electronics

Article

Deep Image Prior for Super Resolution of Noisy Image

Sujy Han [1], Tae Bok Lee [1] and Yong Seok Heo [1,2,*]

[1] Department of Artificial Intelligence, Ajou University, Suwon 16499, Korea; tn0502wl@ajou.ac.kr (S.H.); dolphin0104@ajou.ac.kr (T.B.L.)
[2] Department of Electrical and Computer Engineering, Ajou University, Suwon 16499, Korea
* Correspondence: ysheo@ajou.ac.kr

Abstract: Single image super-resolution task aims to reconstruct a high-resolution image from a low-resolution image. Recently, it has been shown that by using deep image prior (DIP), a single neural network is sufficient to capture low-level image statistics using only a single image without data-driven training such that it can be used for various image restoration problems. However, super-resolution tasks are difficult to perform with DIP when the target image is noisy. The super-resolved image becomes noisy because the reconstruction loss of DIP does not consider the noise in the target image. Furthermore, when the target image contains noise, the optimization process of DIP becomes unstable and sensitive to noise. In this paper, we propose a noise-robust and stable framework based on DIP. To this end, we propose a noise-estimation method using the generative adversarial network (GAN) and self-supervision loss (SSL). We show that a generator of DIP can learn the distribution of noise in the target image with the proposed framework. Moreover, we argue that the optimization process of DIP is stabilized when the proposed self-supervision loss is incorporated. The experiments show that the proposed method quantitatively and qualitatively outperforms existing single image super-resolution methods for noisy images.

Keywords: image restoration; deep image prior; super-resolution

check for updates

Citation: Han, S.; Lee, T.B.; Heo, Y.S. Deep Image Prior for Super Resolution of Noisy Image. *Electronics* **2021**, *10*, 2014. https://doi.org/10.3390/electronics10162014

Academic Editor: Ngai-Man (Man) Cheung

Received: 14 July 2021
Accepted: 18 August 2021
Published: 20 August 2021

1. Introduction

Single image super-resolution (SISR) aims to generate a high-resolution (HR) image from a low-resolution (LR) image. SISR has become one of the important tasks in computer vision. Unlike most deep learning models that are trained on large-scale datasets, Ulyanov et al. [1] recently proposed a deep image prior (DIP) that utilizes a deep neural network (DNN) as a strong prior for image restoration by using only a single image. The results of DIP show that the DNN is useful for capturing meaningful low-level image statistics. With the success of DIP [1], it has been utilized in several ways due to its usefulness for a variety of purposes. DIP has significance in the applications where collecting large-scale of datasets is difficult and expensive, such as hyperspectral image processing [2,3]. Furthermore, DIP can be used for optimization methods when solving inverse problems such as super-resolution, deblurring and denoising [4,5].

In particular, it was demonstrated that the super-resolution (SR) problem for a given target image x_0 can be solved using DIP by minimizing the following reconstruction loss term:

$$E(x; x_0) = ||DS(x) - x_0||^2, \tag{1}$$

where $DS(\cdot)$ is a downsampling operation and x is the restored HR image. By using the downsampling operation, the spatial resolution of x becomes the same as that of x_0.

In practice, the images taken from cameras equipped in the mobile embedded system are prone to have low-resolution and be corrupted by noise due to the small sizes of the camera sensors and apertures [6]. In such situations, the performance of DIP in the SR task (DIP-SR) [1] is significantly degraded (see Figure 1a). The degradation is attributable to the following two reasons. First, the reconstruction loss (Equation (1)) of DIP-SR does not consider the noise in x_0. The loss term only minimizes the pixel-wise difference between $DS(x)$ and x_0; hence, $DS(x)$ tends to be noisy. As $DS(x)$ is dependent only on x, the fact that $DS(x)$ contains noise implies that x also contains noise. Therefore, DIP-SR requires an additional constraint to handle noise effectively. Second, the DIP optimization process is unstable and sensitive to noise. It has been shown that, for a noisy input image, DIP needs early-stopping during the optimization process in order to avoid overfitting the generated image to the noise so that a clean image can be obtained. However, DIP is limited in the absence of a ground-truth image because it cannot be determined whether the result of the early-stopping is the optimal solution. Therefore, it is essential to obtain a method for DIP to achieve noiseless results through a reliable optimization process without early-stopping.

Herein, we propose a novel DIP-based SR framework that can restore a clean HR image from a noisy LR image. As mentioned earlier, one of the main drawbacks of DIP-SR [1] is that it does not consider noise when minimizing the reconstruction loss in the LR space. Since the noisy LR image contains both signal and noise, the signal needs to be learnt by separating the noise from the LR image. However, separating the noise from an image is very challenging in the absence of ground-truth information. In order to overcome this, we propose a framework to learn the distribution of noise, even when the ground-truth of noise is unknown. As shown in Chen et al. [7], generative adversarial networks (GANs) [8] have the capacity to learn the complex distribution of noise. Inspired by this finding, we employ the GAN framework to estimate noise. Specifically, our framework consists of a generator and a discriminator. The generator aims to reconstruct a clean HR image from a noisy LR image. If a clean HR output is restored by the generator, the downsampled result is also a noise-free LR image. Thus, the difference between the downsampled result and the noisy target image must follow the distribution of the real noise. Based on this, we trained our discriminator to determine whether the distribution of the extracted noise follows the real noise distribution. We sampled the real noise sample from Gaussian distribution because it is one of the most common noise models in the image restoration field [9]. The use of an adversarial framework allows our generator to learn how to reconstruct noiseless HR image. In contrast to [7], which utilizes large scale datasets, our framework is trained to extract the noise for only a single image. In addition, we propose a self-supervision loss to increase the stability of the optimization process and prevent early-stopping. In general, signals tend to have high self-similarity (low entropy and low patch diversity) whereas noise has low self-similarity (high entropy and high patch diversity) [1,10]. Ulyanov et al. [1] showed that the parameters of convolutional neural networks (CNNs) have high impedance to noise and low impedance to signals. Owing to this characteristic, when the target image is noisy, the signals are learnt by the CNN in the early stages of optimization before overfitting to noise occurs. In other words, the results of the early stages of optimization are noiseless and significant. Thus, we assume that the result of the early stage of optimization can be used as an effective regularizer for noise-free signal reconstruction. Based on this assumption, we propose a self-supervision loss that utilizes the result of the previous iteration step during the optimization process. By comparing the output image of the current step with that of the previous step, the reconstructed image can retain the learned signal without following the noise in the target image. Thus, the proposed loss prevents the reconstructed image from becoming noisy and it results in stable optimization process without the need for early-stopping.

Extensive experiments on the SISR task in various scenarios show that our method achieves the best quantitative and qualitative results in comparison to the existing SISR methods. Figure 1 exemplifies that our method generates realistic and clean HR image, whereas DIP-SR [1] suffers from noise.

Our main contributions can be summarized as follows:

- We present a GAN [8] framework to estimate the noise in a target image. Given only a noisy LR image without the ground truth, our generator reconstructs a clean HR image. The noise is estimated by learning the noise distribution in the LR image.

- We introduce the self-supervision loss (SSL), a novel approach for resolving the dependency on early-stopping and instability in the DIP [1] optimization process.
- We achieve competitive results in various experiments on Set5 [11] and Set14 [12] datasets. The proposed method outperforms the existing SISR methods.

