

Dual Reconstruction with Densely Connected Residual Network for Single Image Super-Resolution

Chih-Chung Hsu

Department of Management Information Systems
 National Pingtung University of Science and Technology
 No. 1, Shuefu Road, Neipu, Pingtung 91201, Taiwan

cchsu@mail.npust.edu.tw

Chia-Hsiang Lin

Department of Electrical Engineering, & Institute of Computer and Communication Engineering
 National Cheng Kung University
 No.1, University Road, Tainan City 701, Taiwan

chiahsiang.steven.lin@gmail.com

Abstract

Deep learning-based single image super-resolution enables very fast and high-visual-quality reconstruction. Recently, an enhanced super-resolution based on generative adversarial network (ESRGAN) has achieved excellent performance in terms of both qualitative and quantitative quality of the reconstructed high-resolution image. In this paper, we propose to add one more shortcut between two dense-blocks, as well as add shortcut between two convolution layers inside a dense-block. With this simple strategy of adding more shortcuts in the proposed network, it enables a faster learning process as the gradient information can be back-propagated more easily. Based on the improved ESRGAN, the dual reconstruction is proposed to learn different aspects of the super-resolved image for judiciously enhancing the quality of the reconstructed image. In practice, the super-resolution model is pre-trained solely based on pixel distance, followed by fine-tuning the parameters in the model based on adversarial loss and perceptual loss. Finally, we fuse two different models by weighted-summing their parameters to obtain the final super-resolution model. Experimental results demonstrated that the proposed method achieves excellent performance in the real-world image super-resolution challenge. We have also verified that the proposed dual reconstruction does further improve the quality of the reconstructed image in terms of both PSNR and SSIM.

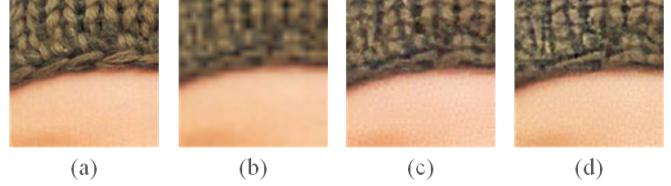


Figure 1. Visual quality comparison between the (a) ground truth and the super-resolved results by using (b) Bicubic interpolation, (c) ESRGAN [26], and (d) proposed method.

1. Introduction

Due to limitations in hardware design, or long distance between the object and the sensors (e.g., in satellite images), spatial resolution of the acquired images is quite often unsatisfactory. Without relying on hardware improvement, we aim at improving the image spatial resolution computationally. One category of image super-resolution (SR) techniques is referred to as panchromatic sharpening [15, 25], for which the camera acquires a single-band image (hence allowing very high-spatial resolution acquisition) of the same spatial region while acquiring the target imagery. Then, the acquired single-band image, known as panchromatic image, provides essential spatial details (or high-frequency components) that are injected/fused into the target imagery, thereby computationally super-resolving the target imagery.

This kind of image fusion approach is of paramount importance in nowadays satellite remote sensing technology, and also appears in other domain applications, such as the hyperspectral super-resolution technique [17]

which fuses a high-resolution multispectral image with a low-resolution hyperspectral image in order to obtain a high spatial/spectral-resolution imagery—critical in military surveillance. However, these approaches all rely on high-resolution counterpart imageries that are, however, not always available, motivating us to pursue single-image super-resolution (SISR) for facilitating the ensuing computer vision tasks [8].

Besides the well-known bicubic interpolation, several conventional SISR methods have been studies, including Lanczos resampling [4] and optimization based algorithms, where the latter usually considers the SISR problem as an imaging inverse problem [2]. SISR has also been widely studied in statistical machine learning area, which requires judiciously transforming suitable prior knowledge into computationally friendly regularization schemes [11], such as self-similarity [16] and total-variation [22].

Recently, deep learning has revealed its potential in achieving high performance SISR task. In the convolutional neural network (CNN) framework, an end-to-end sparse coding CNN (SRCNN), not requiring any engineered features, has also been proposed in [3]. Though SRCNN is a pioneering work introducing the deep learning in solving SR, it converges slowly and relies on the context of small regions. Furthermore, a very deep neural network SR (VDSR) adopting the deeper convolution neural network (CNN) was proposed to obtain high-quality reconstruction result and achieves state-of-the-art performance [9, 27, 28]. Among these methods, the ResNet [5] based SR model [28] shows the state-of-the-art performance, demonstrating that a stronger baseline network should be able to improve the performance as well. However, it is still hard to reconstruct the fine spatial details (such as textures) solely based on VDSR or its improved versions.

