As in Eq.(3), the transition state of convolutive degradations $B^\tau$ can be explicitly represented by two primary degradations $B^0$ and $B^1$. Similar to the transitional transformation function of additive degradations in Eq.(13), we expect to explore a transitional transformation function to approximate the transition state in convolutive degradations.

However, in Eq.(3), unlike the additive degradations, a transition state of convolutive degradation is a non-linear transformation of the primary states. Intuitively, as in Fig. 3...
we design a deep network with non-linear activation to approximately fit this transitional transformation function as

$$y = \mathcal{F}_m^\tau(\cdots \mathcal{S}(\mathcal{F}_1^\tau(x^\tau; \Theta_1^\tau)); \Theta_m^\tau)$$

(17)

where \( \mathcal{S} \) represents a non-linear activation function, e.g., ReLU. For generalization, we use this non-linear network to fit the transitional transformation function for both additive and convolutive degradations.

Thus, for a mini-batch of size \( B \), the objective function of our TSLR is formulated as

$$\mathcal{L}_{\text{TSLR}} = \frac{1}{B} \sum_{b=1}^{B} \| \mathcal{F}_{\text{TSLR}}(x_b; \Theta_{\text{TSLR}}^\tau) - y_b \|_1$$

(18)

where \( y_b \) denotes the HR ground truth.

Specifically, as Fig. 3 shows, the LR images are firstly fed into a homogeneous feature extraction network and parameter-shared for degradations with any DoTs. Then, the homogeneous features are fed into the transitional learning module, which stacks \( m \) transitional transformation functions to infer an adaptive model for the transitional degradation. Furthermore, since the tenability of this conclusion is under the assumption that the transition state can be well represented using the real DoT, it is crucial to estimate DoT \( \tau \), and make the transitional learning module more adaptive to blind SR task.

### 3.3 DoT Estimation

As above, the transitional model can be formulated as the interpolation of finite primary models. Nevertheless, since the real DoT \( \tau \) is generally unknown in blind SR tasks, estimating the desired transitional model is difficult. Thus in this section, we mainly describe an effective way to estimate an appropriate DoT \( \hat{\tau} \), the architecture is shown in Fig. 3 and named DoTNet.

**3.3.1 Random cropping**

Since DoTNet is designed as a score regression network with fully connected layers (FC) in tail, fixed-size image patches need to be fed into the DoTNet. Therefore, we first introduce a random cropping module to obtain a group of fixed-sized patches from the input LR image. The reasons for patch-cropping are that: (1) in practice, the image sizes are generally unified and (2) either convolutive or additive degradation acts on a local receptive field, then each local patch in a single image commonly have the same representation for degradation.

Specifically, taking convolutive degradations (e.g., Gaussian blur effects) into consideration, Gaussian blur kernel of size \( w \) is sliding on each pixel of patch. Since the convolutive degradation has a sliding receptive field of size \( w \), to capture the pattern of degradation effects, we need fixed-size image patches with \( \min(\text{width, height}) \geq w \). Besides, as the considered AWGN is independent of the image and acts on isolated pixels, we only need \( \min(\text{width, height}) \geq 1 \) to capture the pattern of additive degradations.

### 3.3.2 DoTNet

Given groups of LR image patches with variational degradations \( \{\{x_{b,t}\}_{t=1}^{T}\}_{b=1}^{B} \), the output DoT is obtained by capturing the activation in the last layer as \( \hat{\tau} = \mathcal{F}_{\text{DoT}}(x; \Theta_{\text{DoT}}) \). We then train the parameters \( \Theta_{\text{DoT}} \) by optimizing the objective function:

$$\mathcal{L}_{\text{DoT}} = \frac{1}{B} \sum_{b=1}^{B} \sum_{t=1}^{T} \| \mathcal{F}_{\text{DoT}}(x_{b,t}; \Theta_{\text{DoT}}) - \tau_b \|_1$$

(19)

where \( \tau_b \) indicates the real DoT of all the patches \( \{x_{b,t}\}_{t=1}^{T} \) from \( b \)-th LR image, and is preset by inferring the transition state mathematically as Section 3.1.

Specifically, we design the DoT estimation module as an assessment-aware score regression network [51] as in Fig. 3 where the feature extraction part (backbone) stacks 4 residual bottleneck blocks [52] with pooling operations. Moreover, in the decision part, we stack two fully connected layers and a Sigmoid function with a one-dimension output to estimate the DoT \( \hat{\tau} \).

### 3.4 Discussions on Transitional Learning

#### 3.4.1 Difference to integrated learning

Integrated learning based SR methods aim at training a single integrated model with multiple prior degradations [22] and improving the generalization of a model with the capability of processing all the prior degradation effectively. For blind SR, the IL-BSR methods [25], [29] additionally exploit a degradation prediction model to correct the unseen degradation into a prior degradation for SR inference. However, as only need to train two primary degradations and exploit the transitionality of degradations to infer the unknown transition states, the proposed transitional learning is more flexible than the IL-BSR methods which need to cover all the prior degradations. In fact, for these IL-BSR methods, a higher dimension of degradation representation indicates higher degree of freedom, and it is challenging to learn an accurate solution.

Furthermore, these IL-BSR methods aim to explore a large solution space to cover the problem space of multiple degradations, but aimlessly enlarging the solution space is lack of flexibility and inefficient. However, our TSLR attempts to explore an adaptive and new solution space by an interpretable nature of the transitional degradations.