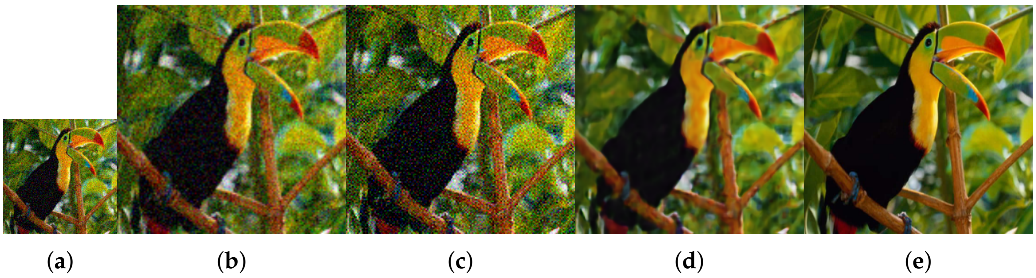

Figure 1. Generated images and PSNR results obtained from bicubic upsampling, DIP-SR [1], and our method when the input LR image is noisy (scaling factor = 2). Note that our method does not suffer from noise, unlike bicubic upsampling and DIP-SR. (**a**) Input (21.05 dB), (**b**) Bicubic (23.66 dB), (**c**) DIP-SR [1] (18.43 dB), (**d**) Ours (26.67 dB) and (**e**) Ground Truth.

2. Related Works

Learning-based approaches using convolutional neural networks (CNNs) have recently achieved excellent performance in image SR. Most CNN-based SR models are trained in a supervised manner using large-scale datasets that contain LR and HR image pairs. Thus, these models learn a well-generalized distribution of the HR images from the training data. SRCNN [13], which learns the mapping from an interpolated LR image to a HR image, was first proposed in the pioneering work. However, the direct mapping of the input image to the target image is difficult to achieve. In order to alleviate this difficulty, a VDSR that learns only the residuals between the input and target images in a process called global residual learning was proposed in [14]. Since the global residual learning greatly reduces the learning difficulty and model complexity [15], it has been used in many SR models including [16–22]. Ledig et al. [16] proposed a SRResNet that combines the ResNet [23] architecture with global residual learning. In addition, the authors applied adversarial training [8] to image SR in order to generate realistic images. EDSR [17] employs a multi-scale architecture with global residual learning and is able to restore HR images with various upscaling factors in a single model. Guo et al. [18] proposed a wavelet prediction network for SR by using residuals. SRDenseNet [24], RDN [20], ESRGAN [19] and DRLN [22] combined DenseNet [25] blocks and global residual learning in order to capture rich features. Benefiting from global residual learning, most existing SR methods are trained to enhance high-frequency information. Due to this characteristic, they also amplify the noise in the LR images. In addition, they do not leverage the information specific to a single image as a prior because they are trained to model the distribution of large external datasets. By contrast, we propose a noise-robust image SR method that focuses on the internal information in a given single image.

Instead of using large scale training datasets, a deep image prior (DIP) [1] framework that requires only a single observation for image SR was recently proposed. The authors found that convolutional layers can be used as a prior for image restoration tasks such as SR, denoising and inpainting. DIP optimizes the CNNs in a self-supervised training scheme without the use of ground-truth image. By minimizing the pixel-wise difference between the reconstructed image and the target image, DIP generates a natural image with fine details. However, DIP-based SR often fails when the target LR image contains noise. Moreover, the performance of DIP relies heavily on early-stopping. In contrast to DIP, our method can restore a clean HR image from a noisy LR image without early-stopping.

3. Proposed Method

Our goal is to restore a clean HR image from a noisy LR image based on the DIP framework. In this section, we first introduce a DIP for a SR task (DIP-SR) [1], which is closely related to our work. We further analyze why DIP-SR fails to restore a high-quality image from a given noisy LR image. We then describe the proposed noise estimation method, which effectively reduces the noise elements while performing image SR. We subsequently describe our novel loss function, called the self-supervision loss (SSL), which helps to provide a stable optimization process in our network. Finally, we introduce the total loss.

3.1. Deep Image Prior (DIP)

Given an input LR image $I^{LR} \in \mathbb{R}^{H\times W\times C}$ and the scaling factor s, DIP-SR [1] generates a HR image $I^{HR} \in \mathbb{R}^{sH\times sW\times C}$. By using a generator G, a code vector $z \in \mathbb{R}^{sH\times sW\times C}$ is mapped to a super-resolved image $\hat{I}^{HR} \in \mathbb{R}^{sH\times sW\times C}$ as $\hat{I}^{HR} = G(z)$. The reconstruction loss for measuring the error between the downsampled generated image and I^{LR} is defined as follows:

$$L_{rec} = ||DS(\hat{I}^{HR}) - I^{LR}||^2, \tag{2}$$

where $DS(\cdot)$ is a downsampler with scaling factor s. Since DIP uses the most common downsampling operators, such as Lanczos, the downsampler is not trainable.

However, when DIP-SR attempts to super-resolve a LR image that has noise, $DS(\hat{I}^{HR})$ is likely to be noisy because a pixel-wise comparison between $DS(\hat{I}^{HR})$ and I^{LR} is performed in the reconstruction loss (Equation (2)). Since DS is not trainable, $DS(\hat{I}^{HR})$ is dependent only on $\hat{I}^{HR}$. Thus, the fact that $DS(\hat{I}^{HR})$ contains noise signifies that $\hat{I}^{HR}$ also contains noise. In addition, we observe that there exists a point at which the quality of the reconstructed image deteriorates as the optimization process proceeds further. From that point, the output is overfitted to the noisy input image and the performance of DIP deteriorates noticeably. This observation emphasizes that DIP early-stopping is required in DIP in order to obtain a reasonable result. However, it is difficult to determine when to stop the optimization process if the clean image is absent.

To this end, both a solution to handle noise in the target image and a method to avoid early-stopping are required for DIP [1]. In order to address these problems, we first propose a noise estimation method to help our generator estimate the noise in the target image using the GAN [8] framework in Section 3.2. We also propose a self-supervision loss, which provides a stable optimization process and is described in detail in Section 3.3. Finally, the total loss and the algorithm of our framework are introduced in Section 3.4.

3.2. Noise Estimation Using GAN

In general, a noisy image I_N can be modeled as the summation of the clean image I_C and noise n as follows.

$$I_N = I_C + n. \tag{3}$$

The noisy LR image can be handled more easily if the noise can be estimated and extracted. We therefore propose a GAN-based [8] noise estimation method to separate the noise from the reconstructed image.

As illustrated in Figure 2, our framework consists of a generator G and discriminator D. Given a noisy LR image I_N^{LR}, our generator G maps a code vector z to the reconstructed image $\hat{I}_C^{HR}$.

$$\hat{I}_C^{HR} = G(z). \tag{4}$$

For comparison with the target LR image, $\hat{I}_C^{HR}$ is downsampled to $\hat{I}_C^{LR}$ through the downsampler $DS(\cdot)$. Our discriminator D is trained to generate the probability y for predicting whether the input noise n_{in} is real or fake as $y = D(n_{in})$. In the case that n_{in} is real, then y becomes y_{real}. If n_{in} is fake, y becomes y_{fake}. While the real noise sample n is generated synthetically, the fake noise sample can be extracted as follows.

