Blind Image Super Resolution Using Deep Unsupervised Learning

Kazuhiro Yamawaki [1] , **Yongqing Sun** [2] **and Xian-Hua Han** [1,*]

1 Graduate School of Science and Technology for Innovation, Yamaguchi University, 1677-1 Yoshida, Yamaguchi 753-8511, Yamaguchi, Japan; b034bv@yamaguchi-u.ac.jp
2 NTT Media Intelligence Laboratories, 1-1 Hikarinooka, Yokosuka 239-0841, Kanagawa, Japan; yongqing.sun.fb@hco.ntt.co.jp
* Correspondence: hanxhua@yamaguchi-u.ac.jp

Abstract: The goal of single image super resolution (SISR) is to recover a high-resolution (HR) image from a low-resolution (LR) image. Deep learning based methods have recently made a remarkable performance gain in terms of both the effectiveness and efficiency for SISR. Most existing methods have to be trained based on large-scale synthetic paired data in a fully supervised manner. With the available HR natural images, the corresponding LR images are usually synthesized with a simple fixed degradation operation, such as bicubic down-sampling. Then, the learned deep models with these training data usually face difficulty to be generalized to real scenarios with unknown and complicated degradation operations. This study exploits a novel blind image super-resolution framework using a deep unsupervised learning network. The proposed method can simultaneously predict the underlying HR image and its specific degradation operation from the observed LR image only without any prior knowledge. The experimental results on three benchmark datasets validate that our proposed method achieves a promising performance under the unknown degradation models.

Keywords: image super resolution; blind unsupervised learning; blur kernel learning; generated network; degradation operation

1. Introduction

Single image super resolution (SISR) aims at recovering a high-resolution (HR) image from a low-resolution (LR) image, and is a fundamental low-level vision task. SISR has received substantial research attention in decades, and has widely been used in different applications [1–3]. However, due to its ill-posed nature with multiple possible HR versions for a specific LR image, SISR is still a challenging task. Numerous SISR methods have been explored to recover the plausible one from many possible solutions. The existing research is mainly categorized into traditional optimization-based methods [4–7] and recent deep learning-based methods [8–14].

Recent deep learning-based methods have made a remarkable performance gain in terms of both the effectiveness and efficiency for SISR, and various network architectures and training strategies [15–19] have been elaborated. Since the pioneering work of employing a convolutional neural network (CNN) for SISR (SRCNN) [20] has proven the feasibility and validity, most subsequent efforts have striven for designing more complicated and deeper network architectures for boosting performance. They usually requires previously prepared large-scale training pairs and well-honed training tricks for generating a stable and good super-resolution (SR) model. However, in terms of the synthesizing of the training pairs, most studies have produced the LR versions simply via bicubic down-sampling of the available HR images [15,19–22], which in general leads to the large deviation from the imaging conditions (degradation operations) of the target real LR images to be required for super resolution. Moreover, the degradation procedure is usually unknown for a specific LR image, and thus the fully supervised deep learning methods cannot be directly adapted to the real LR images captured under diverse imaging conditions.

To handle the real LR image SR problem with the unknown degradation operations, several blind SR methods have been proposed, and therein most fall in the model-based research line [23–26]. The model-based blind SR generally follows two steps of a paradigm with the blur kernel (degradation operation) estimation via exploring the self-similarity properties of natural images [27], and the subsequent optimization procedure for recovering the latent HR image. However, the reliability of the estimated blur kernel is greatly affected by the noise in the input image, and then results in the deterioration of the recovered HR image in the following optimization step. More recently, a few deep learning-based blind SR methods, such as CAB [28] and SRMD [29], assumed that the the blur kernel for a specific LR image is known, and combined the LR observation and its corresponding blur kernel as the input for deep network training. The learned model can be used for predicting the latent HR image from the real LR image conditioned on the blur kernel, which is usually required to be estimated in a separate step. In addition, Ulyanov et al. [30,31] exploited a high-quality image generating framework, dubbed as deep image prior (DIP), from a noisy input via leveraging the observed degraded image only, and applied to several image restoration tasks. Via extensive experiments for natural image generation, DIP argued that the network architecture itself possess a large amount of low-level image statistics (image priors), and is prospected to reconstruct the high-quality HR image from its degraded version only via searching the parameter space of a generative network. Without the requirement to previously learn the reconstruction model with large-scale training dataset, DIP can be considered as the unsupervised SR problem. Furthermore, since the DIP learns the optimal network parameter separately for each individual observation, it inherently has ambitious potential to be adapted for an arbitrary LR observation captured by different imaging conditions (diverse blurring kernels). However, DIP requires the known degradation operations (blurring and down-sampling) to be implemented.

This study proposes a novel blind image super-resolution framework using deep unsupervised learning for adaptively super resolving the LR observations captured under diverse imaging conditions. Specifically, inspired by the fact that the image priors are owned in the network architecture itself, we construct a generative encoder–decoder network for automatically learning the inherent priors of the latent HR image from an interference noisy input without any additional training paired samples, and establish an unsupervised deep learning framework. To adaptively deal with arbitrary LR observations with unknown degradation operations, we further propose a learnable depth-shared convolutional layer (learnable degradation module: LDM) for automatically learning the degradation operations such as blurring kernel and down-sampling operations, and then configure a blind image SR paradigm. Via inputting the learned HR image of the generative network to the designed LDM, we obtain the approximated LR image to formulate the loss function of our proposed blind SR network, and form an end-to-end blind HR image unsupervised learning network from an LR observation only. Moreover, a joint optimization strategy is investigated to solve the unconstrained deep blind SR model for simultaneously estimating the degradation operations and the latent HR image. The experimental results on several benchmark datasets validated that our proposed method illustrates an impressive performance with the known degradation and manifest reasonable reconstruction with little, even no knowledge about the degradation model.

In summary, the main contributions of our work are as the follows:

(1) A novel blind SR method with deep unsupervised learning, i.e., BSR-DUL, is proposed for simultaneously learning the latent HR image and the degradation operations without any external training samples and prior knowledge.

(2) We leverage an encoder–decoder-based generative network for modeling the prior of the latent HR image, and a learnable depth-shared convolutional layer for automatic estimation of the degradation operation. Moreover, via combining these two components, we obtain an approximated LR image for formulating the loss function of the proposed unsupervised network with the LR observation only.

(3) We investigate a joint optimization strategy to solve the BSR-DUL model for simultaneously generating the latent HR image, learning blur kernel and implementing the degradation operation, and thus establish an end-to-end blind SR learning framework, which can be adapted to super resolve the diverse LR observation captured under arbitrary imaging conditions.

The rest of this paper is organized as follows. Section 2 surveys the related work including supervised and unsupervised CNN-based image super-resolution approaches and Section 3 presents the proposed blind SR method with deep unsupervised learning (BSR-DUL). Extensive experiments are conducted in Section 4 to compare the proposed BSR-DUL with the state-of-the-art image SR methods on three benchmark datasets. Section 5 summaries this work.

2. Related Work

In this section, we briefly survey the relevant works, including fully supervised deep learning-based methods and deep unsupervised learning approaches for single image super resolution.