#### 3.4.2 Difference to zero-shot learning

Zero-shot learning based SR methods depend on the internal samples of a single image, and tend to train an image-specific model with a specific prior degradation. For blind SR, the ZSL-BSR methods [28], [29], [30], [41] aim at iteratively optimizing the model with a specific initial degradation to adapt the internal samples with unseen degradations, so they suffer from redundant iterations and are highly dependent on the initial degradation. However, the proposed transitional learning is an end-to-end network without additional iterations and only needs the type of degradation, which is more efficient than the ZSL-BSR.
4 Experiments

4.1 Datasets and Evaluation

In the training phase, high-quality images with 2K resolution from DIV2K and Flickr2K [53] datasets are considered as HR ground truth, and are then downsampled with specific degradations (e.g., bicubic, isotropic Gaussian blur, AWGN) to generate the corresponding LR images. In the testing phase, we conduct the experiments on several benchmark datasets for evaluation, including Set5 [54], Set14 [5], BSD100 [55], and Urban100 [56] with general and structural scenes.

To evaluate SR performance, we apply two common full-reference image quality assessment criteria to evaluate the discrepancies of super-resolved results and HR ground truth, i.e., peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). For super-resolution, the luminance channel is selected for full-reference image quality assessment because the intensity of the image is more sensitive to human vision than the chroma. Furthermore, following [57], the inference time is calculated via torch.cuda.Event to evaluate the computational complexities on GPU.

4.2 Implementation Details

As in Fig. 3, we set $n = 32$ residual blocks [16] with $c = 128$ channels for homogeneous feature extraction, and sample $T = 8$ local patches with random cropping boxes of size $32 \times 32$ for DoT estimation. Furthermore, we feed the estimated DoTs and homogeneous features into the transitional learning module which stacks $m = 8$ residual blocks for network interpolation, and finally use sub-pixel convolutions [10] for feature upsampling.

As discussed in Section 3.1 and following [27], two typical degradations are applied to generate the corresponding LR images, including convolutive degradations (isotropic Gaussian blur kernels of size $21 \times 21$, where $\sigma \sim U(0.2, 2.0)$ for $\times 2$ upsampling and $\sigma \sim U(0.2, 4.0)$ for $\times 4$ upsampling), and additive degradations (AWGN with noise levels $N \sim U(0, 30)$). As illustrated in Eq. 3, since the additive degradation $N$ is independent of the convolutive degradation $B$, we consider them separately in this paper.

All models are optimized using the Adam optimizer [58] with a mini-batch of size $B=16$, and each mini-batch contains $B$ paired groups of $48 \times 48$ RGB LR patches and the corresponding $48s \times 48s$ HR ground truth for $\times s$ upsampling. These LR-HR pairs are augmented with random horizontal and vertical flips and 90 rotations. All LR and HR images are preprocessed by subtracting the mean RGB value of $48 \times 48$ paired groups of $B$ for either additive or convolutive degradations.

4.3 Effect of Transitional Learning

As illustrated in Fig. 4, the transitional learning is built on at least two primary models with $\Theta^0$ and $\Theta^1$. Remarkably, each primary model acts as a specific SISR baseline, with a specific solution space for a given degradation, and shows strong performance on its specific degradation but weak capacities on other unseen degradations.

To investigate the effects of transitional learning, we conduct two baseline models without transitional learning for either additive or convolutive degradations:

1) baseline-$D^0$: $\hat{y} = F_m^0 \left( ... S(F^0(x^\tau; \Theta^0)); \Theta^0 \right)$
2) baseline-$D^1$: $\hat{y} = F_m^1 \left( ... S(F^1(x^\tau; \Theta^1)); \Theta^1 \right)$

where $D$ indicates the degradation type, i.e., $D$ represents $B$ for isotropic Gaussian blur, or $N$ for AWGN. Moreover, these baseline models are trained on their corresponding degradations.

From Fig. 5, each baseline performs the best on the strictly corresponding degradation but gets worse performances on other degradations. For example, baseline-$D^0$ (additive degradations) achieves higher PSNR on $N^\tau = 0$, reaching about 4dB gains against on $N^\tau = 30$. By applying the proposed transitional learning, the final TLSR obtains substantial improvements for various degradations. Specifically, our TLSR (w/ $\tau$) closes the gap of performances between $N^\tau = 0$ and $N^\tau = 30$ from 7dB to 2.6dB.

Furthermore, since the real DoTs ($\tau$) are not given in blind applications, the effects of transitional learning would
depend on the accuracy of DoT estimation. Nevertheless, as only a 1-D scalar should be estimated, our DoTNet is capable enough. Although without the real DoTs, our final model TLSR with the estimated DoTs ($\hat{\tau}$) also shows comparable performances as the non-blind TLSR ($w/\tau$).

### 4.4 Effect of DoT Estimation

As illustrated above, since the DoT estimation network is designed to estimate the degree of transitionality and affects the effectiveness of the transitional learning, we then conduct a visualization experiment on the distribution of the estimated DoTs to illustrate the effect of DoTNet. As in Fig. 6(a), even though the degradation process is acting on the local regions in a single image, we can still accurately capture the DoTs by applying the random cropping and DoTNet. Nevertheless, as we mainly feed the primary degradations as the baselines and consider the transitional degradations as transition states in the training phase, the estimated DoTs are more accurate in the primary states but somewhat discrepancy in the transition states.

Therefore, to further illustrate the effect of DoT estimation, we visualize the activation responses of DoT on the different types of degradations using GradCAM [59]. As shown in Fig. 6(b), the low-level response of DoT on additive degradation focuses on the smoothing regions of the image, which are easier to detect the noises. Contrastively, the high-level GradCAM responses of DoTs on convolutive degradations focus on the textures, which are easier to detect the blur effects. Particularly, these responses are consistent to the mechanism of human vision system, that noises in the smooth region and blurry effects in the complex textures have higher saliencies for visual perception.

### 4.5 Experiments on Unmixed Degradations

In this subsection, we evaluate quantitative and qualitative comparisons on the unmixed synthetic (additive or convolutive) degradations, and the comparative methods are

1) non-blind SISR method, as RCAN [17];
2) integrated learning based blind SR methods (IL-BSR), including DnCNN [24], VDN [25], IKC [26] and DASR [27];
3) zero-shot learning based blind SR methods (ZSL-BSR), including DIP [29], ZSSR [28], KernelGAN [30] and MZSR [31].