$$\hat{n} = I_N^{LR} - DS(G(z)) = I_N^{LR} - \hat{I}_C^{LR}. \tag{5}$$

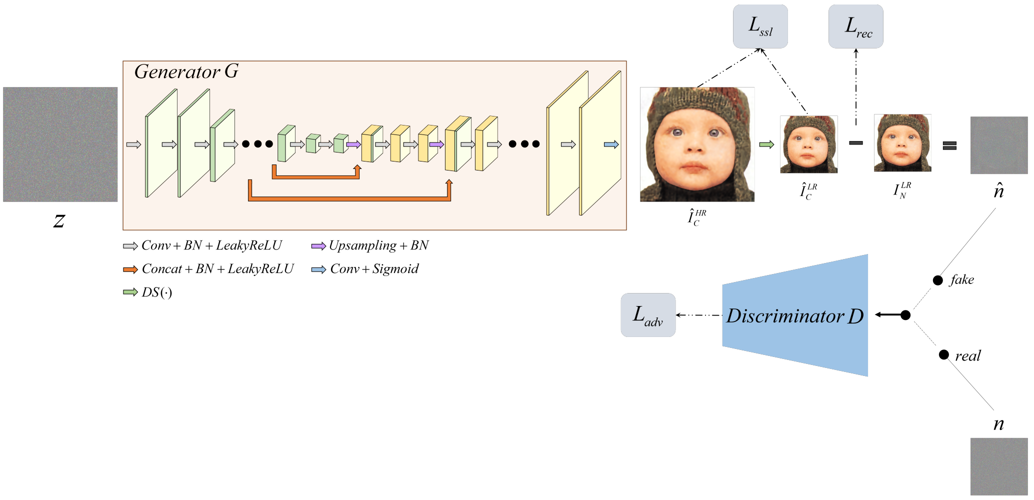

Figure 2. Overall architecture of the proposed model. Through a generator G, the input code tensor z is mapped to a noiseless HR image. For comparison with the target image, the generated image is downsampled via the downsampler. The discriminator D encourages G to learn the noise distribution in the target image.

The extracted noise $\hat{n}$ is made to follow the distribution of the real noise using the GAN framework. Adopting the WGAN loss [26], which stabilizes the optimization, the min-max game between the generator G and discriminator D is defined as follows:

$$\min_G \max_D \mathbb{E}_n[D(n)] - \mathbb{E}_{\hat{n}}[D(\hat{n})], \tag{6}$$

where $\mathbb{E}[\cdot]$ represents the expectation operation. Finally, the adversarial loss is defined as the following.

$$L_{adv} = -\mathbb{E}_{\hat{n}}[D(\hat{n})]. \tag{7}$$

This adversarial loss L_{adv} penalizes the generator G by using the distance between the distribution of n and the distribution of the extracted sample $\hat{n}$.

3.3. Self-Supervision Loss (SSL)

In general, noise has low self-similarity and high entropy because it contains no structure. Unlike noise, signals have high self-similarity and low entropy [27]. In a previous study on DIP [1], it was found that the parameters of CNNs have high impedance to noise and low impedance to signals. Due to this, when the target image is noisy, CNNs learn the signals in the early stage of the DIP optimization process before learning the noise components.

Inspired by this property of CNNs, we present a novel loss function called the self-supervision loss. The proposed framework is optimized through several iterations. In each optimization step, the proposed network outputs the reconstructed image. We hypothesize that the result of an earlier stage can be used as a constraint to reconstruct a noiseless HR image for the following stage. Accordingly, our self-supervision loss utilizes the output of the previous iteration step during training. SSL compares the output image of the current step with that of the previous step. By performing this, the reconstructed image maintains the learned signal without following the noise in the target image. In other words, by adding a constraint to the output image to preserve the learned signal, we avoid early-stopping and a dependency on the number of steps. The SSL for each step is defined as follows:

$$L_{ssl} = ||\hat{I}^{HR}_{C,i} - \hat{I}^{HR}_{C,i-1}||^2 + ||\hat{I}^{LR}_{C,i} - \hat{I}^{LR}_{C,i-1}||^2, \tag{8}$$

where $\hat{I}^{HR}_{C,i}$ and $\hat{I}^{LR}_{C,i}$ represent $\hat{I}^{HR}_{C}$ and $\hat{I}^{LR}_{C}$ at the ith optimization step, respectively.

3.4. Total Loss Functions

Our total loss function L_{total} consists of the reconstruction loss L_{rec} (Equation (2)), the adversarial loss L_{adv} (Equation (7)) and the self-supervision loss L_{ssl} (Equation (8)) as follows:

$$L_{total} = L_{rec} + \lambda_{adv} L_{adv} + \lambda_{ssl} L_{ssl}, \tag{9}$$

where λ_{adv} and λ_{ssl} are hyperparameters that are empirically set as 1.2 and 1, respectively.

The proposed algorithm for our framework is summarized in Algorithm 1. z and n are sampled from the uniform distribution U and Gaussian distribution G, respectively. We solve the SR problem with a noisy image in the case where the noise distribution and noise level σ are known. The code tensor z is perturbed with additional noise before z enters the network. At each iteration, we first train the discriminator. Our generator is then trained using Equation (9). Note that randomly-initialized parameters are used in the downsampler $DS(\cdot)$.

Algorithm 1: Training scheme of proposed method.

```
Require: Maximum iteration number $T$, noise level $\sigma$, noisy LR image $I^{LR}_N$,
    randomly-initialized Generator $G^0$, randomly-initialized Downsampler $DS$,
    randomly-initialized Discriminator $D^0$
$z \leftarrow U(0, 0.1)$
$n \leftarrow N(0, \sigma)$
for $i = 0$ to $T$ do
    perturb $z$
    $\hat{n} \leftarrow I^{LR}_N - DS(G^i(z))$
    Calculate the discriminator loss using Equation (6)
    Compute the gradient w.r.t. $D^i$
    Update the parameters of $D^i$
    perturb $z$
    $\hat{I}^{LR}_C \leftarrow DS(G^i(z))$
    Calculate the reconstruction loss using Equation (2)
    $\hat{n} \leftarrow I^{LR}_N - DS(G^i(z))$
    Calculate the adversarial loss using Equation (7)
    if $i = 0$ then
        $L_{ssl} \leftarrow 0$
        $\hat{I}^{HR}_{C,0} \leftarrow G^i(z)$
        $\hat{I}^{LR}_{C,0} \leftarrow DS(G^i(z))$
    else
        $\hat{I}^{HR}_{C,i} \leftarrow G^i(z)$
        $\hat{I}^{LR}_{C,i} \leftarrow DS(G^i(z))$
        Calculate the self-supervision loss using Equation (8)
    end if
    Calculate the total loss for generator using Equation (9)
    Compute the gradient w.r.t. $G^i$
    Update the parameters of $G^i$
end for
$I^{HR}_C \leftarrow G^T(z)$
return Clean HR image $I^{HR}_C$
```

4. Experimental Results

4.1. Dataset

We evaluate our method on the general SR test sets including Set5 [11] and Set14 [12]. Unlike the existing SR methods, our approach considers the degradation of the given LR image by noise. Therefore, we prepare LR noisy images by downsampling the HR images by a factor s and then adding Gaussian noise of level σ. In order to evaluate the general SR performance for various degradations, we use multiple upsampling factors (i.e., $s = \times 2, \times 4$) and noise levels (i.e., $\sigma = 15, 25$).

4.2. Implementation Details

Our framework is implemented in Pytorch [28]. The proposed generator is similar to U-net [29] and the discriminator is the same as a Markovian discriminator [30] with a patch size of 11×11. In order to train both the generator and the discriminator, we adopt the Adam optimizer [31]. The learning rates are set to 1×10^{-2} for the generator and 1×10^{-4} for the discriminator. We optimize the generator and discriminator by using our objectives for 2000 iterations in the same manner as DIP-SR [1]. We use a single NVIDIA TITAN XP GPU for every single image in all the experiments.