Supervised deep learning-based image super resolution: Motivated by the great success of deep convolutional neural networks (DCNNs) in image classification and object detection, DCNN has widely been applied for SISR, and has made significant progress in terms of the recovery performance. Recently, various network architectures and training strategies [15–19] have been elaborated for performance boosting. Dong et al. [20] firstly employed a three-layer fully convolutional neural network (CNN) for directly modeling the mapping relation between the observed LR and the HR images, and later extended to the faster version (Faster-SRCNN) [32] via expanding the spatial resolution of feature maps in the final stage for accelerating the computational speed. The continuous research attention focuses on exploring more complicated and deeper network architectures for SR performance boosting, which usually causes remarkable difficulty for training a stable model. Kim et al. increased the depth of the SR model to 20 convolutional layers in very deep convolutional networks (VDSR) [8] and further integrated the advanced residual learning structure to ease the training difficulty in deep reconstruction-classification networks (DRCN) [9]. Shi et al. proposed an efficient subpixel convolutional neural network (ESPCN) [15] to reduce the used memory and computational cost via adopting efficient subpixel convolutional layer to upscale the learned LR features to HR output at the end of the SR network. Later, Lim et al. investigated a very deep and wide network EDSR [15] by stacking residual blocks without the batch normalization (BN) layers while Ledig et al. exploited the SRResNet [11] and further integrated the dense connections [33] for boosting performance. Moreover, to improve the perceptual quality of the SISR results, several works [11,33,34] combined the perceptual loss [35] and adversarial loss [36] with the commonly used fidelity loss for the SR network training. However, all of the above SR networks are realized in a fully supervised manner and require large-scale training pairs for reconstructing a robust model. In addition, the top-performing LR-to-HR reconstruction models are generally learned with the previously prepared training sample pairs under a fixed degradation model (blurring and down-sampling operations), such as bicubic down-sampling, and face difficulty to be generalized to the LR observations captured by the real imaging sensors. Therefore, Cai et al. [37] made an effort to generate LR–HR image pairs under a realistic setting via tuning the focal length of DSLR cameras, and collected the real training images with different resolutions for learning the SR model. It is well known that different imaging sensors usually have various imaging settings, and thus the learned models, even while using the real captured image pairs by a specific sensor, may be incapable of being generalized well to the LR observations by other imaging sensors. More recently, to learn a more robust and generalized SR model for dealing with the LR observations captured under diverse imaging conditions, several works prepared the training LR/HR image pairs with a different degradation operation, such as diverse blur-kernels, and then constructed the SR model in the fully supervised learning manner [29,38,39].

However, the generalization of these constructed SR models greatly depends on the used blur-kernels in the prepared training datasets, and thus have an insufficient modeling capability to handle arbitrary blur kernels.

Unsupervised deep learning-based method: To tackle the limited generalization issue of the fully supervised deep learning methods on real scenarios, recently unsupervised learning methods have been explored for image super resolution [40]. The research based on generative adversarial networks (GAN) [36] has illustrated that the image data with the same content but different styles can be mutually translated, generally called image-to-image translation, without using the paired training samples [41,42]. Via treating the LR images as the source domain and the HR image as the target domain, image super resolution can be categorized as a special image translation task. Yuan et al. [43] proposed to solve the image SR problem using cycle-in-cycle GAN (CinGan) consisting of two translation cycles, where one cycle is adopted for translating between the real LR and synthetic LR images while the other is used between the real LR and HR images. CinGan utilized the fixed degradation model in the translation cycle from the HR images to real LR images, and therefore has a deficiency to generate diverse and real-world LR images. To increase the diversity of the degradation operations between the HR images and the real LR images, Zhao et al. [44] integrated unsupervised learning of the degradation procedure for image SR, and established the cycle for predicting the HR reconstruction and degradation models via leveraging an additional perceptual loss on the LR domain instead of HR domain. Lugmayr et al. [45] investigated two stages of the SR framework via separating the image pair synthesizing and the HR image restoration model training, where the fist stage leveraged the unsupervised image translation model to generate realistic image pairs, and the second stage learned the HR image restoration model with the synthesize image pairs. Later, Fritsche et al. [46] extended the two-stage SR method to separately deal with the low and high frequency components. In addition, Bulat et al. [47] proposed an end-to-end learning framework using the high-to-low and low-to-high networks for simultaneously modeling the relation of the LR–HR image and learning the degradation from HR images to real LR images. Moreover, to improve the SR performance, Chen et al. [48] attached another cycle learning network to model the subtle distinctiveness between the real and synthetic LR images for aiding the reconstruction of the HR images. Although these unsupervised methods manifested great potential for dealing with the real LR images, they still have to be trained previously using external image samples.

Instead of resorting to the external data, another research line make use of the observed LR image only to generate internal training samples according to the inter-scale similarity in natural images, which can be categorized into the zero-shot learning (ZSL) paradigm. Shocher et al. [49] firstly proposed a zero-shot super-resolution network, dubbed as ZSSR, which synthesized the training pairs via treating the LR observation as HR supervision and its down-sampled images as the corresponding LR version, and then learned a specific CNN model for the under-studying scene. Via varying degradation (blur kernel and down-sampling) operations in preparing the internal training samples according to the imaging conditions of the observed LR image, ZSSR is capable of addressing different blur kernels but the degradation model for the under-studying LR image is assumed to be known. Soh et al. [50] integrated meta-learning into ZSSR methods, and leveraged the advantages of both internal and external learning for improving the SR performance. These ZSL-based SR pipelines treat the observed LR image as the HR supervision ("HR father") and synthesize the "LR son" via down-sampling the LR observation to extract training paired samples for internal learning, and thus result in insufficient paired samples to train a stable model, especially for a large upscale factor. Therefore, these methods are generally adapted only to small, upscale SR tasks such as 2–4. Moreover, Ulyanov et al. [30] proposed to leverage the powerful modeling ability of deep CNN for capturing the inherent structure of nature images, and exploited a 'self-supervised' SR learning paradigm without any external and internal training pairs. DIP adopted a generative network to directly estimate the latent HR image using only the observed LR image, and demonstrated an impressive

performance even for large upscale factor. However, DIP assumed that the observed LR image is a "bicubic" down-sampling version of the latent HR image, and implemented this fixed degradation operation with mathematical computation, which restricts the wide applicability to the real scenarios. This study proposes a blind unsupervised SR framework for being adapted to the LR images captured under different imaging conditions, and closely relates to DIP [30] but has a distinctive difference. We propose to model not only the latent HR image with a generative network but also the degradation kernel with a learnable depth-shared convolutional module to construct an end-to-end blind zero-shot SR learning framework.

Summarized limitations of the existing methods: We briefly summarize the limitations of the existing SR methods and clarify the key challenges in the SR task. On one hand, the popularly used deep learning methods are implemented in a fully supervised learning way using a previously collected external dataset. Most methods synthesize the training pairs via bicubic down-sampling the available HR image to give the corresponding LR images [15–19], and only learn the models for approximating the inverse transformation of the bicubic down-sampling operation. Therefore, the applicability to the LR image captured under uncontrolled conditions would lead to a great performance degradation. On the other hand, although the blind SR methods have been actively explored recently, they are mainly realized in two separated steps via firstly estimating the blur kernel and then constructing the deep supervised models guided by the kernel. These separated strategies would cause a complicated training procedure, and the incorrectly estimated kernel possibly results in unstable super-resolving images. Moreover, the unsupervised methods (ZSSR and DIP) [30,31,49] do not require off-line model training using the external dataset, and has remarkable flexibility for dealing with the LR image captured under various conditions. However, the degradation operations of the LR observation are still required to be known, and cannot simultaneously reconstruct the latent HR image and predict the blur kernel in an end-to-end manner. This study aims to alleviate the above-mentioned difficulty in the real SR problem, and exploits an unsupervised SR method with high generalization.

3. Blind Image SR Framework with Deep Unsupervised Learning

This section first describes the problem formulation of the blind SR task, and then presents the proposed blind SR framework with deep unsupervised learning (BSR-DUL), including the motivation, the detailed generative network for modeling a latent HR image, the learnable depth-shared convolutional module for implementing the degradation operation and the joint optimization algorithm for network training.

3.1. Problem Formulation

With an observed LR image $\mathbf{I}^{LR} \in \mathbb{R}^{w \times h}$, the goal of the SR problem is to reconstruct an HR image $\mathbf{I}^{HR} \in \mathbb{R}^{W \times H}$ with $w \ll W$ and $h \ll H$. Generally, the degradation procedure of the observed $\mathbf{I}^{LR}$ is mathematically formulated as the follows:

$$\mathbf{I}^{LR} = (\mathbf{I}^{HR} \otimes \mathbf{k}) \downarrow_s + \mathbf{n}, \tag{1}$$

where $\otimes$ denotes the 2D convolution operation, $\mathbf{k}$ and $\downarrow_s$ represent the blur kernel and down-sampling operation with factor s, respectively, while $\mathbf{n}$ is the additive white Gaussian noise. Most current deep learning-based SR methods synthesize the external training dataset by using the available natural images and their simulated LR versions with a simple bicubic down-sampling operation without noise term, and construct a deep LR–HR prediction model using the prepared LR–HR pairs. Therefore, the super-resolved HR results with these learned models for the LR images captured under uncontrolled imaging conditions would be greatly degraded. This study proposes a blind SR method using deep unsupervised learning, and simultaneously learns the latent HR image and the unknown degradation model (the blur kernel and down-sampling operation) using only the observed LR image.