In [14], the authors first adopt generative adversarial networks (GAN) to synthesize the high-quality fine details, though these reconstructed details may slightly differ from the original ones, and this technique is known as SRGAN [14]. An enhanced version of SRGAN, termed as ESRGAN, is proposed to further improve the visual quality of the reconstructed image based on several learning tricks and Residual-in-Residual Dense Block (RRDB) [26].

Although the RRDB takes the advantages of both ResNet [5] and DenseNet [7], some shortcomings can be further improved. In ESRGAN, the shortcut connection is established between the input data and the feature map of the last convolution layer in an RRDB. In ResNet [5], the shortcut is connected between every two layers, which can fully take advantage of the effective gradient information. However, the shortcut connections in RRDB is relative few, compared to ResNet. To overcome this shortcoming, we add one more shortcut in an RRDB to ensure more effective learning of the SR model, called densely connected residual module.

On the other hand, several upsampling modules have been studied recently. However, it is hard to distinguish which upsampler should we use in the SR model. In this paper, instead, two upsamplers are selected to simultaneously super-resolve the feature maps of the base network. Afterward, two super-resolved feature maps will be fused by the last convolution layer to obtain the SR image in RGB channels, termed as dual reconstruction module. As shown in Fig. 1, the proposed method can provide more fine details of the reconstructed image based on training set provided by [21, 20]. In summary, the main contribution of the proposed method is two-fold:

1. We proposed a densely connected residual (DCR) network to better capture the hierarchical feature representations.
2. The proposed dual reconstruction module (DRM) can effectively super-resolve low-resolution (LR) image based on two different upsampler modules.

The rest of this paper is organized as follows. Some most relevant works are surveyed in Sec.2. Sec.2 presents the proposed SR model based on DCR and DRM. In Sec.3, experimental results are demonstrated. Finally, conclusions are drawn in Sec.4.

2. Proposed Super-Resolution Method

The overall flowchart of the proposed method is illustrated in Fig.2. First, we build the base network by composing several proposed DCR blocks to extract useful features for the LR image. Second, the extracted feature maps will be upsampled by the proposed DRM simultaneously. Afterward, two super-resolved feature maps can be estimated and added to each other. Finally, the super-resolved image is obtained by the last convolution layer with 3 channels.

Three loss functions are used in the proposed method. The first one is pixel loss, which is a common way to measure the similarity between the reconstructed image and its ground truth. Usually, the pixel loss is measured by l_1 norm distance [26, 14, 28], which is defined as

$$L_p(\mathbf{x}_{HR}, \mathbf{x}_{SR}) = \frac{1}{N^2 - 1} \sum_{i=1}^{N^2} |\mathbf{x}_{SR} - \mathbf{x}_{HR}|, \quad (1)$$

where N^2 is number of pixels of HR/SR image, \mathbf{x}_{SR} and \mathbf{x}_{LR} present the SR and LR images, respectively.

It is well-known that the GAN-based SR models can achieve better visual-quality of the reconstructed image, such as SRGAN [14] and ESRGAN [26]. In GAN-based SR models, the adversarial loss L_{GAN} incorporated in a GAN is defined as

$$\begin{aligned} & L_{GAN}(D, G) \\ &= E_D [\log D(\mathbf{x}_{HR})] + E_G [\log (1 - D(G(\mathbf{x}_{LR})))], \end{aligned} \quad (2)$$