4.3. Comparison with Existing Methods

We compare our approach with various SR methods such as DIP [1] and data-driven DL methods (i.e., DRLN [22], HAN [32] and SAN [33]). Two different sets of experiments with DIP were performed because the denoising problem and the SR problem were solved individually by DIP using different architectures and optimization methods. The first set involved the use of DIP for the SR task with noisy LR images; this is denoted as DIP-SR. The second set involved the sequential applications of two DIP networks, which were used for noise removal and SR, and denoted as DIP-Seq. For DIP-Seq, we optimize DIP for noise reduction and the SR task over 1800 iterations and over 2000 iterations, respectively. All experiments are performed with the authors' official code.

4.3.1. Quantitative Comparison

We evaluate the performance of our method using PSNR, SSIM [34] and FSIM [35], which are widely used in image quality assessment. Table 1 shows the quantitative comparisons for Set5 [11] and Set14 [12] at scaling factors of ×2 and ×4 and the noise levels $\sigma = 15$ and $\sigma = 25$. From the results, it can be observed that our method significantly outperforms the existing methods and achieves the best performance at all scaling factors and noise levels, except at $s = 2$ and $\sigma = 15$ on the Set14 dataset. The results for the SR methods (i.e., DIP-SR [1], DRLN [22], HAN [32] and SAN [33]) show that the existing approaches are vulnerable to noise in images. Even when the noise level is low (i.e., $\sigma = 15$), their performances are significantly worse than that of our method (see Table 1). When DIP was sequentially applied for noise removal and DIP-SR, the performance improves compared to DIP-SR (compare the results of DIP-SR and DIP-Seq in Table 1). However, the performance is still not as good as that of our method. We attribute the superior performance of our method to the effects of our GAN [8] framework in which the discriminator encourages the generator to reconstruct a clean output image and estimates the noise. In addition, the results show that the proposed self-supervision loss $\mathcal{L}_{ssl}$ in Equation (8) permits a more reliable optimization of the existing DIP algorithm for image restoration.

4.3.2. Qualitative Comparison

Visual comparisons are shown in Figures 3–6. The results of the bicubic upsampling method suffer significantly from noise. This is because the resulting images are generated using the pixel values of the given image, which contains the unexpected noise. The results of DIP-SR, DRLN, HAN and SAN clearly show the side effects of the existing SR algorithms that amplifies the noise when input images are contaminated by noise (see the second, fourth, fifth and sixth columns in Figures 3–6). By contrast, the proposed method restores clean SR images that are close to ground truth. As shown in third columns in Figures 3–6, the results of DIP-Seq are less noisy than those from existing SR methods. However, noise artifacts still exist prominently in the resulting images. This indicates that the sequential optimization using two DIP networks for noise and SR is insufficient for handling both noise and SR. By contrast to the results of existing methods, we effectively remove the noise during the SR process and achieve clean HR image.

Table 1. Quantitative comparisons on Set5 [11] and Set14 [12]. The best results are highlighted in bold.

Method	**Scale**	**Noise**	**Set5**			**Set14**		
			PSNR	**SSIM**	**FSIM**	**PSNR**	**SSIM**	**FSIM**
Bicubic	×2	$\sigma = 15$	25.74	0.8447	0.8620	24.44	0.7723	0.8831
DRLN [22]	×2	$\sigma = 15$	22.03	0.7136	0.7545	21.40	0.6592	0.8241
HAN [32]	×2	$\sigma = 15$	21.81	0.7055	0.7519	21.19	0.6488	0.8206
SAN [33]	×2	$\sigma = 15$	22.06	0.7162	0.7573	21.36	0.6575	0.8237
DIP-SR [1]	×2	$\sigma = 15$	23.07	0.7680	0.7881	22.59	0.7125	0.8561
DIP-Seq [1]	×2	$\sigma = 15$	26.97	0.9050	**0.8926**	**25.64**	**0.8253**	**0.9100**
Ours	×2	$\sigma = 15$	**27.81**	**0.9127**	0.8886	24.96	0.7871	0.8658
Bicubic	×2	$\sigma = 25$	22.91	0.7473	0.7882	22.06	0.6703	0.8212
DRLN [22]	×2	$\sigma = 25$	17.71	0.5438	0.6181	17.38	0.4925	0.7187
HAN [32]	×2	$\sigma = 25$	17.73	0.5413	0.6273	17.29	0.4850	0.7214
SAN [33]	×2	$\sigma = 25$	17.73	0.5444	0.6284	17.24	0.4858	0.7214
DIP-SR [1]	×2	$\sigma = 25$	18.35	0.5676	0.6478	18.44	0.5330	0.7469
DIP-Seq [1]	×2	$\sigma = 25$	22.36	0.7695	0.7872	23.08	0.7367	**0.8643**
Ours	×2	$\sigma = 25$	**26.72**	**0.8906**	**0.8806**	**24.15**	**0.7631**	0.8495
Bicubic	×4	$\sigma = 15$	22.81	0.7862	0.7945	21.81	0.6553	0.7954
DRLN [22]	×4	$\sigma = 15$	20.77	0.6913	0.7425	19.85	0.5931	0.7513
HAN [32]	×4	$\sigma = 15$	20.92	0.6909	0.7453	19.88	0.5900	0.7538
SAN [33]	×4	$\sigma = 15$	20.58	0.6804	0.7430	19.75	0.5745	0.7533
DIP-SR [1]	×4	$\sigma = 15$	21.43	0.7153	0.7627	20.69	0.6241	0.7874
DIP-Seq [1]	×4	$\sigma = 15$	22.86	0.7960	0.8084	22.23	0.6988	0.8372
Ours	×4	$\sigma = 15$	**25.13**	**0.8710**	**0.8457**	**23.26**	**0.7742**	**0.8414**
Bicubic	×4	$\sigma = 25$	21.04	0.7025	0.7563	20.31	0.5933	0.7549
DRLN [22]	×4	$\sigma = 25$	16.91	0.5312	0.6234	16.15	0.4359	0.6373
HAN [32]	×4	$\sigma = 25$	17.31	0.5371	0.6360	16.66	0.4466	0.6529
SAN [33]	×4	$\sigma = 25$	16.95	0.5242	0.6330	16.29	0.4343	0.6463
DIP-SR [1]	×4	$\sigma = 25$	17.58	0.5421	0.6479	17.16	0.4610	0.6753
DIP-Seq [1]	×4	$\sigma = 25$	18.83	0.6150	0.6976	18.76	0.5481	0.7428
Ours	×4	$\sigma = 25$	**22.03**	**0.7696**	**0.7909**	**21.10**	**0.6589**	**0.7931**

4.3.3. Runtime Comparison

As shown in Table 2, we compare the runtime of our method with those of existing methods. The measured runtime is the average value over 10 images with the size of 256 × 256 × 3 on a PC with a single NVIDIA Titan XP GPU. Even though data-driven DL methods show fast inference time, they require training time for a large dataset. Note that since our method optimizes the network only for a given image, we do not need additional training time. The runtime of our method is similar to DIP-SR. However, our runtime is faster than that of DIP-Seq because DIP-Seq sequentially performs noise removal and SR, while our method efficiently generates noise-free SR images.

Table 2. Comparison of the averaged runtime when the size of the input LR image is 256 × 256 × 3.