Specifically, VDN and KernelGAN are the representative blind methods for noise modeling and kernel estimation, respectively, and we naturally apply the ZSSR with the estimated noises or kernels for blind SR as in [30]. Besides, for general blind SR: 1) IKC iteratively estimates and corrects the representation vector of blur kernel to guide a conditional integrated model for blind deblurring and super-resolution; 2) DASR directly predict a discriminative degradation representation vector to guide a conditional integrated model, which is efficient as a non-iterative blind SR method; 3) DIP is an image-specific heuristic method to progressively explore an ideal transition state between random noises and the corrupted LR images; 4) ZSSR is a representative zero-shot

---

1. As specified in [29], on synthetic degradations, we use the ground-truth to select the number of iterations with the best PSNR, and on unknown degradations, we fix the number of iterations to 2000.
TABLE 1: Quantitative comparisons of the proposed TLSR and SOTAs on blind SR with additive degradations for ×4 upscaling. Particularly, VDN [25] is a blind noise modelling method, we then feed the denoised image into RCAN [17] for SR, or use only the S-Net [25] of VDN to estimate the noises for post-processing of SR using either ZSSR [28] or SRMD [22]. The inference time (sec) is averagely calculated on Set14 dataset.

| Scale | Method | Time↓ (sec) | Set5 PSNR↑ | Set5 SSIM↑ | Set14 PSNR↑ | Set14 SSIM↑ | BSD100 PSNR↑ | BSD100 SSIM↑ | Urban100 PSNR↑ | Urban100 SSIM↑ |
|-------|--------|------------|-----------|-----------|-------------|-------------|---------------|---------------|---------------|---------------|
|       | Bicubic | -          | 29.28 0.7437 | 27.44 0.7034 | 26.92 0.6749 | 25.11 0.6866 |
|       | RCAN [17] | 0.139      | 27.38 0.6027 | 26.11 0.5981 | 25.47 0.5786 | 22.20 0.5286 |
|       | DIP [29] | 143.6      | 28.71 0.7911 | 26.45 0.6862 | 26.20 0.6599 | 25.02 0.6698 |
|       | ZSSR [28] | 82.15      | 27.62 0.6265 | 26.17 0.6163 | 25.56 0.5937 | 24.31 0.6158 |
|       | VDN [25]+ZSSR [28] | 82.16      | 31.19 0.8659 | 29.15 0.8065 | 27.92 0.7568 | 27.40 0.8174 |
|       | VDN [25]+MZSR [31] | 0.744      | 31.10 0.8177 | 28.80 0.7664 | 27.95 0.7390 | 27.41 0.7865 |
|       | S-Net [25]+SRMD [22] | 0.023     | 32.40 0.8782 | 29.82 0.8039 | 28.80 0.7649 | 28.17 0.8327 |
|       | DnCNN [24]+RCAN [17] | 0.146      | 32.94 0.8443 | 29.59 0.8137 | 29.39 0.8679 |
|       | TLSR (ours) | 0.039      | 33.67 0.9063 | 30.95 0.8488 | 29.73 0.8166 | 29.77 0.8770 |
|       | Bicubic | -          | 26.22 0.6904 | 24.58 0.5998 | 24.45 0.5614 | 22.23 0.5606 |
|       | RCAN [17] | 0.086      | 24.90 0.5798 | 23.62 0.5221 | 23.09 0.4824 | 22.20 0.5286 |
|       | DIP [29] | 144.2      | 26.55 0.7280 | 24.68 0.6095 | 24.51 0.5713 | 22.60 0.5648 |
|       | ZSSR [28] | 88.06      | 25.48 0.6183 | 23.81 0.5319 | 23.54 0.4857 | 22.49 0.5111 |
|       | VDN [25]+ZSSR [28] | 88.11      | 28.29 0.7918 | 26.13 0.6932 | 25.72 0.6512 | 23.84 0.6899 |
|       | VDN [25]+MZSR [31] | 0.578      | 28.21 0.7975 | 26.18 0.6927 | 25.76 0.6511 | 23.68 0.6811 |
|       | S-Net [25]+SRMD [22] | 0.009     | 28.45 0.8049 | 26.20 0.6831 | 25.64 0.6348 | 23.94 0.6862 |
|       | DnCNN [24]+RCAN [17] | 0.094      | 28.62 0.8105 | 26.41 0.6989 | 25.96 0.6562 | 24.21 0.6994 |
|       | VDN [25]+RCAN [17] | 0.092      | 29.07 0.8238 | 26.73 0.7118 | 26.09 0.6646 | 24.63 0.7188 |
|       | DASR [27] | 0.033      | 28.73 0.8252 | 26.55 0.7122 | 25.99 0.6648 | 24.16 0.7111 |
|       | TLSR (Ours) | 0.025      | 29.42 0.8349 | 27.01 0.7216 | 26.22 0.6706 | 24.91 0.7348 |

Fig. 7: Qualitative visual comparisons of the proposed TLSR and state-of-the-arts blind SR methods for ×4 upscaling on “img004” and “img099” from Urban100 dataset with additive degradations. Our TLSR reconstructs the edges and structures better than other methods.
TABLE 2: Quantitative comparisons of the proposed TLSR and SOTAs on blind SR with convolutive degradations for ×4 upscaling. Particularly, KernelGAN [30] is a kernel estimation method for blind image deblurring, so we apply either ZSSR [28] or SRMD [22] with the estimated kernel for post-processing of SR. The inference time (sec) is averagely calculated on Set14 dataset.