3.2. Motivation of the Proposed BSR-DUL

Given the previously synthesized external LR–HR image pairs $(\mathbf{x}_n^{LR}, \mathbf{y}_n^{HR})$ $(n = 1, 2, \cdots, N)$, where $\mathbf{x}_n^{LR}$ is generally a bicubic down-sampled version of $\mathbf{x}_n^{HR}$, the fully supervised CNN methods learn an off-line SR model via minimizing the reconstruction errors of training HR images to obtain the optimal network parameters as follows:

$$\theta^* = \arg \min_{\theta} \sum_{n=1}^{N} \|\mathbf{x}_n^{HR} - f_{\theta}^{CNN}(\mathbf{x}_n^{LR})\|^2, \tag{2}$$

For a test LR image $\mathbf{I}^{LR}$ to be super resolved, the latent HR image $\mathbf{I}^{HR}$ is predicted using the learned model with the optimal network parameters θ^* as: $\mathbf{I}^{HR} = f_{\theta^*}^{CNN}(\mathbf{I}^{LR})$. Unlike the above paradigm, this study appeals only to the observed LR image instead of previously preparing the external and internal training paired samples for learning an HR image reconstruction model whilst leveraging the powerful capability of deep network architecture for capturing enough low-level image statistics [30] to reconstruct the arbitrary high-quality natural images, which can simultaneously learn the latent HR image and the degradation model (the blur kernel and the down-sampling operation) from a noisy input. Specifically, we construct a generative network $\mathcal{G}$ with the unknown network parameters θ, and search a set of optimal θ in the network parameter space for capturing the latent HR image's prior: $\mathbf{I}^{HR}$. Moreover, following after the generative network, we design a learnable degradation module (LDM) using a depth-shared convolutional layer to automatically predict the blur kernel for a specific LR observation, and establish the blind unsupervised SR network in an end-to-end learning manner. The conceptual structure of our proposed BSR-DUL is shown in Figure 1a. According to the the loss function of the fully supervised CNN-based SR network in Equation (2), we formulate the objective function of the BSR-DUL framework as follows:

$$\theta^* = \arg \min_{\theta} \|\mathbf{I}^{LR} - \mathcal{G}_{\theta}(\mathbf{z}) \otimes \mathbf{k} \downarrow_s \|^2,$$
$$s.t. \quad 0 \le \mathcal{G}_{\theta}(\mathbf{z})_{i,j} \le 1, \forall i, j \tag{3}$$

where $\mathbf{z}$ is the input of the generative network, and $\mathcal{G}_{\theta}(\mathbf{z})_{i,j}$ denotes the intensity of the estimated HR image on the i-th row and j-th column pixel. In Equation (3), instead of optimization directly on the latent HR image, we search the parameter space of the generative network $\mathcal{G}_{\theta}$ for pursuing a set of optimal θ^* to well reconstruct the target as $\hat{\mathbf{I}}^{HR} = \mathcal{G}_{\theta^*}(\mathbf{z})$ using only the observed LR image. In the following subsection, we describe the detailed design of the proposed BSR-DUL including the encoder–decoder architecture of the generative network $\mathcal{G}_{\theta}$, the used input of $\mathcal{G}_{\theta}$, the learnable degradation block for automatically estimating the blur kernel and down-sampling operation and the joint optimization algorithm for network training.

3.3. The Detailed Implementation of the Proposed BSR-DUL

As shown in Figure 1a, our proposed BSR-DUL mainly includes the modeling module of the latent HR image with the generative network and the learnable degradation module (LDM) for automatically learning the blur kernel and down-sampling operation related to the imaging conditions of the LR observation. We substantiate the encoder–decoder architecture of the generative network, the specifically designed depth-shared convolutional layer for LDM and the joint optimization algorithm for network training.

The encoder-decoder-based architecture of the generative network: To handle diverse images containing salient structures and rich textures, the generative network is required to have sufficient modeling capacity. Inspired by the successful generation of high-quality image with the encoder–decoder network in various adversarial learning [42,51], we exploit a symmetric encoder–decoder architecture with skip connections, to serve as $\mathcal{G}_{\theta}$ for its multi-level feature learning nature and simplification. Both the encoder and decoder have five blocks for learning multiple scales of contexts, and the outputs of the five blocks in the

encoder are skip connected to the corresponding blocks of the decoder for feature reuse. Each block consists three convolutional layers following the RELU activation function while the max-pooling layer with a 2×2 kernel is used for decreasing the feature map size between blocks of the encoder and the up-sampling layer is employed for doubly recovering the feature map size between blocks of the decoder. Finally, a convolutional output layer is adopted to generate a latent HR image. The encoder–decoder architecture of our generative network is shown in Figure 1b.

(**a**)

(**b**)

(**c**)
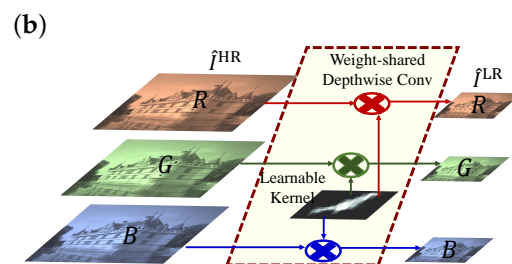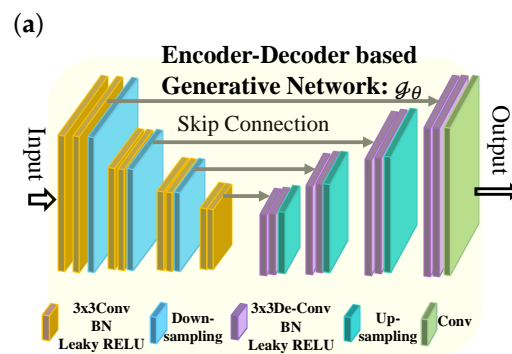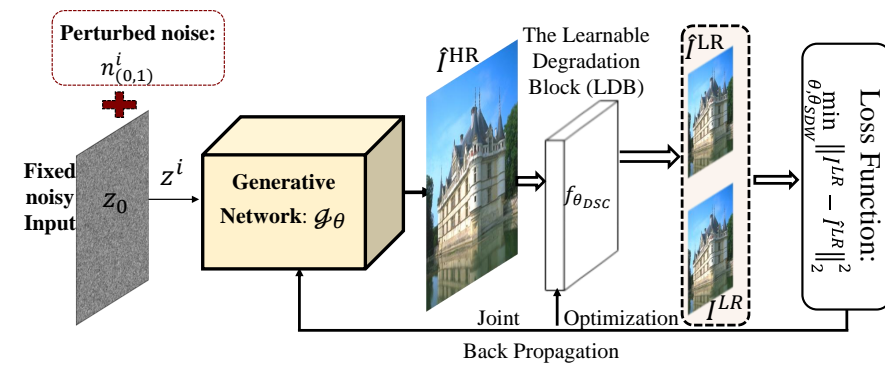
Figure 1. The conceptual scheme of the proposed blind SR framework with deep unsupervised learning (BSR-DUL). (**a**) The overall architecture of the proposed BSR-DUL. (**b**) The generative network. (**c**) The learnable degradation module.