Method	**DRLN [22]**	**HAN [32]**	**SAN [33]**	**DIP-SR [1]**	**DIP-Seq [1]**	**Ours**
Runtime (s)	0.663	1.258	0.946	149.815	225.087	146.334

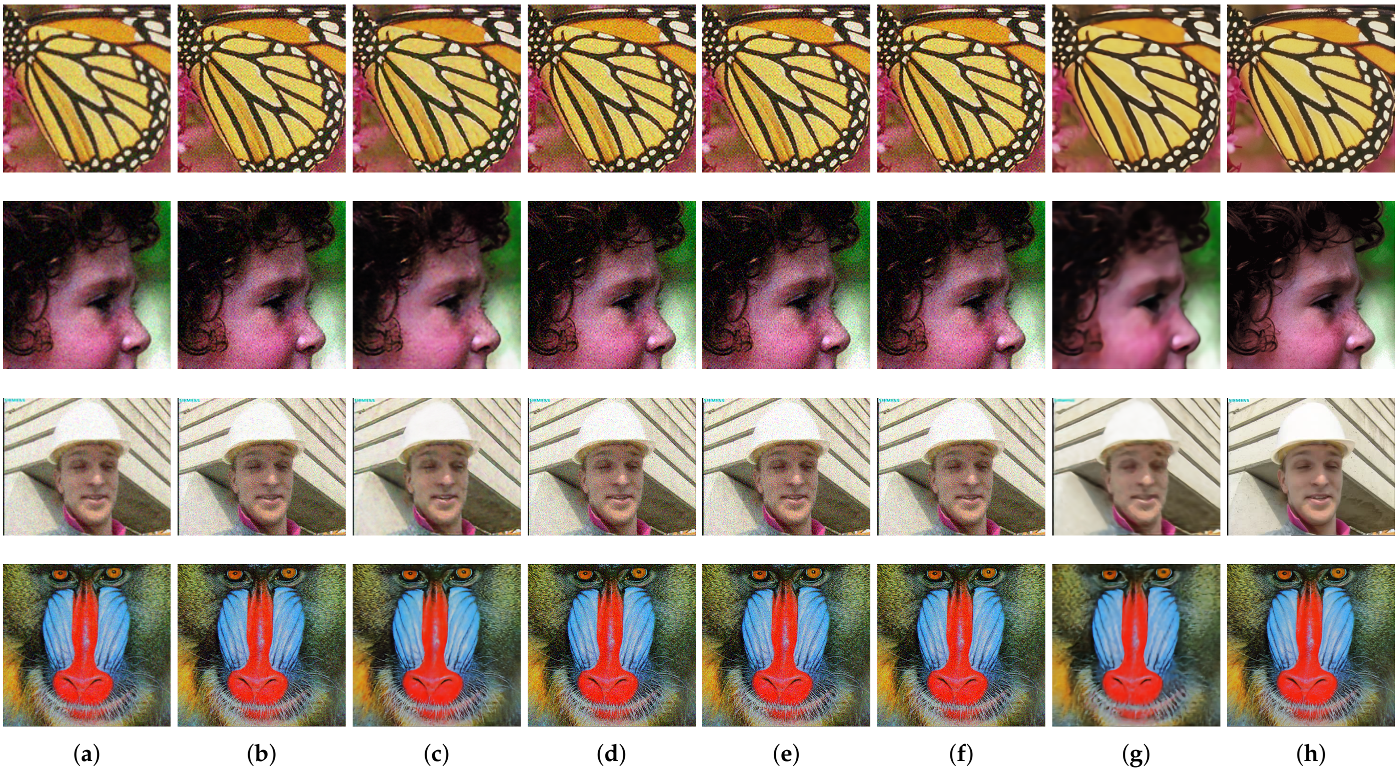

Figure 3. Qualitative comparisons of Set5 [11] and Set14 [12] ($\times 2, \sigma = 15$). (**a**) Bicubic, (**b**) DIP-SR [1], (**c**) DIP-Seq, (**d**) DRLN [22], (**e**) HAN [32], (**f**) SAN [33], (**g**) Ours and (**h**) Ground truth.

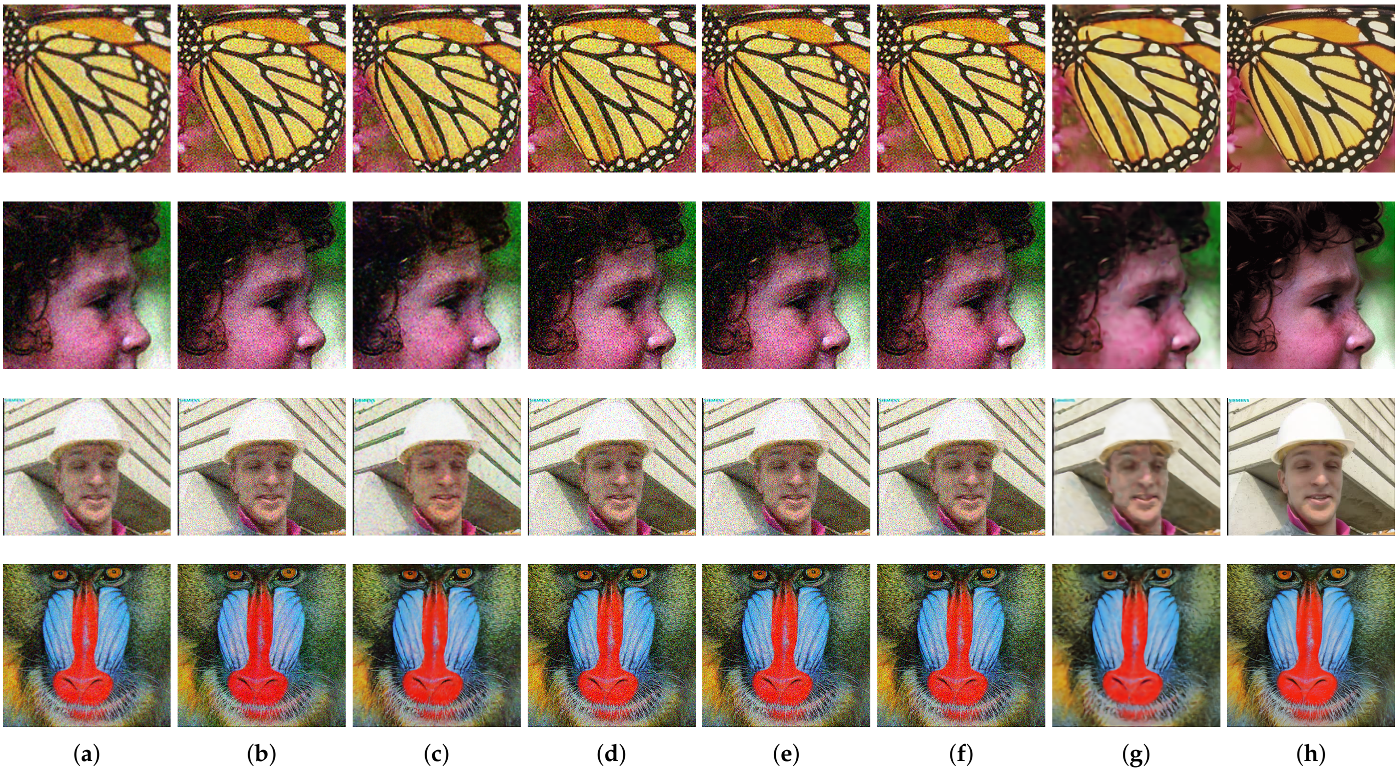

Figure 4. Qualitative comparisons of Set5 [11] and Set14 [12] ($\times 2, \sigma = 25$). (**a**) Bicubic, (**b**) DIP-SR [1], (**c**) DIP-Seq, (**d**) DRLN [22], (**e**) HAN [32], (**f**) SAN [33], (**g**) Ours and (**h**) Ground truth.