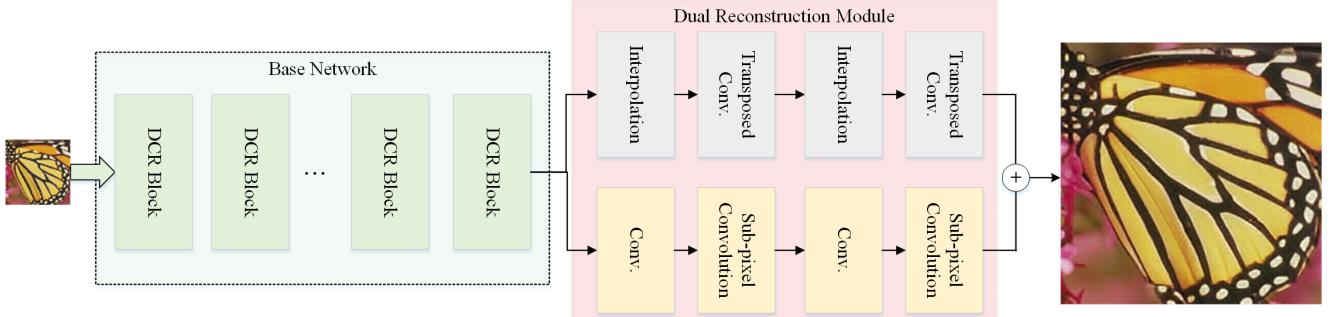


Figure 2. Diagram of the proposed method with the proposed dual reconstruction technique.

where D and G indicate the discriminator and generator, respectively, $G(\mathbf{x}_{LR})$ is the generative model used for generating the super-resolved image \mathbf{x}_{SR} , and $D(\mathbf{x})$ is the classification of data sample \mathbf{x} being visually-qualified.

Inspired by [26], the feature distance/similarity between the super-resolved image and its ground truth is usually used to measure the perceptual loss L_{feat} , as follows:

$$L_{feat}(\mathbf{x}_{HR}, \mathbf{x}_{SR}) = \frac{1}{M} \sum_{i=0}^{M-1} \|f_e(\mathbf{x}_{SR})(i) - f_e(\mathbf{x}_{HR})(i)\|_1, \quad (3)$$

where M is the number of feature vector and f_e is the feature extractor based on pre-trained Vgg16 model [24].

2.1. Densely Connected Residual Block

Assume that the input feature map (or data) is \mathbf{x} with n_c channels. In ESRGAN, each RRDB contains four convolutions $D = [f_{c1}, f_{c2}, f_{c3}, f_{c4}]$. The number of channels of the output feature maps at first three convolution layers are n_g and last convolution layer in a dense block is n_c . In the first three convolution layers will be concatenated [7] so that the number of the channels of outputs at the third convolution layer will be $3n_g$, while the last transition layer will squeeze the input feature map to the n_c -channeled feature map. Then, the shortcut will be constructed $\mathbf{x} + f_{c4}(f_{c3}(f_{c2}(f_{c1}(\mathbf{x})))$. To provide more gradient information and to accelerate the learning process, we propose to insert a transition layer f_{m2} to squeeze the feature map of $f_{c2}(f_{c1}(\mathbf{x}))$ to the n_c -channeled feature map. In this way, the second shortcut between f_m and \mathbf{x} can be constructed, as depicted in Fig. 3.

Similarly, the base network is constructed by several DCR blocks (3 DCR blocks in this paper), implying that the shortcut connection can also be added between DCR blocks. Since the number of channels of the feature maps of each DCR will be the same to each other, it is easy to construct the shortcut. In this study, the bypass is established between the first and second DCRs to take back-propagate

the gradient information easily, resulting the faster learning process of our SR model.

2.2. Dual High-Resolution Image Reconstruction

It is a well-known trick that the image resolution can be improved by learning the residual of the high-frequency component, instead of directly learn the SR result in RGB domain [10]. Also, there is a recent research trend in the remote sensing area (see, e.g., [17]), aiming at fusing the high-resolution spatial details (probably extracted from some counterpart panchromatic image) with an input low-resolution image. To take the advantages from different super-resolved results (or information from LR image), we need to have different SR models first. However, it is hard to train multiple SR models due to very time-consuming as well as design a frequency decomposition function to the deep neural network. Instead, we adopt two different upsamplers including the traditional interpolation function and sub-pixel convolution operation [23] to learn different super-resolved images from the feature maps of our base network, called double reconstruction technique.