The encoder–decoder network is used to generate the latent HR image. In natural image generation research of adversarial learning, such as DCGAN [52] and its variants [53–56], most methods use the randomly sampled noisy vectors or (observed knowledge) conditioned noisy vectors for the network input, and the quality of the generated images with the help of adversarial learning is continually improved. However, GAN-based methods aim at learning the inherent structures (priors) of the latent images with a specific concept, and expect more diverse generated samples having the same distribution with the real samples via an additional real/fake discriminator. In our unsupervised SR problem, we want to generate the corresponding HR image of a specific LR observation

instead of diverse HR samples, and thus the network input in the training procedure should be fixed. Similarly in [30], we adopt a random generated noise $\mathbf{z}_0$ at the beginning of the network learning as the base input of our generative network $\mathcal{G}_\theta$. However, the fixed input possibly leads to the generative network failing into a local minimum status. Thus, we make a small random perturbation (a randomly generated noise with a uniform distribution of the value range (0, 1)) on the initialized noise for each step of the network training, and the input to the generative network in i-th training step is formulated as:

$$\mathbf{z}^i = \mathbf{z}_0 + \beta \mathbf{n}^i_{(0,1)}, \tag{4}$$

where β denotes the perturbation degree on the base noise, and $\mathbf{n}^i_{(0,1)}$ represents the randomly generated noise in the i-th training step. We adopt the perturbation noise $\mathbf{n}^i_{(0,1)}$ to prevent the dropping into a local solution of the network training, and the perturbation degree β should be small enough to avoid the loss oscillation. In our experiments, we set β from 0.01 to 0.08, which usually provides the stable training procedure for our BSR-DUL network. With the learned generative network, we estimate the latent HR image from the initial fixed noise $\mathbf{z}_0$ as $\hat{\mathbf{I}}^{HR} = \mathcal{G}_\theta(\mathbf{z}_0)$.

The learnable degradation module: With the generated HR image of the generative network, it is needed to employ the degradation operations to approximate its corresponding LR image for conducting an evaluation of the network training. With the known blur kernel $\mathbf{k}$ and down-sampling operation, we can employ a mathematical formula to approximate the degradation model, which limits the applicability on the observation with unknown degradations. Moreover, the mathematical implementation of the degradation model is usually difficult to integrate into the learning network as an end-to-end framework. Thus, this study designs a special learnable module to implement the degradation model after the generative backbone, and constructs an end-to-end SR framework for flexibly accommodating the known and unknown degradation. Specifically, we alter a vanilla depth-wise convolutional layer to realize the blurring and down-sampling transformation. It is well known that the same blurring and down-sampling operations are conducted for all of the RGB channels in a real scenario, and then we impose the depth-wise convolutional layer on different color bands to share the same kernel with zero bias; stride parameter: 1 for blurring the operation and spatial expanding factor for down-sampling operation, which construct our proposed depth-shared convolutional (DSC) block as shown in Figure 1c. The specifically designed DSC block is expressed as:

$$\hat{\mathbf{I}}^{LR} = f_{\theta_{DSC}}(\hat{\mathbf{I}}^{HR}) \tag{5}$$

where $\hat{\mathbf{I}}^{LR}$ denotes the degraded LR version of the estimated HR image using $\mathcal{G}_\theta$. With the learnable $f_{\theta_{DSC}}$ module, it is prospected for being flexibly adapted to different real settings. Via substituting the mathematical transformation in Equation (3) with the learnable DSC block, the loss function for training our end-to-end blind unsupervised SR network can be rewritten as:

$$(\theta^*, \theta^*_{DSC}) = \underset{\theta, \theta_{DSC}}{\arg\min} \|\mathbf{I}^{LR} - f_{\theta_{DSC}}(\mathcal{G}_\theta(\mathbf{z}))\|^2, \tag{6}$$

where θ^*_{DSC} denotes the learnable parameters of the depth-shared convolutional layer for approximating the degradation model. Via minimizing Equation (6), we can jointly optimize parameters of the generative network and the degradation module. The optimization process of the blind SR network can be explained as a kind of "zero-shot" self-supervised learning [49], where the generative networks $\mathcal{G}_\theta$ and the degradation block are trained using only the observed image (i.e., the observed LR image) and no ground truth HR image is available.

Joint optimization algorithm: The optimization problem of the constructed model in Equation (6) for our BSR-DUL is unconstrained and highly non-convex. Most of the existing solutions, such as to solve the traditional MAP-based framework, often utilize

an alternating minimization strategy, which may lead the solution to be stuck in saddle points [57]. We investigate a joint optimization method instead of using alternating optimization, taking advantage of the powerful modeling capacity of $\mathcal{G}_\theta$, which can avoid an invalid and trivial HR solution. In the joint optimization, we derive the gradients w.r.t. θ and θ_{DSC} using the automatic differentiation techniques [58], and simultaneously update the parameters of the generative networks $\mathcal{G}_\theta$ and $f_{\theta_{DSC}}$. The proposed joint optimization algorithm is summarized in Algorithm 1, which jointly updates the parameters θ and θ_{DSCk} using the ADAM algorithm [59]. ADAM is an efficient optimization algorithm with an adaptive learning rate, which is specifically designed for training deep neural networks, and has been proven to achieve a relatively stable training procedure compared with other optimization methods. Moreover, ADAM is also an efficient method as it only requires first-order gradients with little memory requirement. In the experiments, we conduct the optimization procedure with T iterations, and then the latent HR image $\mathbf{I}^{HR}$ can bey generated as $\hat{\mathbf{I}}^{HR} = \mathcal{G}_\theta(\mathbf{z}_0, \theta^T)$.

Algorithm 1 Joint Optimization for BSR-DUL.

Input: the LR observation $\mathbf{I}^{LR}$
Output: the latent HR image $\mathbf{I}^{HR}$
　　Sample the base noise $\mathbf{z}_0^0$ from uniform distribution
　　for $i = 0$ to max. iter. (T) **do**
　　　Sample the noise $\mathbf{n}_{(0,1)}^i$ from uniform distribution
　　　Perturb $\mathbf{z}_0$ with $\mathbf{n}_{(0,1)}^i$: $\mathbf{z}^i = \mathbf{z}_0 + \beta \mathbf{n}_{(0,1)}^i$
　　　$\hat{\mathbf{I}}^{HR} = \mathcal{G}_\theta(\mathbf{z}^i, \theta^{i-1})$
　　　$\hat{\mathbf{I}}^{LR} = f_{DSC}(\hat{\mathbf{I}}^{HR}, \theta_{DSC}^{i-1})$
　　　Loss function of Equation (6): $\|\mathbf{I}^{LR} - \hat{\mathbf{I}}^{LR})\|^2$
　　　Compute the gradients w.r.t θ and θ_{DSC}
　　　Update θ and θ_{DSC} using the ADAM algorithm [59]
　　end for
　　$\mathbf{I}^{HR} = \mathcal{G}_\theta(\mathbf{z}_0, \theta^T)$

---

4. Experimental Results

4.1. Experimental Settings

We conducted experiments on three widely used benchmark datasets: Set5 [6], Set14 [5] and BSD100 [60]. The Set5 dataset has 5 test data including baby, bird, butterfly, head and woman images while the Set14 dataset consists of 14 data with Baboon, Barbara, bridge, coastguard, comic, face, flowers, foreman, lenna, man, monarch, pepper, ppt3 and zebra images. BSD100 is a widely used classical dataset for both for image denoising and super resolution, and has 100 test images. The dataset is composed of a large variety of images ranging from natural images to object-specific images, such as plants, people, food etc. All of the datasets are commonly used for testing the performance of image super-resolution models. We consider the original images in all of the datasets as the HR images, and synthesize the LR observations using different blurring kernels and down-sampling operations. For simple implementation, we first top-left cropped the HR image with the pixel numbers in both horizontal and vertical directions to be the multiply of 32. All of the experiments are performed with a scale factor of $4\times$ or $8\times$ between low- and high-resolution images. The quantitative metrics: peak signal-to-noise ratio (PSNR) [dB] and structural similarity index measure (SSIM) have been adopted for evaluating the SR performance, and for a fair comparison, all are computed on three RGB channels instead of the y-channel only. We calculate the average quantitative values of all of the images in each dataset, and provide the fair comparisons with the state-of-the-art (SoTA) methods and a different experimental setting in our proposed BSR-DUL.

The proposed BSR-DUL is implemented using Pytorch. We set the learning rates for $\mathcal{G}_\theta$ and $f_{\theta_{DSC}}$ as 0.01 and 1×10^{-4}, respectively, and adopt the ADAM optimization strategy.