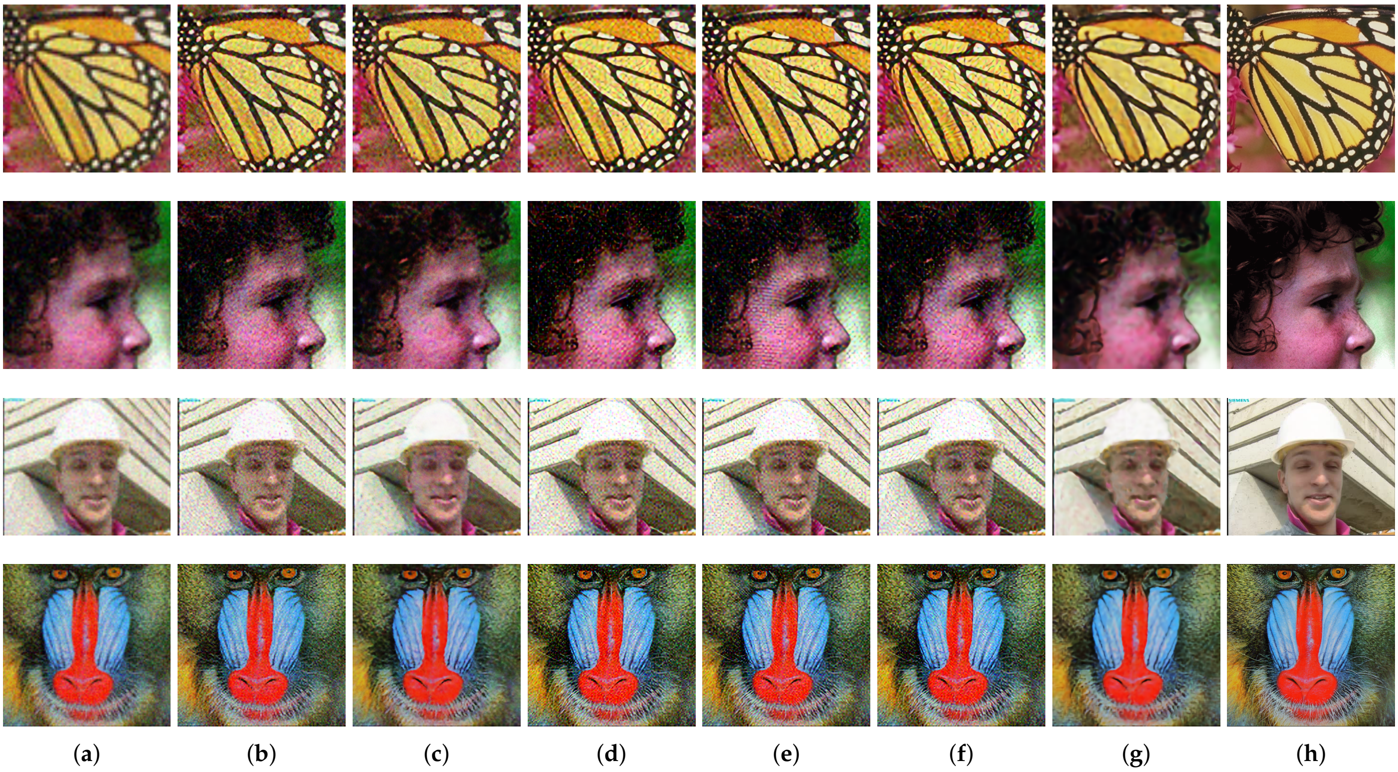

Figure 5. Qualitative comparisons of Set5 [11] and Set14 [12] ($\times 4, \sigma = 15$). (**a**) Bicubic, (**b**) DIP-SR [1], (**c**) DIP-Seq, (**d**) DRLN [22], (**e**) HAN [32], (**f**) SAN [33], (**g**) Ours and (**h**) Ground truth.

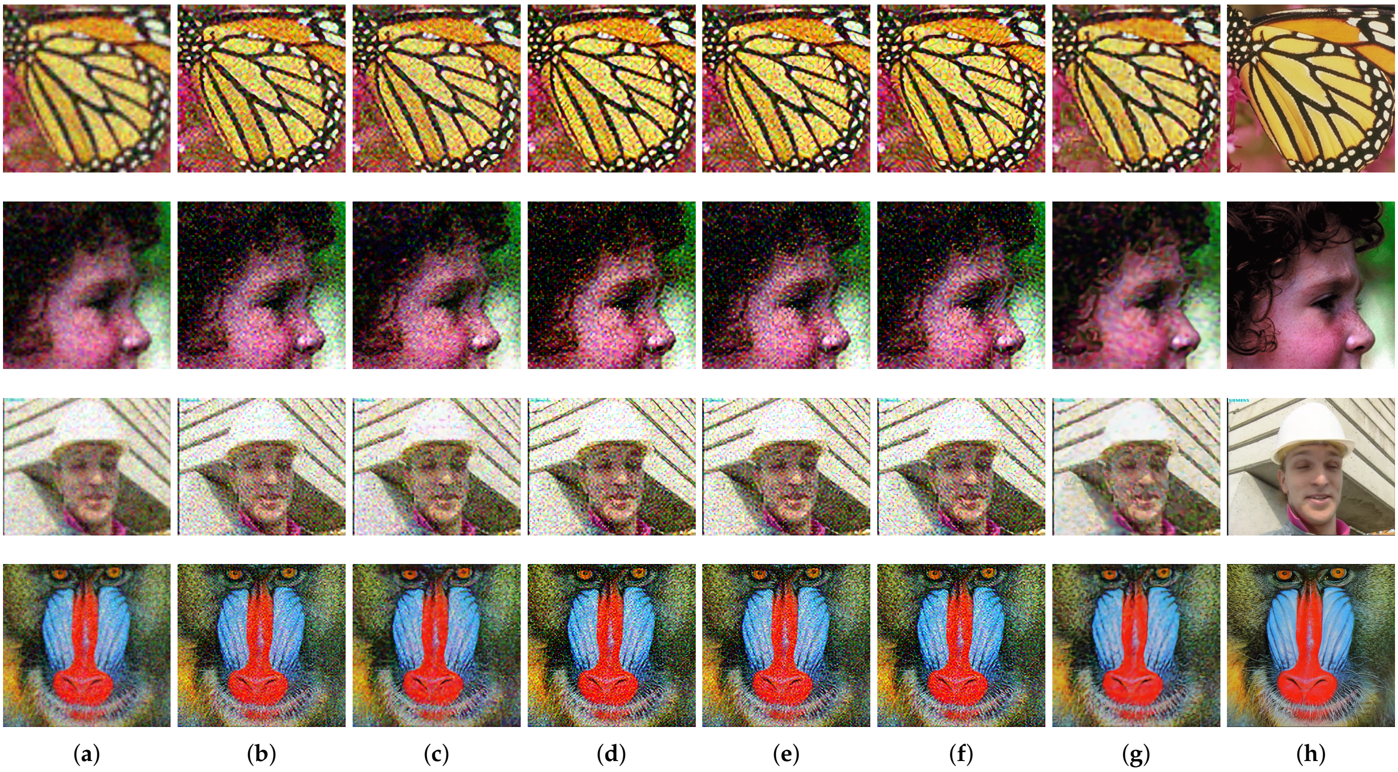

Figure 6. Qualitative comparisons of Set5 [11] and Set14 [12] ($\times 4, \sigma = 25$). (**a**) Bicubic, (**b**) DIP-SR [1], (**c**) DIP-Seq, (**d**) DRLN [22], (**e**) HAN [32], (**f**) SAN [33], (**g**) Ours and (**h**) Ground truth.