Since sub-pixel convolution can be used to upsample the feature maps by aggregating feature responses at different channels, the information in feature maps can be fully discovered [23]. For example, the upscaling factor is 4, we need to give 16-channeled feature maps to upsample. Assumed that the number of the output feature map of our base network is n_c , a convolution layer was added to transfer the n_c -channeled input feature maps to $4n_c$ -channeled feature maps. Then, the upsampled feature map based on sub-pixel convolution will become n_c when the scaling factor is 2. In this manner, we can ensure that the number of channels of the upsampled feature map is n_c . Note that we can repeat the above process to obtain the upsampled results with scaling factor 4. Afterward, a convolution layer will be added to obtain the n_c -channeled feature maps. Finally, the super-resolved feature map can be obtained by concatenating two upsampled feature maps U_1 and U_2 .

In summary, the details of our DRM are listed as follows:

- First upsampler U_1 : It contains two transposed con-

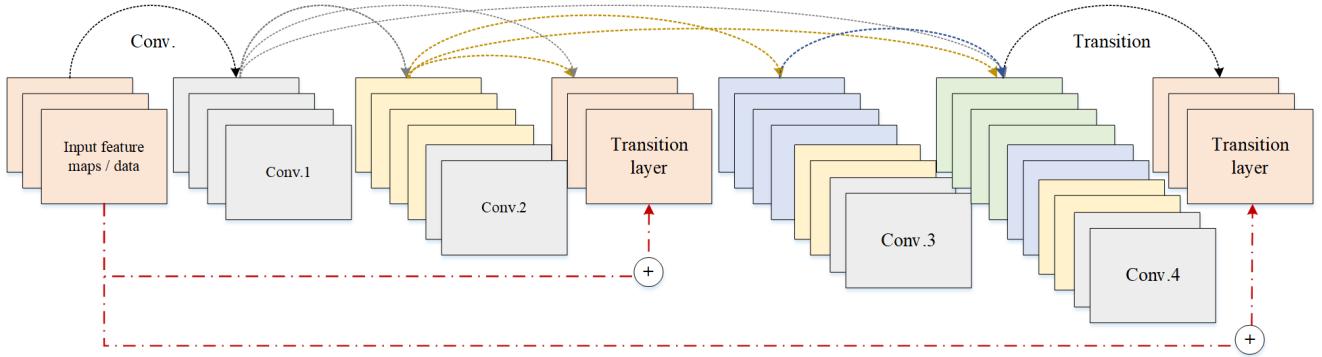


Figure 3. Network architecture of the proposed densely connected residual block.

volution layers and Bicubic Interpolation function, as depicted in the upper pipeline in Fig. 2.

- Second upsample U_2 : Three convolution layers and sub-pixel convolution [23] is used to upsample the input feature map of our base network, as shown in the bottom pipeline in Fig. 2.

Note that the DRM technique is not used in the AIM-Real-World-Super-Resolution-Challenge-Track1 (AIMRWSR).

2.3. Learning Tricks

In this paper, several useful learning-tricks are adopted to have better and faster network training, as follows:

- Inspired by ESRGAN [26], the adversarial learning is used to fine-tune our SR model based on the GAN loss in (2), where the Vgg-16 architecture [24] is used as our discriminator.
- We adopt H-Swish [6] activation in our discriminator to accelerate the learning process.
- We use unbalanced learning rates for the generator and discriminator in the second model (determined by manual check).
- The fully connected layers are removed from the discriminator. The global averaging pooling is used instead to better capture the spatial information of feature maps [18].
- Larger learning rate at the first 10,000 iterations and followed by relative small learning rate to train our SR model.
- Since the first SR model (say, SNR-aware model, SAM) is learned by (1) and the fine-tuned SR model (say, visual-aware model, VAM) is learned by (1)(2)(3), it is worth to fuse the results of VAM

and SAM to obtain the better super-resolved image. Inspired by [26], the final SR model can be fused by merging the parameters in tow models such that $\alpha \text{SAM}(\mathbf{x}_{LR}) + (1 - \alpha) \text{VAM}(\mathbf{x}_{LR})$ for some $\alpha \in [0, 1]$.