| Scale D | Method | Time | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ | PSNR↑ | SSIM↑ |
|---------|--------|------|-------|-------|-------|-------|-------|-------|-------|-------|
| ×2 | Bicubic | - | 30.95 | 0.8778 | 28.21 | 0.7954 | 27.79 | 0.7633 | 25.04 | 0.7592 |
| | RCAN [17] | 0.086 | 33.09 | 0.9001 | 30.03 | 0.8307 | 29.22 | 0.8019 | 27.68 | 0.8155 |
| | DIP [29] | 143.6 | 31.70 | 0.8782 | 28.42 | 0.7768 | 27.90 | 0.7486 | 25.04 | 0.7197 |
| | ZSSR [28] | 82.15 | 32.52 | 0.8960 | 29.45 | 0.8232 | 28.60 | 0.7923 | 26.03 | 0.7901 |
| | KernelGAN [30] + ZSSR [28] | 374.9 | 33.64 | 0.9301 | 29.95 | 0.8615 | 29.32 | 0.8403 | 27.31 | 0.8478 |
| | KernelGAN [30] + MZSR [31] | 293.5 | 34.03 | 0.9252 | 30.48 | 0.8597 | 29.64 | 0.8328 | 27.15 | 0.8345 |
| | KernelGAN [30] + SRMD [22] | 292.7 | 33.12 | 0.9244 | 29.84 | 0.8680 | 28.99 | 0.8449 | 27.11 | 0.8456 |
| | IKC [26] | 0.428 | 36.38 | 0.9471 | 32.39 | 0.8923 | 31.05 | 0.8674 | 29.63 | 0.8833 |
| | DASR [27] | 0.043 | 36.80 | 0.9492 | 32.49 | 0.8934 | 31.23 | 0.8710 | 30.18 | 0.8978 |
| | TLSR (ours) | 0.039 | 37.23 | 0.9520 | 33.08 | 0.9019 | 31.57 | 0.8793 | 30.95 | 0.9092 |

Fig. 8: Qualitative comparisons of the proposed TLSR and state-of-the-arts blind SR methods for ×4 upscaling on “img079” and “img096” from Urban100 dataset with convolutive degradations. Our TLSR has fewer distortions and better structures/edges than other methods.
Ascribing to the vital capacity of transitional learning to rebuild transitional networks, TLSR is more flexible and capable of processing various degradations as long as the degradations are transitional. Therefore, as reported in TABLE 1 and TABLE 2, without any additional iterative operations for correcting $\hat{\tau}$, our TLSR achieves best performances on both additive and convolutive degradations, and is faster than most of the comparative SOTA methods. For example, for $\times 2$ upscaling on Set14 dataset with convolutive degradations, our TLSR gains PSNR of 0.69dB and 3.63dB, and is $\times 10$ and $\times 2000$ times faster than IKC and ZSSR, respectively. Moreover, from the detailed quantitative comparison on computational complexities reported in TABLE 3, our TLSR is with relatively the fewest complexities and fastest speed. Besides, the qualitative comparisons (for $\times 4$ upscaling) are shown in Fig. 7 and Fig. 8. Obviously, for the additive degradations $N \in \{0, 15, 30\}$, our TLSR reconstructs the noisy edges and structures better than other methods, especially for $N \geq 15$, the existing state-of-the-art blind SR methods tend to generate more distortions and smoothen the details. For the convolutive degradations $B \in \{0.2, 2.0, 4.0\}$, our TLSR shows higher fidelity of color and details against other blind SR methods, particularly limited by the iterative operations, IKC shows comparative performance but are far slower than our TLSR.

**4.6 Experiments on Mixed Degradations**

Besides the unmixed synthetic degradations, we further conduct experiments on more complex (real-world) degradations. Since the degradation of the real-world image is generally unknown and mixed, but our TLSR models are trained on a single type of transitional degradations (e.g., F_TLSR$_N$ for additive degradations, and or F_TLSR$_B$ for convolutive degradations). As in Eq. (3) and following [27], since the additive degradation $N$ is independent of the convolutive degradation $B$, it is feasible to plug-in the denoiser as a preprocessing module.
TABLE 4: Quantitative comparisons (PSNR↑) of the proposed TLSR<sub>real</sub> and SOTAs on blind SR with mixed synthetic and unknown real-world degradations for ×4 upscaling.

| Dataset | Method          | Timeʃ | N<sup>x0</sup> | θ | N<sup>x5</sup> | θ | N<sup>x10</sup> | θ | N<sup>x10</sup> | θ | N<sup>x10</sup> | θ | N<sup>x10</sup> | θ | N<sup>x10</sup> | θ |
|---------|-----------------|-------|----------------|---|---------------|---|-----------------|---|-----------------|---|-----------------|---|-----------------|---|-----------------|---|
| Set14   | ZSSR [28]       | 88.23 | 27.33          | 26.59 | 25.18         | 26.98         | 26.30         | 24.96 | 25.53         | 25.07 | 24.06         | 25.07 | 24.06         | 25.07 | 24.06         |
|         | DIP [29]        | 144.2 | 27.11          | 26.33 | 25.52         | 26.82         | 26.12         | 25.27 | 25.46         | 24.92 | 24.39         | 24.92 | 24.39         | 24.92 | 24.39         |
|         | DnCNN [24]+IKC [26] | 0.221 | 27.90          | 27.36 | 26.27         | 28.00         | 27.53         | 26.60 | 27.59         | 26.12 | 25.61         | 26.12 | 25.61         | 26.12 | 25.61         |
|         | VDN [25]+IKC [26] | 0.227 | 28.36          | 27.73 | 26.95         | 28.40         | 27.60         | 26.83 | 27.91         | 26.40 | 25.84         | 26.40 | 25.84         | 26.40 | 25.84         |
|         | DASSR [27]      | 0.033 | 28.08          | 27.66 | 27.07         | 28.15         | 27.58         | 26.91 | 27.79         | 26.79 | 26.08         | 26.79 | 26.08         | 26.79 | 26.08         |
|         | TLSR<sub>real</sub>(Ours) | 0.051 | 28.79          | 27.80 | 27.04         | 28.73         | 27.68         | 26.87 | 28.65         | 26.95 | 26.09         | 26.95 | 26.09         | 26.95 | 26.09         |