4.4. Ablation Study

We propose a noise estimating framework using GAN [8] to estimate the noise and a SSL to provide stable optimization for DIP [1]. In order to demonstrate the effectiveness of our method, we conduct ablation studies by gradually adding the noise estimation method (i.e., Equation (7)) and SSL (Equation (8)) based on the reconstruction loss (Equation (2)). For the ablation studies, the scale factor and noise level are set to 2 and 25, respectively.

As depicted in Figure 7, when only reconstruction loss is used, the optimization process becomes overfitted within approximately 500 iterations, resulting in poor performance. When the noise estimation method is applied, the optimization process performs stably without overfitting in the early stage. Furthermore, after 800 iterations, our framework with the noise estimation method outperforms the case that only reconstruction loss is used. The final proposed model, which includes both the noise estimation method and SSL, not only shows the most stable optimization process but also achieves the best performance. At the 2000th iteration, our final model shows the best performance compared to the other methods. Although the number of iterations is set to 2000 in DIP [1], we optimized each algorithm to run for 3500 iterations to show the independence from early-stopping. We can therefore confirm that even when the number of iterations exceeds 2000, the performance of our method improves steadily.

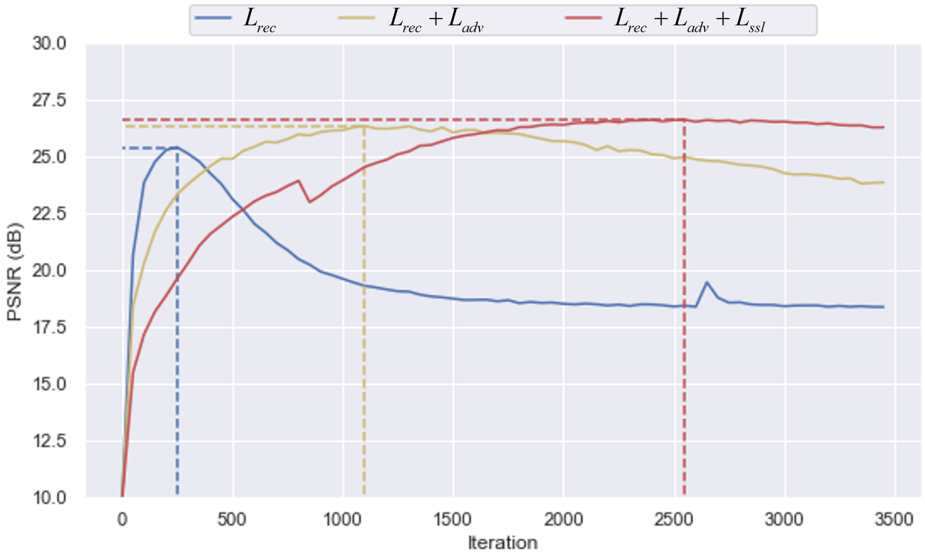

Figure 7. PSNR vs. iteration plot. The plot demonstrates the instability of DIP and the ability of our self-supervision loss to stabilize the optimizing process and avoid overfitting.

The results in Figure 8 clearly show the qualitative effectiveness of our proposed method. In the early stages, when only the reconstruction loss is used, the results are optimized more quickly than those from our method (see the results of 100 and 600 iterations in Figure 8). However, as the iteration progresses, the generator reconstructs more unwanted noise elements, resulting in unpleasant images. Therefore, its results suffer from the presence of noise components in the target image. By contrast, when only the noise estimation method is applied, the results were reliably restored as the iterations proceeded. In this case, the noise elements are observed at approximately 1300 iterations. In comparison, our final model, which adopts both the noise estimation method and SSL, can restore the details well without generating noise elements until 2000 iterations. The PSNR, SSIM and FSIM results are shown in Table 3. When we additionally use the noise estimation method, our method performs much better than when only the reconstruction loss is used with average increase in PSNR of 7.5 dB, SSIM of 0.3094 and FSIM of 0.2217. After the additional adoption of SSL, our final method generates higher quality HR images with average increase in PSNR of 0.87 dB, SSIM of 0.0136 and FSIM of 0.0111.

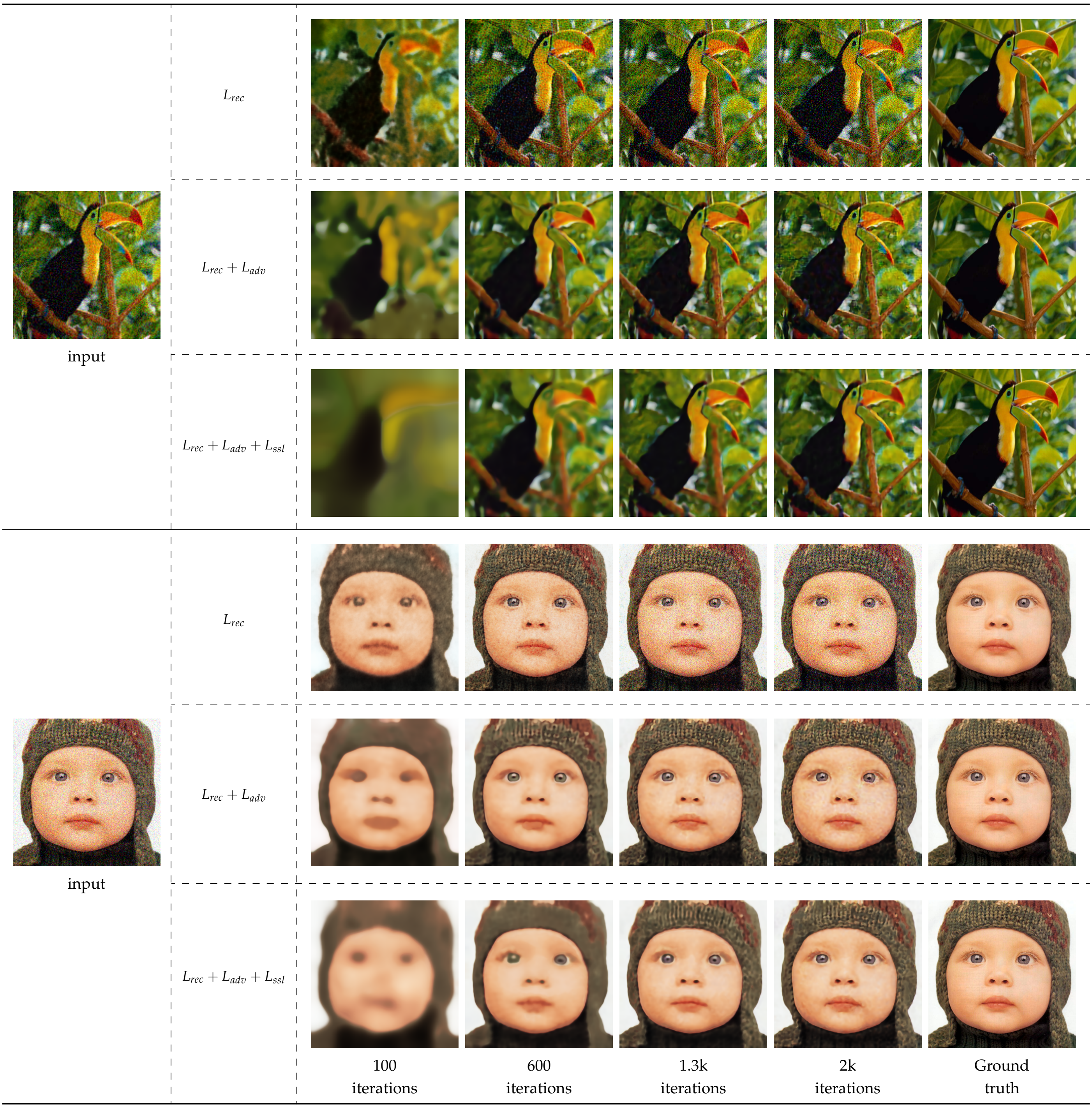

Figure 8. Ablation study for "bird" image and "baby" image in SET5 dataset ($s = 2$, $\sigma = 25$).