3. Experimental Results

3.1. Experiments Settings

In our solution, we adopt Python as the programming language. Since our method is based on deep learning, we use PyTorch as our backbone. We train the SR model based on a personal computer equipped with two GTX 2080Ti GPUs, Intel i9-9900 CPU, and 64 GB system memory. In this paper, two SR models are learned sequentially. The first model is based on (1). In this way, it should be able to learn the SAM SR network. Afterward, adversarial learning is adopted to learn the VAM SR model based on the trained SAM SR model.

In our SAM SR model, the initial learning rate is $3e - 4$ with SGDR learning decay policy [19]. The size of the ground truth at the first 400,000 iterations is 128×128 , while that at the following 400,000 iterations is 160×160 . The total iterations is 800,000 with Adam optimizer [12]. In the proposed VAM SR model, the initial learning rates of the generator and the discriminator are respectively $1e - 3$ and $1e - 4$. Since we found that the discriminator is faster to converge, we use the unbalance learning rate setting to avoid the learning difficulty. The size of the ground truth at the first 40,000 is 160×160 while that at the following 40,000 iterations is 128×128 . The total iterations are 80,000 with the learning rate decay by a ratio of 0.5 for every 5,000 iteration. The same experiment settings are adopted in ESRGAN [26] to have a fair comparison. Therefore, we select the trained model at iteration 80,000 as our final ESRGAN model.

The proposed SR model is trained based on AIMRWSR training set (source domain) [21, 20]. The scaling factor is 4. Therefore, the size of LR image is 128×128 if the size of

	Method	\uparrow PSNR	\uparrow SSIM	\downarrow LPIPS	\downarrow MOS
Participants	MadDemon	22.65	0.48	0.36	2.22
	IPCV IITM	25.15	0.60	0.66	2.36
	Nam	25.52	0.63	0.65	2.46
	CVML	24.59	0.62	0.61	2.47
	ACVLab-NPUST (ours)	24.77	0.61	0.60	2.49
	SeeCout	25.30	0.61	0.74	2.50
Baselines	Image Specific NN for RWSR	24.31	0.60	0.69	2.56
	ESRGAN Superv.	24.22	0.55	0.30	1.81
	Bicubic	25.34	0.61	0.73	2.43
	EDSR PT	25.14	0.60	0.67	
	ESRGAN PT	22.57	0.51	0.58	
	ESRGAN FT-Src	24.54	0.56	0.45	

Table 1. Challenge results for Track 1: Source domain on the final test set [21, 20]

Table 2. Performance comparison among the different SR methods evaluated on Set5 and Set14.

Method	Set5		Set14	
	PSNR	SSIM	PSNR	SSIM
ESRGAN [26]	27.13	0.648	24.62	0.589
SAM-SR (ours)	27.64	0.648	24.85	0.596
VAM-SR (ours)	27.85	0.648	25.25	0.607
SAM+VAM (ours)	28.03	0.649	25.60	0.616

corresponding HR image is 512×512 . To verify the performance of the reconstructed image during training phase, we evaluate the PSNR and SSIM based on the 6 selected images from DIV2k[1]. To further evaluate the performance, Set5 and Set14 in [13] are used as the test set.

3.2. Objective Quality Comparison

In AIMRWSR Challenge, the proposed DRM is disabled because it is hard to train the proposed SR model with DRM in a limited time. Table 3.2 shows the performance comparison between the proposed method (named ACVLab-NPUST) and other methods proposed by other participants [21, 20]. We show good performance in terms of PSNR/SSIM/LPIPS. Since the test set did not release yet, we compare the proposed method with enabled DRM with other methods based on Set5 and Set14. Table 2 shows the performance comparison between the proposed method and other SR models for Set5 and Set14. Our proposed method achieves better performance in terms of SSIM and PSNR, compared to ESRGAN [26]. In our VAM-SR model, the objective quality is lower than ESGAN because the GAN-based SR models tend to super-resolved the HR image with fine details, whereas the synthesized details may differ from that in the ground truth. Note that the result from ESRGAN is generated by fusing the ESRGAN and its pre-trained SR-ResNet model with $\alpha = 0.8$ [26].

3.3. Subjective Quality Comparison

Figure 4 presents the subjective quality comparison between the proposed method and ESRGAN. It is clear that

Table 3. Ablation study on the proposed method for Set5 and Set14.