TABLE 5: Quantitative comparisons (PSNR↑) of the proposed TLSR and SOTAs on blind SR with anisotropic Gaussian blur degradations for ×4 upscaling on Urban100 dataset.

| Method | Anisotropic Blur Kernel |
|--------|-------------------------|
| Bicubic | ![Icon](icon1.png)       |
| ZSSR [28] | ![Icon](icon2.png)       |
| DIP [29] | ![Icon](icon3.png)       |
| IKC [26] | ![Icon](icon4.png)       |
| DASSR [27] | ![Icon](icon5.png)       |
| TLSR (Ours) | ![Icon](icon6.png)       |

Therefore, we build a model as TLSR<sub>real</sub> by firstly using TLSR<sub>x</sub> for denoising and then feeding the denoised image x′ into the TLSR<sub>y</sub> for deblurring and super-resolution:

\[
x' = \mathcal{F}_{\text{TLSR}_x}(x; \Theta_{\text{DoT}_x}(x), \uparrow 1)
\]

\[
y = \mathcal{F}_{\text{TLSR}_y}(x'; \Theta_{\text{DoT}_y}(x'), \uparrow 4)
\]

Additionally, as illustrated in Section 4.4, the deblurring and denoising tasks have opposite effects on DoT representation. Thus, it is feasible to solve the complex degradations using this two-stage framework.

As in Tables 4 and Fig. 9, ascribing to the flexible transitional learning, on both complex degradations and unknown real-world degradations, our TLSR<sub>real</sub> achieves comparable performances against other blind SR methods. Nevertheless, as a two-stage framework, TLSR<sub>real</sub> costs twice of a single TLSR framework, and it is promising to solve DoT estimation on complex degradations for implementing one-stage TLSR<sub>real</sub>.

5 DISCUSSIONS

Apart from the aforementioned additive and convolutive degradations, other types of degradation also exist in realistic scenes. In this section, we mainly investigate the intermediate variants of a single type of degradation experimentally, but the corresponding theoretical demonstrations are expected for further exploring.

5.1 Anisotropic Gaussian Blur

Different from the isotropic Gaussian kernel, anisotropic Gaussian kernel is more flexible to represent the blur effects by a linear moving camera, which is formulated as \( B = \mathcal{N}(0, \Sigma) \). Unlike a single scalar \( \sigma \) for representing isotropic Gaussian kernels, \( \Sigma \) indicates a covariance matrix determined by two scalars \( (\sigma_u, \sigma_v) \) and a rotation angle \( \theta \). Since \( \Sigma \) has higher degree of freedom and is too complicated for the present TLSR with a single DoT, we then fix \( \sigma_u = 1.3, \sigma_v = 3.25 \) and randomly sample the rotation angles \( \theta \sim \mathcal{U}(0, \pi/2) \). The other implementation details are the same as Section 4.2.

As reported in Table 3, unlike IKC and DASSR that estimate a high-dimension representation vector to distinguish different degradations, our TLSR explores only a 1-D scalar \( \tau \) to represent the different rotation angles, which is easier to train and achieves the best performance on the Urban100 dataset with \( \theta \in \{0, \pi/6, \pi/4, \pi/3, \pi/2\} \). For qualitative comparisons, we post the visual comparisons as shown in Fig. 10, it is observed that the degraded LR images suffer from seriously motion blurry effects in v-axis, and only our TLSR could restore them well for all the rotation angles.

5.2 JPEG Compression Artifacts

In practice, especially social media, JPEG compression is utilized to speed up image transmission, leading to the loss
Fig. 10: Qualitative comparisons of the proposed TLSR and state-of-the-arts blind SR methods on “img002”, “img052” and “img072” from Urban100 dataset for ×4 upscaling with different anisotropic Gaussian blur kernels.

Fig. 11: Qualitative comparisons of the proposed TLSR and SOTAs on “buildings”, “sailing3” and “carnivaldolls” from LIVE1 dataset for JPEG compression artifacts removal with Q=10, Q=20 and Q=30, respectively.

of high-frequency details. Fortunately, since JPEG compression is based on the non-overlapping 8 × 8 pixel blocks and is determined by a single scalar of quality factor $Q$, the DoT estimation with random cropping boxes of size $\min(\text{width}, \text{height}) \geq 8$ is appropriate enough. Thus, same as Section 4.2, we also set the box size to 32 × 32 but remove the sub-pixel convolutions for upscaling, and random sample the quality factors $Q \sim U(10, 30)$ to train a TLSR model for blind compression artifacts removal.

As reported in TABLE 6, our TLSR achieves the best performance on both Classic5 and LIVE1 datasets for various JPEG compression artifacts removal. For qualitative comparisons, we post the visual comparisons as shown in Fig. 11, it is observed that the compressed images suffer from seriously smoothing in each 8 × 8 blocks. After applying our TLSR, both the flattened regions and textures are restored well.

6 Conclusion

In this paper, we firstly analyze and demonstrate the transitionality of degradations theoretically, which indicates that any transitional degradation can be well represented by two or more primary degradations. Under this condition, we then propose a transitional learning method for blind super-resolution on both additive degradation (AWGN) and convolutive degradations (isotropic Gaussian blur kernel). Specifically, the SR performance and generalization are significantly improved by introducing the DoT estimation and constructing an adaptively transitional transformation function using transitional learning under the guidance of the estimated DoT. Theoretical and experimental demonstrations indicate that the proposed TLSR achieves superior performance on synthetic unmixed/mixed degradations and
unknown real-world degradation, and consumes relatively less time against the state-of-the-art blind SR methods. Furthermore, we also discuss the performance of our TSLR on blind SR with anisotropic Gaussian blur kernel and blind JPEG compression artifacts removal tasks, but the corresponding theoretical demonstrations on the transitionality of these degradations are expected for further exploring.