The iteration step in all of the experiments is set as $T = 4000$ for scale factor 8 and $T = 2000$ for scale factor 4. The noises $\mathbf{z}_0$ and $\mathbf{n}^i_{(0,1)}$ are sampled from the uniform distribution with fixed random seed 0 while the perturbed parameter β is set as 0.03 for most experiments. We also change the values of β to validate the effect on the SR performance. All of the proposed models with different experimental settings are run on the computer with Ubuntu OS, 8GB memory and Tesla K80 GPU.

To verify the effectiveness of our proposed BSR-DUL framework, we firstly conduct experiments on different simulated LR images to analyze the effect of the learnable degradation module (LDM) for approximating different degradation operations. Then, we compare the SR performance of our unsupervised non-blind/blind method with the state-of-the-art methods, including the fully supervised non-blind methods and unsupervised non-blind SR methods.

4.2. Compared Results on Different Degraded LR Images

Without a lack of generalization, we simulate the LR inputs from the benchmark datasets: Set5 [6], Set14 [5] and BSD100 [60] with different degradation operations including the simple bicubic down-sampling only (without the blur kernel) and the combined bicubic down-sampling and Gaussian blur kernels with different standard deviation values (σ from 1.0 to 3.0). As mentioned in Section 3, the kernel weights of the learnable degradation module $f_{\theta_{DSC}}$ can previously be defined and fixed in the network training procedure to establish a non-blind SR framework. For the bicubic down-sampled LR images, we firstly conduct the experiments in the non-blind setting via initially fixing the LDM's kernel weights as the correct kernel (Lanczos kernel for approximating bicubic down-sampled operation) and a wrong kernel, such as the Gaussian kernel, for validating the learning capability of the generative network $\mathcal{G}_\theta$. Moreover, we assume no prior knowledge about the degradation procedure, and automatically learn the kernel weights to verify the potential of the kernel modeling capability of the $f_{\theta_{DSC}}$. The quantitative comparisons on all three datasets from the bicubic down-sampled LR images with upscale factor 4 and 8 are manifested in Table 1. It can be seen from Table 1 that the learnable kernel under the blind setting of our proposed method illustrates comparable results with the correct kernel (here Lanczos kernel for bicubic down-sampling) under the non-blind setting. Two recovered HR images of two samples with a wrong kernel, the correct kernel and the automatically learned kernel are shown in Figure 2.

Next, we simulate the LR image using both the blur kernel $\mathbf{k}$ and the bicubic down-sampling operation, and conduct experiments to verify the feasibility of the our blind SR method. Without a loss of generality, Gaussian blur kernels with different standard deviations from 1.0 to 3.0 are used. Experiments have been conducted under varieties of settings, including non-blind (known blur kernel and down-sampling), semi-blind, where there is little knowledge about the blur kernel such as only the known kernel type (Gauss) is known, and a complete-blind paradigm without any prior knowledge about the blur kernel. In the semi-blind experimental setting with the known kernel type but unknown deviation value σ, we simply set three values: 0 (assume no blur kernel), 1 and the true value to get the Gaussian kernel, and set them as the weights of the LDM, respectively, while learning the parameters θ of the generative network $\mathcal{G}_\theta$ only to estimate the latent HR image $\mathbf{I}^{HR}$. Moreover, with an unknown kernel type, the blur kernel is automatically learned via setting the stride of the DSC layer as 1 in the LDM to give a non-down-sampled blurred version of the estimated HR image, and further adopt another DSC layer with the fixed Lanczos kernel to produce the approximated LR image. Table 2 manifests the quantitative comparisons on the Set5 and Set14 datasets using different experimental settings with the upscale factor 4 and 8, respectively. From Table 2, it can be seen that the correct kernel can provide the best results while the learned blur kernel with the known down-sampling operation can only give the second best results for most different types of LR images. Figure 3 gives the visualization of the reconstructed HR images with different experimental settings for the simulated LR images using a Gaussian blur kernel with $\sigma = 1.0$.

Table 1. Comparison of quantitative evaluation on Bicubic downsampled LR images using f_{DSC} the pre-defined and learnable kernels. The first and second numerals represent the PSNR and SSIM values, respectively.

Dataset	Factor	Correct Kernel	Wrong Kernel	Learnable $f_{\theta_{DSC}}$
Set5	X4	28.36/0.9049	19.10/0.6965	27.31/0.9053
	X8	24.25/0.7944	19.00/0.6475	23.35/0.7750
Set14	X4	25.14/0.8144	18.31/0.6398	23.41/0.8107
	X8	23.37/0.7046	18.48/0.5950	20.84/0.6896
B100	X4	25.16/0.7869	19.60/0.6452	23.11/0.7858
	X8	23.02/0.6824	20.02/0.6083	20.82/0.6751

Table 2. Quantitative comparison for super-resolving the LR images with Gaussian blur kernels (different standard deviation values) and the bicubic down-sampling (DS) operation. The first and second numerals represent the PSNR and SSIM values, respectively.

Dataset	σ	Semi-Blind			Blind
		Known DS and Gaussian Kernel with Different σ			Unknown Kernel
		$\sigma = 0$	$\sigma = 1.1$	True σ	$f_{\theta_{DSC}}$
Set5	$\sigma = 1.0$	24.17/0.7895	24.34/0.7962	24.39/0.7976	24.07/0.7875
	$\sigma = 1.2$	24.00/0.7846	24.34/0.8087	24.44/0.8000	23.84/0.7789
	$\sigma = 1.5$	23.83/0.7786	24.24/0.7911	24.36/0.7962	23.62/0.7812
	$\sigma = 2.0$	23.73/0.7732	24.25/0.7918	24.38/0.7968	23.84/0.7886
	$\sigma = 2.5$	21.42/0.6913	21.84/0.7055	23.73/0.7716	21.54/0.7000
	$\sigma = 3.0$	20.77/0.668	21.05/0.6776	23.09/0.7464	20.80/0.6719
Set14	$\sigma = 1.0$	22.17/0.6951	22.30/0.6912	22.45/0.7052	22.10/0.6971
	$\sigma = 1.2$	22.12/0.6925	22.38/0.7030	22.46/0.7041	21.87/0.6902
	$\sigma = 1.5$	22.05/0.6898	22.28/0.6988	22.45/0.7043	20.88/0.6897
	$\sigma = 2.0$	21.99/0.6867	22.33/0.6995	22.41/0.7029	21.12/0.6940
	$\sigma = 2.5$	20.43/0.6314	20.74/0.6407	22.03/0.6821	19.66/0.6355
	$\sigma = 3.0$	19.92/0.6145	19.92/0.6145	21.69/0.6673	19.27/0.6163

4.3. Comparison with State-of-the-Arts

Most of the existing methods typically super resolve the bicubic down-sampled LR images to measure the quality of the recovered HR images. To provide a fair comparison, we also conduct experiments on the bicubic down-sampled LR images using our proposed method and the state-of-the-art methods, including the unsupervised/non-blind pipeline (bicubic up-sampling, TV_Prior: unsupervised optimization-based method, DIP [30]), ZSSR [49], fully supervised deep network: LapSRN [10] and EDSR [15]. It should be noted that the degradation operations should be known for realizing the ZSSR [49] method (denoted as ZSSR_CK). Since it has to first obtain the training samples of the synthesized LR images and the original LR observation to begin the specific CNN model training in ZSSR, it is difficult to extend the ZSSR for the blind SR task. As introduced above, our proposed BSR-DUL method is a generalized unsupervised framework, and can be implemented in non-blind, semi-blind and complete-blind ways. Thus, we give the compared quantitative results of our non-blind and blind unsupervised implementation with the existing methods in Table 3, which manifests that our non-blind implementation can achieve an acceptable performance. Although the complete-blind implementation on the more challenging conditions leads to the performance degradation, it demonstrates the