Table 3. Ablation study on the Set5 [11] dataset ($s = 2, \sigma = 25$). The best results are highlighted in bold.

Method	**Loss**	**Baby**			**Bird**			**Butterfly**			**Head**			**Woman**			**Avg.**		
		PSNR	**SSIM**	**FSIM**	**PSNR**	**SSIM**	**FSIM**	**PSNR**	**SSIM**	**FSIM**	**PSNR**	**SSIM**	**FSIM**	**PSNR**	**SSIM**	**FSIM**	**PSNR**	**SSIM**	**FSIM**
Baseline	L_{rec}	19.32	0.5766	0.7806	18.43	0.6129	0.6041	17.36	0.7302	0.6175	18.54	0.3808	0.6051	18.12	0.5375	0.6319	18.35	0.5676	0.6478
+ noise estimation	$L_{rec} + L_{adv}$	27.94	**0.9036**	**0.9335**	25.49	0.8921	0.8382	23.92	0.9253	0.8580	26.50	0.7661	**0.8449**	25.38	0.8980	0.8729	25.85	0.8770	0.8695
+ noise estimation + SSL	$L_{rec} + L_{adv} + L_{ssl}$	**28.09**	0.8983	0.9226	**26.67**	**0.9129**	**0.8824**	**24.91**	**0.9401**	**0.8854**	**27.23**	**0.7840**	0.8233	**26.68**	**0.9175**	**0.8893**	**26.72**	**0.8906**	**0.8806**

5. Conclusions

In this paper, we propose a DIP based noise-robust SR method. Our framework combines a noise estimation method and the self-supervision loss with DIP-SR. By adopting the proposed noise estimating method, the noise in the given LR target image can be estimated. The use of the self-supervision loss increases the stability of the optimization process. By using extensive experiments, it can be concluded that our method achieves outstanding performance both quantitatively and qualitatively.

Author Contributions: Conceptualization, S.H., T.B.L. and Y.S.H.; software, S.H.; validation, S.H.; investigation, S.H. and T.B.L.; writing—original draft preparation, S.H. and T.B.L.; writing—review and editing, S.H., T.B.L. and Y.S.H.; supervision, Y.S.H. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Ministry of Science and ICT (MSIT), South Korea, under the Information Technology Research Center (ITRC) Support Program supervised by the Institute for Information and Communications Technology Promotion (IITP) under Grant IITP-2021-2018-0-01424.

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

References

1. Ulyanov, D.; Vedaldi, A.; Lempitsky, V. Deep image prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 9446–9454.
2. Ma, X.; Hong, Y.; Song, Y. Super resolution land cover mapping of hyperspectral images using the deep image prior-based approach. *Int. J. Remote Sens.* **2020**, *41*, 2818–2834. [CrossRef]
3. Sidorov, O.; Yngve Hardeberg, J. Deep hyperspectral prior: Single-image denoising, inpainting, super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, 27–28 October 2019.
4. Sagel, A.; Roumy, A.; Guillemot, C. Sub-Dip: Optimization on a Subspace with Deep Image Prior Regularization and Application to Superresolution. In Proceedings of the ICASSP 2020—IEEE International Conference on Acoustics, Barcelona, Spain, 4–8 May 2020; pp. 2513–2517. [CrossRef]
5. Mataev, G.; Milanfar, P.; Elad, M. Deepred: Deep image prior powered by red. In Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops, Seoul, Korea, 27–28 October 2019.
6. Abdelhamed, A.; Lin, S.; Brown, M.S. A high-quality denoising dataset for smartphone cameras. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1692–1700.
7. Chen, J.; Chen, J.; Chao, H.; Yang, M. Image blind denoising with generative adversarial network based noise modeling. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018; pp. 3155–3164.
8. Goodfellow, I.J.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial networks. *arXiv* **2014**, arXiv:1406.2661.
9. Cattin, D.P. Image restoration: Introduction to signal and image processing. *MIAC Univ. Basel Retrieved* **2013**, *11*, 93.
10. Gandelsman, Y.; Shocher, A.; Irani, M. "Double-DIP": Unsupervised Image Decomposition via Coupled Deep-Image-Priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 11026–11035.
11. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi Morel, M.L. Low-Complexity Single-Image Super-Resolution based on Nonnegative Neighbor Embedding. In Proceedings of the British Machine Vision Conference, Surrey, UK, 3–7 September 2012; pp. 135.1–135.10. [CrossRef]
12. Zeyde, R.; Elad, M.; Protter, M. On Single Image Scale-Up Using Sparse-Representations. In Proceedings of the International Conference on Curves and Surfaces, Avigon, France, 24–30 June 2010; pp. 711–730.
13. 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.
14. 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.
15. Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef] [PubMed]
16. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photo-realistic 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; pp. 4681–4690.
17. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.

18. Guo, T.; Seyed Mousavi, H.; Huu Vu, T.; Monga, V. Deep wavelet prediction for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 104–113.
19. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision Workshops, Munich, Germany, 8–14 September 2018.
20. 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, 18–22 June 2018; pp. 2472–2481.
21. 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, Munich, Germany, 8–14 September 2018; pp. 286–301.
22. Anwar, S.; Barnes, N. Densely residual laplacian super-resolution. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [CrossRef] [PubMed]
23. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
24. Tong, T.; Li, G.; Liu, X.; Gao, Q. Image super-resolution using dense skip connections. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 4799–4807.
25. Huang, G.; Liu, Z.; Van Der Maaten, L.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
26. Arjovsky, M.; Chintala, S.; Bottou, L. Wasserstein generative adversarial networks. In Proceedings of the International Conference on Machine Learning, Sydney, Australia, 6–11 August 2017; pp. 214–223.
27. Fan, W.; Yu, H.; Chen, T.; Ji, S. OCT Image Restoration Using Non-Local Deep Image Prior. *Electronics* **2020**, *9*, 784. [CrossRef]
28. Paszke, A.; Gross, S.; Massa, F.; Lerer, A.; Bradbury, J.; Chanan, G.; Killeen, T.; Lin, Z.; Gimelshein, N.; Antiga, L.; et al. PyTorch: An Imperative Style, High-Performance Deep Learning Library. In Proceedings of the Advances in Neural Information Processing Systems, Vancouver, BC, Canada, 8–12 December 2019; pp. 8026–8037.
29. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
30. Li, C.; Wand, M. Precomputed real-time texture synthesis with markovian generative adversarial networks. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 702–716.
31. Kingma, D.P.; Ba, J. Adam: A method for stochastic optimization. In Proceedings of the International Conference on Learning Representations (Poster), San Diego, CA, USA, 7–9 May 2015.
32. Niu, B.; Wen, W.; Ren, W.; Zhang, X.; Yang, L.; Wang, S.; Zhang, K.; Cao, X.; Shen, H. Single image super-resolution via a holistic attention network. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 191–207.
33. 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.
34. Wang, Z.; Bovik, A.C.; Sheikh, H.R.; Simoncelli, E.P. Image quality assessment: From error visibility to structural similarity. *IEEE Trans. Image Process.* **2004**, *13*, 600–612. [CrossRef] [PubMed]
35. Zhang, L.; Zhang, L.; Mou, X.; Zhang, D. FSIM: A feature similarity index for image quality assessment. *IEEE Trans. Image Process.* **2011**, *20*, 2378–2386. [CrossRef] [PubMed]