References

[1] R. Keys, “Cubic convolution interpolation for digital image processing,” IEEE Trans. Acoustics, Speech, and Signal Process., vol. 29, no. 6, pp. 1153–1160, 1981.
[2] A. Marquina and S. Osher, “Image super-resolution by tv-regularization and bregman iteration,” J. Sci. Comput., vol. 37, no. 3, pp. 367–382, Dec. 2008.
[3] W. Dong, L. Zhang, G. Shi, and X. Wu, “Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization,” IEEE Trans. Image Process., vol. 20, no. 7, pp. 1836–1857, Jul. 2011.
[4] J. Yang, J. Wright, T. S. Huang, and Y. Ma, “Image super-resolution via sparse representation,” IEEE Trans. Image Process., vol. 19, no. 11, pp. 2861–2873, 2010.
[5] R. Zeyde, M. Elad, and M. Protter, “On single image scale-up using sparse-representations,” in Curves and Surfaces, 2010.
[6] Y. Hu, N. Wang, D. Tao, X. Gao, and X. Li, “Serf: A simple, effective, robust, and fast image super-resolver from cascaded linear regression,” IEEE Trans. Image Process., vol. 25, no. 9, pp. 4091–4102, 2016.
[7] Y. Huang, J. Li, X. Gao, L. He, and W. Lu, “Single image super-resolution via multiple mixture prior models,” IEEE Trans. Image Process., vol. 27, no. 12, pp. 5904–5917, 2018.
[8] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Eur. Conf. Comput. Vis., 2014, pp. 184–199.
[9] C. Dong, C. C. Loy, K. He, and X. Tang, “Image super-resolution using deep convolutional networks,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 38, no. 2, pp. 295–307, 2016.
[10] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, and Z. Wang, “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 1874–1883.
[11] C. Dong, C. C. Loy, and X. Tang, “Accelerating the super-resolution convolutional neural network,” in Eur. Conf. Comput. Vis., 2016, pp. 491–497.
[12] Y. Tai, J. Yang, X. Liu, and C. Xu, “Memnet: A persistent memory network for image restoration,” in Int. Conf. Comput. Vis., 2017, pp. 4549–4557.
[13] T. Tong, G. Li, X. Liu, and Q. Gao, “Image super-resolution using dense skip connections,” in Int. Conf. Comput. Vis., 2017, pp. 4799–4807.
[14] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, “Photo-realistic single image super-resolution using a generative adversarial network,” in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 105–114.
[15] Y. Tai, J. Yang, and X. Liu, “Image super-resolution via deep recursive residual network,” in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 3147–3155.
[16] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, “Enhanced deep residual networks for single image super-resolution,” in IEEE Conf. Comput. Vis. Pattern Recog. Worksh., 2017, pp. 136–144.
[17] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, “Image super-resolution using very deep residual channel attention networks,” in Eur. Conf. Comput. Vis., 2018, pp. 294–310.
[18] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image super-resolution,” in IEEE Trans. Comput. Vis. Pattern Recog., 2018, pp. 2472–2481.
[19] Y. Hu, J. Li, Y. Huang, and X. Gao, “Channel-wise and spatial feature modulation network for single image super-resolution,” IEEE Trans. Circuits Syst. Video Technol., vol. 30, no. 11, pp. 3911–3927, 2020.
[20] Y. Huang, J. Li, X. Gao, Y. Hu, and W. Lu, “Interpretable detail-attentiveness network for single image super-resolution,” IEEE Trans. Image Process., vol. 30, pp. 2325–2339, 2021.
[21] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, “Residual dense network for image restoration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 43, no. 7, pp. 2480–2495, 2021.
[22] K. Zhang, W. Zuo, and L. Zhang, “Learning a single convolutional super-resolution network for multiple degradations,” in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 3262–3271.
[23] Y.-S. Xu, S.-Y. R. Tseng, Y. Tseng, H.-K. Kuo, and Y.-M. Tsai, “Unified dynamic convolutional network for super-resolution with variational degradations,” in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 12493–12502.
[24] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, “Beyond a gaussian denoiser: Residual learning of deep cnns for image denoising,” IEEE Trans. Image Process., vol. 26, no. 7, pp. 3142–3155, 2017.
[25] Z. Yue, H. Yong, Q. Zhao, D. Meng, and L. Zhang, “Variational denoising network: Toward blind noise modeling and removal,” in Adv. Neural Inform. Process. Syst., 2019.
[26] J. Gu, H. Lu, W. Zuo, and C. Dong, “Blind super-resolution with iterative kernel correction,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 1604–1613.
[27] L. Wang, Y. Wang, X. Dong, Q. Xu, J. Yang, W. An, and Y. Guo, “Unsupervised degradation representation learning for blind super-resolution,” in IEEE Conf. Comput. Vis. Pattern Recog., 2021, pp. 10581–10590.
[28] A. Shocher, N. Cohen, and M. Irani, “Zero-shot super-resolution using deep internal learning,” in IEEE Conf. Comput. Vis. Pattern Recog., 2018, pp. 3118–3126.
[29] D. Ulyanov, A. Vedaldi, and V. Lempitsky, “Deep image prior,” in IEEE Conf. Comput. Vis. Pattern Recog., 2017, pp. 9446–9454.
[30] S. Bell-Kligler, A. Shocher, and M. Irani, “Blind super-resolution kernel estimation using an internal-gan,” in Adv. Neural Inform. Process. Syst., 2019.
[31] J. W. Soh, S. Cho, and N. I. Cho, “Meta-transfer learning for zero-shot super-resolution,” in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 3516–3525.
[32] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 1646–1654.
[33] Y. Wang, Y. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, C. C. Loy, Y. Qiao, and X. Tang, “Esrgan: Enhanced super-resolution. generative adversarial networks,” in Eur. Conf. Comput. Vis. Worksh., 2018.
[34] K. Zhang, W. Zuo, and L. Zhang, “Deep plug-and-play super-resolution for arbitrary blur kernels,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 1671–1681.
[35] K. Wei, A. I. Aviles-Rivero, J. Liang, Y. Fu, C.-B. Schönlieb, and H. Huang, “Tuning-free plug-and-play proximal algorithm for inverse imaging problems,” in Int. Conf. Mach. Learn., 2020.
[36] K. Zhang, Y. Li, W. Zuo, L. Zhang, L. V. Gool, and R. Timofte, “Plug-and-play image restoration with deep denoiser prior,” IEEE Trans. Pattern Anal. Mach. Intell., 2021.
[37] K. Zhang, L. V. Gool, and R. Timofte, “Deep unfolding network for image super-resolution,” in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 3214–3223.
[38] K. He, H. Fan, Y. Wu, S. Xie, and R. Girshick, “Momentum contrast for unsupervised visual representation learning,” in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 9729–9738.
[39] J. Heitnann, J. Munkberg, J. Hasselgren, S. Laine, T. Karras, M. Aittala, and T. Aila, “Noise2Noise: Learning image restoration without clean data,” in Int. Conf. Mach. Learn., 2018.
[40] C. Finn, P. Abbeel, and S. Levine, “Model-agnostic meta-learning for fast adaptation of deep networks,” in Int. Conf. Mach. Learn., 2017.
[41] S. Park, J. Yoo, D. Cho, J. Kim, and T. H. Kim, “Fast adaptation to super-resolution networks via meta-learning,” in Eur. Conf. Comput. Vis., 2020.
[46] J. Ho, A. Jain, and P. Abbeel, “Denoising diffusion probabilistic models,” in Adv. Neural Inform. Process. Syst., 2020.