feasibility and potential of the proposed generalized framework on super resolving the LR images captured under diverse imaging conditions. Figure 4 shows the comparison of the visualization results of the HR images restored by our methods and various SoTA methods. Moreover, in order to evaluate the effect of the perturbation degree β and the optimizing strategies, we further provide the compared results with different values of β and different optimizers in Tables 4 and 5.
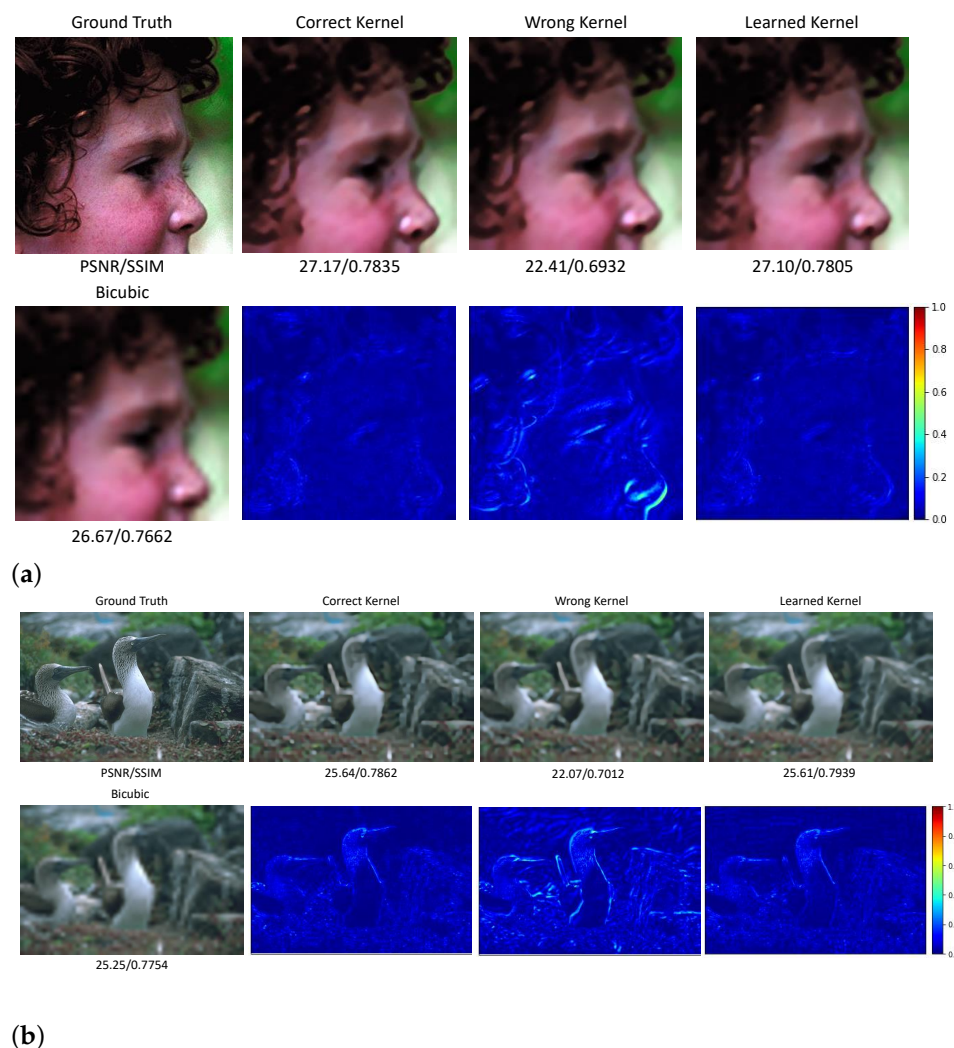
Figure 2. The experimental results of two images from the Bicubic down-sampled LR observation with the upscale factor 8. The first row gives the recovered HR images with different kernels, and the second row visualizes the difference in image between the ground truth and their corresponding estimations. (**a**) The 'head' image in Set5. (**b**) The '103070' image in BSD100.

4.4. Discussion

As validated in the above section, our proposed BSR-DUL method can simultaneously learn the latent HR image and the adaptive degradation operations on the LR observation, and thus has a high generalization ability for dealing with real diverse images. In spite of the proved feasibility on the LR images captured under uncontrolled conditions, the degradation operations are naively learned using a depth-wise convolution layer, which may result in the irrational parameters inconsistent with the real imaging scenario. As we know, the mathematical transformation parameters of the degradation in the real optical systems should be non-negative and equality. However, the learned parameters in the depth-wise convolution layer are not always conformed to the optical constraints. Thus, it would be encouraged to incorporate the optical constraints in learning the degradation operation

in the future work. Moreover, the proposed method requires us to train an image-specific CNN model for each under-studying image, and thus leads to additional on-line training time, including the inference time. In our experiments, recovering a 512×512 image takes the training and inference time about 3 and 5 minutes for the upscale factors 4 and 8, respectively. We are going to improve the efficiency of the network training by exploiting different strategies, such as constructing a common model as the initial state of the image-specific CNN.
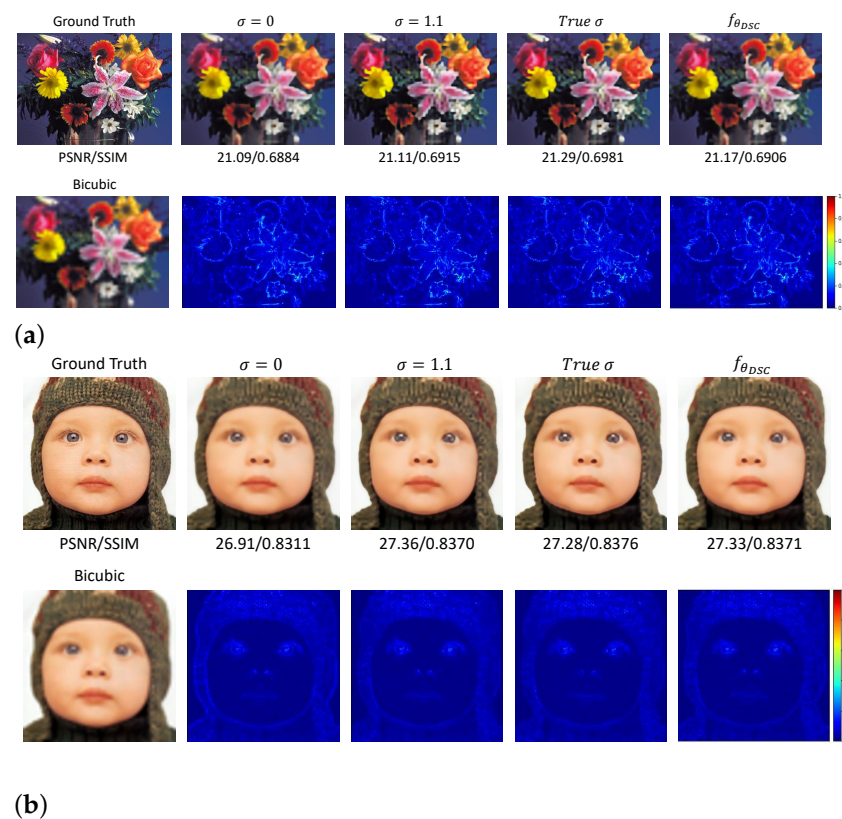
Figure 3. The compared visualization of the reconstructed HR images with different experimental settings. (**a**) The 'flower' image in Set14 from its simulated LR images using a Gaussian blur kernel with $\sigma = 1.0$. (**b**) The 'baby' image in Set5 from its simulated LR image using a Gaussian blur kernel with $\sigma = 2.0$. (**a**) The 'flower' image in Set14. (**b**) The 'baby' image in Set5.

Table 3. Quantitative comparison of our proposed BSR-DUL with the sate-of-the-art methods on the simulated LR images of all three benchmark datasets. The LR images are simulated via bicubic down-sampling the ground truth HR images for a fair comparison with fully supervised methods, such as LapSRN [10] and EDSR [15]. Similarly, the first and second numerals represent the PSNR and SSIM values, respectively.