Method	Set5		Set14	
	PSNR	SSIM	PSNR	SSIM
Baseline-I (SAM)	27.61	0.640	24.89	0.600
Baseline-I (VAM)	27.32	0.639	24.32	0.590
Baseline-I (VAM+SAM)	27.61	0.641	25.30	0.607
Baseline-II (SAM)	27.57	0.641	25.52	0.612
Baseline-II (VAM)	27.21	0.639	24.62	0.589
Baseline-II (VAM+SAM)	27.82	0.640	25.47	0.611
SAM-SR (ours)	27.64	0.648	24.85	0.596
VAM-SR (ours)	27.85	0.648	25.25	0.607
SAM+VAM (ours)	28.03	0.649	25.60	0.616

the proposed method (SAM+VAM) can provide fine details, while the ESRGAN also generates fine synthesized textures. However, the details in the reconstructed image based on our method is more realistic, compared to ESRGAN. It demonstrates that the DRM module indeed can be used to improve the visual quality of the reconstructed image further. In the limited iterations, ESRGAN can not learn the good visually-pleased results from LR images. Although the textures in the reconstructed image based on SAM-SR model cannot be fully recovered, the super-resolved results will be closer to the ground truth.

3.4. Ablation Study

In this subsection, the DCR and DRM are removed respectively to form the Baseline-I and Baseline-II. Two Baselines are trained on the same training set provided by AIMRWSR challenge [21, 20]. It is verified that the proposed DCR and DRM can be used to improve the visual and objective quality of the reconstructed image. It also shows that the DRM can fuse different HR details within the network, which may exist some potential cues to synthesize the better high-quality SR images.

On the other hand, we also show the PSNR curve for the Set5 during the training phase in Fig. 5. It is clear that the proposed SAM-SR model is relatively faster to converge, compared to the default number of iterations in ESRGAN [26] (say, 1,000,000 iterations), which indicates that the more shortcut connections should be able to accelerate the learning process as well as slightly improve the performance of the super-resolution task. In the proposed SAM-SR model, the PSNR value can be approximately improved to 28dB at 550,000 iteration, which is significantly smaller than the default iterations used in ESRGAN [26]. We also evaluate our method for the validation set in AIMRWSR challenge [21] based on the online validation server. The PSNR values of the reconstructed images of the proposed method and Baseline-I (i.e., the submitted version in the AIMRWSR challenge for final test set) are 25.60dB and 25.49dB, implying that our DRM can further improve the