[47] M. A. de Gosson, “From weyl to born-jordan quantization: The schrodinger representation revisited,” Physics Reports, vol. 623, pp. 1–58, Mar. 2016.

[48] P. A. M. Dirac, “Foundations of quantum mechanics,” Nature, vol. 203, pp. 115–116, Jul. 1964.

[49] F. H. Rhodes, “Gradualism, punctuated equilibrium and the origin of species,” Nature, vol. 305, no. 22, pp. 269–272, Sep. 1983.

[50] X. Wang, K. Yu, C. Dong, X. Tang, and C. C. Loy, “Deep network interpolation for continuous imagery effect transition,” in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 1692–1701.

[51] H. Zhua, L. Li, J. Wu, W. Dong, and G. Shi, “Metaqa: Deep meta-learning for no-reference image quality assessment,” in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 14143–14152.

[52] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in IEEE Conf. Comput. Vis. Pattern Recog., 2016, pp. 1063–6919.

[53] R. Timofte, E. Agustsson, L. V. Gool, M.-H. Yang, L. Zhang, and et al, “Nitre 2017 challenge on single image super-resolution: Methods and results,” in IEEE Conf. Comput. Vis. Pattern Recog. Worksh., 2017, pp. 1110–1121.

[54] M. Bevilacqua, A. Roumy, C. Guillemot, and M.-L. A. Morel, “Low-complexity single-image super-resolution based on nonnegative neighbor embedding,” in Brit. Mach. Vis. Conf., 2012.

[55] P. Arbelaez, M. Maire, C. C. Fowlkes, and J. Malik, “Contour detection and hierarchical image segmentation,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 33, no. 5, pp. 898–916, 2011.

[56] J.-B. Huang, A. Singh, and N. Ahuja, “Single image super-resolution from transformed self-exemplars,” in IEEE Conf. Comput. Vis. Pattern Recog. Worksh., 2015, pp. 5197–5206.

[57] K. Zhang, S. Gu, R. Timofte et al., “Aim 2019 challenge on constrained super-resolution: Methods and results,” in Int. Conf. Comput. Vis. Worksh., 2019.

[58] D. P. Kingma and J. Ba, “Adam: A method for stochastic optimization,” in Int. Conf. Learn. Represent., 2014.

[59] R. R. Selvaraju, M. Cogswell, A. Das, R. Vedantam, D. Parikh, and D. Batra, “Grad-cam: Visual explanations from deep networks via gradient-based localization,” in Int. Conf. Comput. Vis., 2017.

[60] A. Lugmayr, M. Danelljan, R. Timofte, N. Ahn, D. Bai, J. Cai, Y. Cao, and et al, “Ntire 2020 challenge on real-world image super-resolution: Methods and results,” in IEEE Conf. Comput. Vis. Pattern Recog. Worksh., 2020, pp. 2058–2076.

[61] Y. Li, F. Guo, R. T. Tan, and M. S. Brown, “A contrast enhancement framework with jpeg artifacts suppression,” in Eur. Conf. Comput. Vis., 2014, pp. 174–188.

[62] C. Dong, Y. Deng, C. C. Loy, and X. Tang, “Compression artifacts reduction by a deep convolutional network,” in IEEE Conf. Comput. Vis. Pattern Recog. 2015, pp. 576–584.

[63] Y. Chen and T. Pock., “Trainable nonlinear reaction diffusion: A flexible framework for fast and effective image restoration,” IEEE Trans. Pattern Anal. Mach. Intell., vol. 39, no. 6, pp. 1256–1272, 2017.