Categories	Methods	Datasets					
		Set5		Set14		BSD100	
		X4	X8	X4	X8	X4	X8
	Bicubic	26.71/0.8660	22.74/0.7278	24.20/0.7860	21.37/0.6624	24.78/0.7725	22.48/0.6618
Unsuper Non-Blind	TV_Prior	26.66/0.8761	23.01/0.7433	24.34/0.7870	21.60/0.6761	-	-
	DIP [30]	27.93/0.8928	24.04/0.7828	25.01/0.8030	22.17/0.6953	25.15/0.7862	23.01/0.6859
	ZSSR_CK [49]	28.85/0.8009	24.18/0.6272	26.86/0.7381	23.07/0.5627	$-/-$	$-/-$

Table 3. *Cont.*

Categories	Methods	Set5		Set14		BSD100	
		X4	X8	X4	X8	X4	X8
	Our_CK	28.36/0.9049	24.25/0.7944	25.14/0.8144	23.37/0.7046	25.19/0.7919	23.02/0.6824
Unsuper Blind	Our_blind	27.31/0.9053	23.73/0.7876	23.41/0.8107	20.84/0.6896	23.11/0.7858	20.82/0.675
Super	LapSRN [10]	29.36/0.9196	24.22/0.7913	25.90/0.8327	22.43/0.7061	25.97/0.8115	23.21/0.6926
Non-Blind	EDSR [15]	29.99/0.9275	24.25/0.7959	26.37/0.8441	22.39/0.7060	26.20/0.8178	23.05/0.6890

Table 4. Performance effect of the perturbation degree β on the Bicubic down-sampled LR images of the Set5 dataset.

Factors	$\beta = 0$	$\beta = 0.01$	$\beta = 0.03$	$\beta = 0.05$	$\beta = 0.08$
X4	25.80/0.8567	26.93/0.8932	27.31/0.9053	26.88/0.8989	26.25/0.8850
X8	22.07/0.7176	23.62/0.7900	23.73/0.7876	23.35/0.7750	23.18/0.7651

Table 5. Performance effect of different optimization strategies on the Bicubic down-sampled LR images of the Set5 dataset.

Optimizers	SGD	Adadelta	Adagrad	ADAM
	17.67/0.5862	18.22/0.5950	21.38/0.7474	23.73/0.7876

(**a**)

Figure 4. *Cont.*

(**b**)
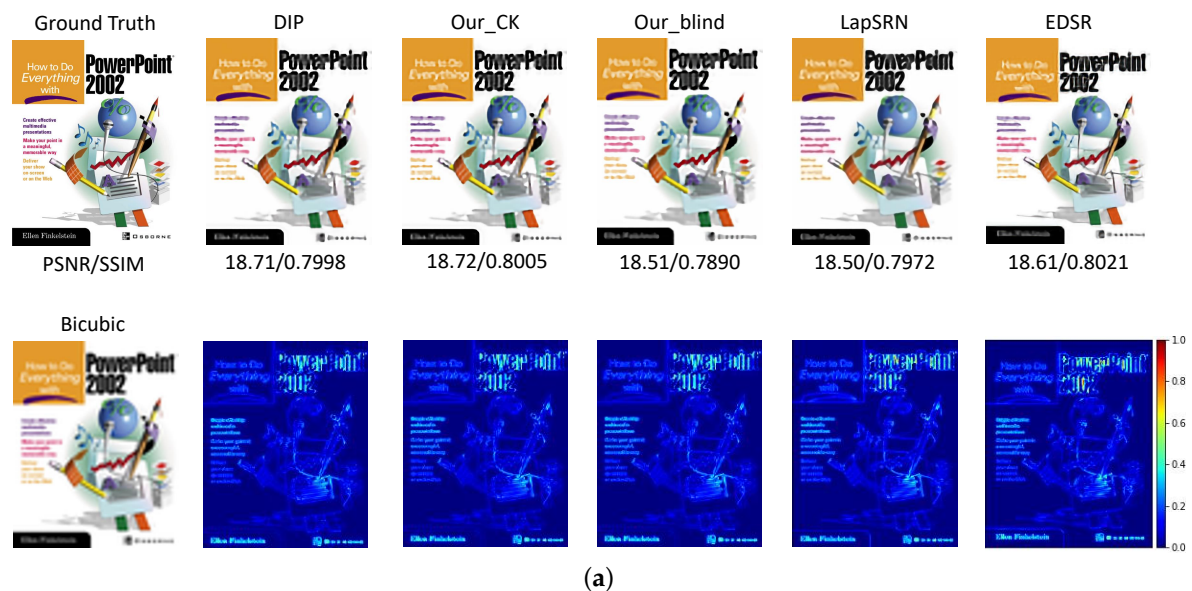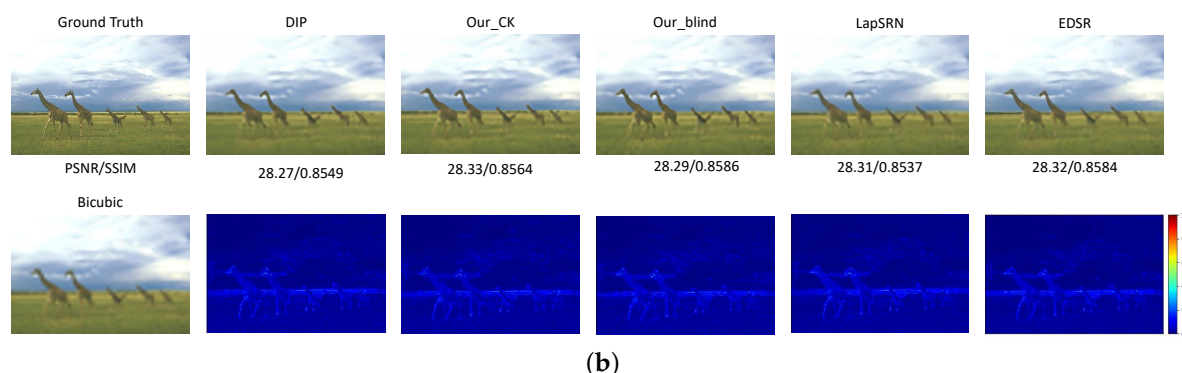
Figure 4. Compared visualization results of the recovered HR images with different SoTA methods. The first row denotes the resulted HR images while the second row gives the difference in images between the recovered and the ground truth images. (**a**) The 'ppt3' image in Set14. (**b**) The '253055' image in BSD100.

5. Conclusions

This study proposed a blind unsupervised learning network for a real SR task from a single LR image. We specifically constructed a generative network for simultaneously learning the inherent priors of the latent HR image and the degradation operations with the under-studying LR observation only. The proposed method is capable of learning any complicated blurring kernel in a general SR framework, and is an end-to-end HR image learning network. The experimental results on three benchmark datasets validated that the proposed method achieved an impressive performance under the unknown degradation model.

Author Contributions: Conceptualization, methodology and writing, K.Y. and X.-H.H.; software, K.Y.; validation and visualization, K.Y.; supervision, project administration and funding acquisition, X.-H.H.; investigation and data curation, Y.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported in part by the Grant-in Aid for Scientific Research from the Japanese Ministry for Education, Science, Culture and Sports (MEXT) under the Grant No. 20K11867, and JSPS KAKENHI Grant Number JP12345678.

Institutional Review Board Statement: Not applicable

Informed Consent Statement: Not applicable.

Data Availability Statement: Data available in a publicly accessible repository.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Nasrollahi, T.; Moeslund, T.B. Super-resolution: A comprehensive survey. *Mach. Vis. Appl.* **2014**, *25*, 1423–1468. [CrossRef]
2. Yang, Q.; Yang, R.; Davis, J.; Nister, D. Spatial-depth super Resolution for range images. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Minneapolis, MN, USA, 18–23 June 2007; pp. 1–7.
3. Zou, W.W.W.; Yuen, P.C. Very Low Resolution Face Recognition Problem. *IEEE Trans. Image Process.* **2011**, *21*, 327–340. [CrossRef] [PubMed]
4. Zhang, K.; Gao, X.; Tao, D.; Li, X. Single image super-resolution with non-local means and steering kernel regression. *IEEE Trans. Image Process.* **2012**, *21*, 4544–4556. [CrossRef] [PubMed]
5. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. *Int. Conf. Curves Surf.* **2010**, *6920*, 711–730.
6. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M.L. Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. *Proc. Br. Mach. Vis. Conf.* **2012**, *135*, 1–10.
7. Begin, I.; Ferrie, F. Blind super-resolution using a learning-based approach. In Proceedings of the 17th International Conference on Pattern Recognition, ICPR 2004, Cambridge, UK, 26 August 2004; Volume 2, pp. 85–89.
8. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.