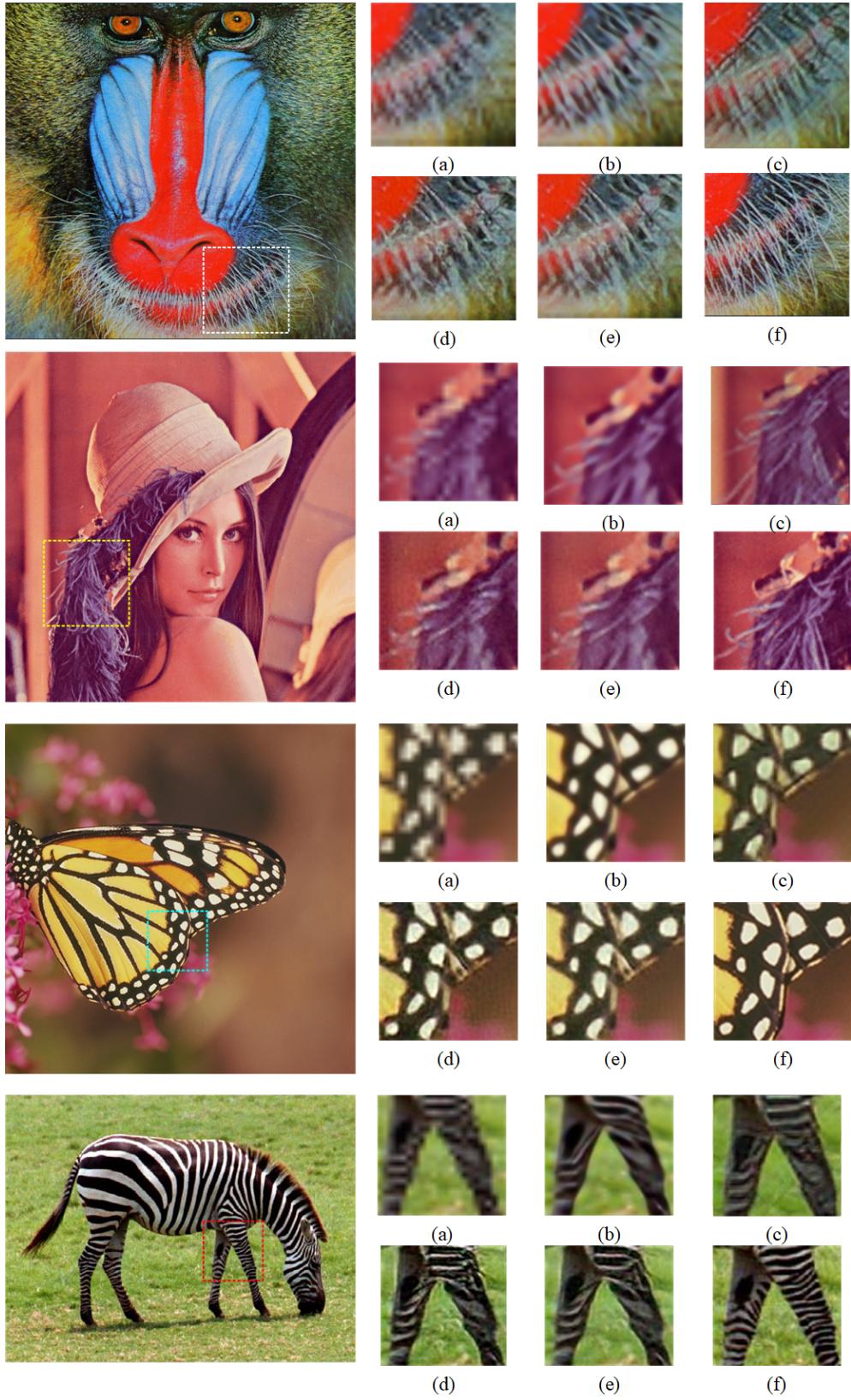


Figure 4. Subjective visual quality comparison of the proposed method and the baselines: (a)–(e) the reconstructed HR faces using (a) bicubic interpolation, (b) ESRGAN [26], (c) SAM-SR, (d) VAM-SR, (e) VAM+SAM SR, and (f) the ground-truths

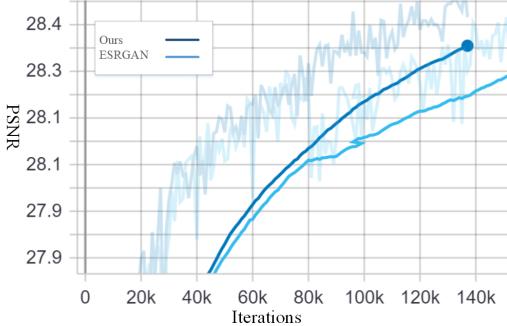


Figure 5. The PSNR curve of the Set5 [13] measured by the ground truth and the super-resolved image during training phase.

performance.

4. Conclusions

In this paper, we have proposed a novel single image super-resolution model based on the densely connected residual (DCR) blocks and dual reconstruction module. The proposed DCR architecture effectively accelerate the learning process because one more shortcut connections between convolution layers and DCR blocks allow more effective back-propagation of the gradient information. Furthermore, our dual reconstruction module fuses counterpart super-resolved images directly from our base network, yielding a better synthesized high-resolution image both quantitatively and qualitatively. Experimental results have demonstrated the effectiveness and efficiency of the proposed super-resolution model.

Acknowledgment

This study was supported in part by the Ministry of Science and Technology, Taiwan, under Grants MOST 108-2634-F-007-009, 107-2218-E-020-002-MY3, and 108-2218-E-006 -052.