[64] X. Fu, Z.-J. Zha, F. Wu, X. Ding, and J. Paisley, “Jpeg artifacts reduction via deep convolutional sparse coding,” in Int. Conf. Comput. Vis., 2019, pp. 2501–2510.

[65] M. Ehrlich, L. Davis, S.-N. Lim, and A. Shrivastava, “Quantization guided jpeg artifact correction,” in Eur. Conf. Comput. Vis., 2020, pp. 293–309.

Yuanfei Huang received the B.Eng. degree in communication engineering from Hebei University of Technology, Tianjin, China, in 2016, and the Ph.D. degree in circuits and systems from the School of Electronic Engineering, Xidian University, Xi’an, China, in 2021. He is currently a Post-Doctoral Research Fellow with the School of Artificial Intelligence, Beijing Normal University. His current research interests include low-level vision, computational imaging, and machine learning. In these areas, he has published technical articles in refereed journals and proceedings including IEEE TIP, IEEE TCSVT, Signal Processing, etc.

Jie Li received the B.Sc., M.Sc. and Ph.D. degrees in Circuits and Systems from Xidian University, Xi’an, China, in 1995, 1998 and 2005 respectively. Since 1998, she joined the School of Electronic Engineering at Xidian University. Currently, she is a Professor of Xidian University. Her research interests include computational intelligence, machine learning, and image processing. In these areas, she has published over 50 technical articles in refereed journals and proceedings including IEEE TIP, IEEE TCVB, IEEE TCSVT, Pattern Recognition, etc.

Xinbo Gao (M’02–SM’07) received the B.Eng., M.Sc. and Ph.D. degrees in electronic engineering, signal and information processing from Xidian University, Xi’an, China, in 2008 and 2019, respectively. Since 2008, she has been a faculty member with the School of Medical Engineering and Technology, Xi’An Jiaotong University, Xi’an, China. Her current research interests include machine learning and computer vision.

Hua Huang (Senior Member, IEEE) received his B.S., M.S. and Ph.D. degrees from Xi’an Jiaotong University, Xian, China, in 1996, 2001, and 2006, respectively. He is currently a Professor in School of Artificial Intelligence, Beijing Normal University. His current research interests include image and video processing, computer graphics, and pattern recognition.

Yanting Hu received the M.Eng. degree in information and communication engineering and Ph.D. degree in pattern recognition and intelligent system from Xidian University, Xi’an, China, in 2008 and 2019, respectively. Since 2008, she has been a faculty member with the School of Medical Engineering and Technology, Xi’An Jiaotong University, Xi’an, China. Her current research interests include machine learning and computer vision.
1 Appendix

1.1 Proof of Proposition 1

**Proposition 1.** Given two Gaussian blur kernels $B^0 \sim \mathcal{N}(0, \sigma_0^2)$, $B^1 \sim \mathcal{N}(0, \sigma_1^2)$ ($\sigma_1 > \sigma_0$), and degree of transitionality $\tau$, then in element-wise $(i,j)$, $B^\tau \sim \mathcal{N}(0, \sigma_\tau^2)$ can be represented as a transition state between $B^0$ and $B^1$:

$$B^\tau(i,j) \propto (B^0(i,j))^{\frac{\sigma_0^2}{\sigma_\tau^2}} (B^1(i,j))^{\frac{\sigma_1^2}{\sigma_\tau^2}}$$

(1)

where $\sigma_\tau = (1-\tau)\sigma_0 + \tau\sigma_1$.

**Proof.** Given two Gaussian blur kernels $B^0 \sim \mathcal{N}(0, \sigma_0^2)$ and $B^1 \sim \mathcal{N}(0, \sigma_1^2)$, where $\sigma_1 > \sigma_0$, then in element-wise $(i,j)$,

$$B^0(i,j) = \frac{1}{\sqrt{2\pi}\sigma_0} e^{-\frac{i^2+j^2}{2\sigma_0^2}}, B^1(i,j) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\frac{i^2+j^2}{2\sigma_1^2}}$$

(2)

given the degree of transition $\tau$, and a transitional Gaussian blur kernel $B^\tau \sim \mathcal{N}(0, ((1-\tau)\sigma_0 + \tau\sigma_1)^2)$, then

$$B^\tau(i,j) = \frac{1}{\sqrt{2\pi}((1-\tau)\sigma_0 + \tau\sigma_1)} e^{-\frac{i^2+j^2}{2((1-\tau)\sigma_0 + \tau\sigma_1)^2}}$$

(3)

for simplifying, we transform it into logarithmic form as

$$\log(B^\tau(i,j)) = -\log(\sqrt{2\pi}((1-\tau)\sigma_0 + \tau\sigma_1)) - \frac{i^2+j^2}{2((1-\tau)\sigma_0 + \tau\sigma_1)^2}$$

$$= -(i^2+j^2)(\frac{\sigma_0^2}{2((1-\tau)\sigma_0 + \tau\sigma_1)^2}) + \frac{\sigma_1^2}{2((1-\tau)\sigma_0 + \tau\sigma_1)^2}) + f(\tau, \sigma_0, \sigma_1)$$

(4)

where $f(\tau, \sigma_0, \sigma_1)$ denotes the constant independent of $(i,j)$. Then, $B^\tau$ can be represented as a transition state between $B^0$ and $B^1$:

$$B^\tau(i,j) = f(\tau, \sigma_0, \sigma_1)(B^0(i,j))^{\frac{\sigma_0^2}{\sigma_\tau^2}} (B^1(i,j))^{\frac{\sigma_1^2}{\sigma_\tau^2}}$$

$$\propto (B^0(i,j))^{\frac{\sigma_0^2}{\sigma_\tau^2}} (B^1(i,j))^{\frac{\sigma_1^2}{\sigma_\tau^2}}$$

(5)

where $\sigma_\tau = (1-\tau)\sigma_0 + \tau\sigma_1$. 

□