9.   Kim, J.; Lee, J.K.; Lee, K.M. Deeply recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.

10.   Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate superresolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; Volume 2.

11.   Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z. Photorealistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

12.   Tai, Y.; Yang, J.; Liu, X.; Xu, C. Memnet: A persistent memory network for image restoration. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4539–4547.

13.   Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

14.   Zhang, Y.; Li, K.; Li, K.; Wang, L.; Zhong, B.; Fu, Y. Image super-resolution using very deep residual channel attention networks. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 286–301.

15.   Lim, B.; Son, S.; Kim, H.; Nah, S.; Lee, K.M. Enhanced Deep Residual Networks for Single Image Super-Resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 1132–1140.

16.   Liu, D.; Wen, B.; Fan, Y.; Loy, C.C.; Huang, T.S. Non-local recurrent network for image restoration. *arXiv* **2018**, arXiv:1806.02919.

17.   Dai, T.; Cai, J.; Zhang, Y.; Xia, S.T.; Zhang, L. Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11065–11074.

18.   Hu, X.; Mu, H.; Zhang, X.; Wang, Z.; Tan, T.; Sun, J. Meta-sr: A magnification-arbitrary network for super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1575–1584.

19.   Li, Z.; Yang, J.; Liu, Z.; Yang, X.; Jeon, G.; Wu, W. Feedback network for image superresolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 3867–3876.

20.   Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 184–199.

21.   Ji, X.; Cao, Y.; Tai, Y.; Wang, C.; Li, J.; Huang, F. Real-World Super-Resolution via Kernel Estimation and Noise Injection. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020.

22.   Chen, H.; He, X.; Qing, L.; Wu, Y.; Ren, C.; Zhu, C. Real-World Single Image Super-Resolution: A Brief Review. *arXiv* **2021**, arXiv:2103.02368.

23.   He, H.; Siu, W.C. Single image super-resolution using gaussian process regression. In Proceedings of the CVPR 2011, Colorado Springs, CO, USA, 20–25 June 2011; pp. 449–456.

24.   He, Y.; Yap, K.H.; Chen, L.; Chau, L.P. A soft map framework for blind super-resolution image reconstruction. *Image Vis. Comput.* **2009**, *27*, 364–373. [CrossRef]

25.   Joshi, N.; Szeliski, R.; Kriegman, D.J. Psf estimation using sharp edge prediction. In Proceedings of the 2008 IEEE Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.

26.   Wang, Q.; Tang, X.; Shum, H. Patch based blind image super resolution. In Proceedings of the Tenth IEEE International Conference on Computer Vision (ICCV'05), Beijing, China, 17–20 October 2005; Volume 1, pp. 709–716.

27.   Michaeli, T.; Irani, M. Nonparametric blind super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 945–952.

28.   Riegler, G.; Schulter, S.; Ruther, M.; Bischof, H. Conditioned regression models for non-blind single image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 522–530.

29.   Zhang, K.; Zuo, W.; Zhang, L. Learning a single convolutional super-resolution network for multiple degradations. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

30.   Dmitry, U.; Andrea, V.; Victor, L. Deep Image Prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.

31.   Dmitry, U.; Andrea, V.; Victor, L. Deep Image Prior. *arXiv* **2017**, arXiv:1711.10925.

32.   Dong, C.; Loy, C.C.; Tan, X. Accelerating the Super-Resolution Convolutional Neural Network. In Proceedings of the European Conference on Computer Vision (ECCV), Amsterdam, The Netherlands, 11–14 October 2016.

33.   Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Loy, C.C. ESRGAN: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision (ECCV) Workshops, Munich, Germany, 8–14 September 2018; pp. 63–79.

34.   Sajjadi, M.S.M.; Scholkopf, B.; Hirsch, M. Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

35.   Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016.

36.   Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.C.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Process. Syst.* **2014**, *27*, 2672–2680.

37. Cai, J.; Zeng, H.; Yong, H.; Cao, Z.; Zhang, L. Toward real-world single image super-resolution: A new benchmark and A new model. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Seoul, Korea, 27–28 October 2019.

38. Gu, J.; Lu, H.; Zuo, W.; Dong, C. Blind super-resolution with iterative kernel correction. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1604–1613.

39. Zhang, K.; Zuo, W.; Zhang, L. Deep plug-and-play super-resolution for arbitrary blur kernels. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019.

40. Lugmayr, A.; Danelljan, M.; Timofte, R.; Fritsche, M.; Gu, S.; Purohit, K.; Irani, M. Aim 2019 challenge on real-world image super-resolution: Methods and results. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019.

41. Zhu, J.Y.; Park, T.; Isola, P.; Efros, A.A. Unpaired image-to-image translation using cycle-consistent adversarial networks. *arXiv* **2017**, arXiv:1703.10593.

42. Yi, Z.; Zhang, H.; Tan, P.; Gong, M. Unsupervised dual learning for image-to-image translation. *arXiv* **2017**, arXiv:1704.02510.

43. Yuan, Y.; Liu, S.; Zhang, J.; Zhang, Y.; Dong, C.; Lin, L. Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018.

44. Zhao, T.; Ren, W.; Zhang, C.; Ren, D.; Hu, Q. Unsupervised degradation learning for single image super-resolution. *arXiv* **2018**, arXiv:1812.04240.

45. Lugmayr, A.; Danelljan, M.; Timofte, R. Unsupervised learning for real-world super-resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019.

46. Fritsche, M.; Gu, S.; Timofte, R. Frequency separation for real-world super-resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision Workshop (ICCVW), Seoul, Korea, 27–28 October 2019.

47. Bulat, A.; Yang, J.; Tzimiropoulos, G. To learn image super-resolution, use a gan to learn how to do image degradation first. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018.

48. Chen, S.; Han, Z.; Dai, E.; Jia, X.; Liu, Z.; Liu, X.; Zou, X.; Xu, C.; Liu, J.; Tian, Q. Unsupervised Image Super-Resolution with an Indirect Supervised Path. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020.

49. Shocher, A.; Cohen, N.; Irani, M. "Zero-shot" super-resolution using deep internal learning. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 3118–3126.

50. Soh, J.W.; Cho, S.; Cho, N.I. Meta-Transfer Learning for Zero-Shot Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3516–3525.

51. Isola, P.; Zhu, J.Y.; Zhou, T.; Efros, A.A. Image-To-Image Translation With Conditional Adversarial Networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.

52. Radford, A.; Metz, L.; Chintala, S. Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks. *arXiv* **2015**, arXiv:1511.06434.

53. Zhang, H.; Goodfellow, I.; Metaxas, D.; Odena, A. Self-Attention Generative Adversarial Networks. *arXiv* **2019**, arXiv:1805.08318.

54. Karras, T.; Aila, T.; Laine, S.; Lehtinen, J. Progressive Growing of GANs for Improved Quality, Stability, and Variation. *arXiv* **2018**, arXiv:1710.10196.

55. Donahue, J.; Krähenbühl, P.; Darrell, T. Adversarial Feature Learning. *arXiv* **2017**, arXiv:1605.09782.

56. Karras, T.; Laine, S.; Aila, T. A Style-Based Generator Architecture for Generative Adversarial Networks. *arXiv* **2019**, arXiv:1812.04948.

57. Tseng, P. Convergence of a block coordinate descent method for nondifferentiable minimization. *J. Optim. Theory Appl.* **2001**, *109*, 475–494. [CrossRef]

58. Paszke, A.; Gross, S.; Chintala, S.; Chanan, G.; Yang, E.; DeVito, Z.; Lin, Z.; Desmaison, A.; Antiga, L.; Lererr, A. Automatic differentiation in pytorch. In Proceedings of the NIPS Workshop: The Future of Gradient-Based Machine Learning Software and Techniques, Los Angles, CA, USA, 8–9 December 2017.

59. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representation, San Diego, CA, USA, 7–9 May 2015.

60. Martin, D.; Fowlkes, C.; Tal, D.; Malik, J. A database of human segmented natural images and its application to evaluating segmentation algorithms and meauring ecological statics. In Proceedings of the International Conference on Computer Vision (ICCV), Vancouver, BC, Canada, 7–14 July 2001; pp. 416–423.