References

- [1] Eirikur Agustsson and Radu Timofte. Ntire 2017 challenge on single image super-resolution: Dataset and study. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017. 5
- [2] Mario Bertero and Patrizia Boccacci. *Introduction to Inverse Problems in Imaging*. CRC Press, 1998. 2
- [3] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Image super-resolution using deep convolutional networks. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 38(2):295–307, 2015. 2
- [4] Claude E Duchon. Lanczos filtering in one and two dimensions. *Journal of Applied Meteorology*, 18(8):1016–1022, 1979. 2
- [5] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. Deep residual learning for image recognition. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 770–778, 2016. 2
- [6] Andrew Howard, Mark Sandler, Grace Chu, Liang-Chieh Chen, Bo Chen, Mingxing Tan, Weijun Wang, Yukun Zhu, Ruoming Pang, Vijay Vasudevan, et al. Searching for mobilenetv3. *arXiv preprint arXiv:1905.02244*, 2019. 4
- [7] Gao Huang, Zhuang Liu, Kilian Q Weinberger, and Laurens van der Maaten. Densely connected convolutional networks. *arXiv preprint arXiv:1608.06993*, 2016. 2, 3
- [8] Jia-Bin Huang, Abhishek Singh, and Narendra Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 5197–5206, 2015. 2
- [9] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 2
- [10] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1646–1654, 2016. 3
- [11] Kwang In Kim and Younghlee Kwon. Single-image super-resolution using sparse regression and natural image prior. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 32(6):1127–1133, 2010. 2
- [12] Diederik P Kingma and Jimmy Ba. ADAM: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014. 4
- [13] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *IEEE Conference on Computer Vision and Pattern Recognition*, 2017. 5, 7
- [14] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, 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*, pages 4681–4690, 2017. 2
- [15] Jonghwa Lee and Chulhee Lee. Fast and efficient panchromatic sharpening. *IEEE Transactions on Geoscience and Remote Sensing*, 48(1):155–163, 2009. 1
- [16] C.-H. Lin and J. M. Bioucas-Dias. An explicit and scene-adapted definition of convex self-similarity prior with application to unsupervised Sentinel-2 super-resolution. *IEEE Transactions on Geoscience and Remote Sensing*, 2019. 2
- [17] Chia-Hsiang Lin, F Ma, Chong-Yung Chi, and Chih-Hsiang Hsieh. A convex optimization-based coupled nonnegative matrix factorization algorithm for hyperspectral and multi-spectral data fusion. *IEEE Trans. Geosci. Remote Sens.*, 56(3):1652–1667, 2018. 1, 3
- [18] Jonathan Long, Evan Shelhamer, and Trevor Darrell. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 3431–3440, 2015. 4

- [19] Ilya Loshchilov and Frank Hutter. SGDR: Stochastic gradient descent with warm restarts. *arXiv preprint arXiv:1608.03983*, 2016. [4](#)
- [20] Andreas Lugmayr, Martin Danelljan, and Radu Timofte. Unsupervised learning for real-world super-resolution. *ICCV Workshops*, 2019. [2](#), [4](#), [5](#)
- [21] Andreas Lugmayr, Martin Danelljan, Radu Timofte, et al. Aim 2019 challenge on real-world image super-resolution: Methods and results. *ICCV Workshops*, 2019. [2](#), [4](#), [5](#)
- [22] Michael K Ng, Huanfeng Shen, Edmund Y Lam, and Liangpei Zhang. A total variation regularization based super-resolution reconstruction algorithm for digital video. *EURASIP Journal on Advances in Signal Processing*, 2007(1):074585, 2007. [2](#)
- [23] Wenzhe Shi, Jose Caballero, Ferenc Huszár, Johannes Totz, Andrew P Aitken, Rob Bishop, Daniel Rueckert, and Zehan Wang. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 1874–1883, 2016. [3](#), [4](#)
- [24] Karen Simonyan and Andrew Zisserman. Very deep convolutional networks for large-scale image recognition. *arXiv preprint arXiv:1409.1556*, 2014. [3](#), [4](#)
- [25] C.-H. Wang, C.-H. Lin, J. M. Bioucas-Dias, W.-C. Zheng, and K.-H. Tseng. Panchromatic sharpening of multispectral satellite imagery via an explicitly defined convex self-similarity regularization. In *Proceedings of the IEEE International Geoscience and Remote Sensing Symposium*, pages 1–4, 2019. [1](#)
- [26] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks. In *The European Conference on Computer Vision Workshops (ECCVW)*, September 2018. [1](#), [2](#), [3](#), [4](#), [5](#), [6](#)
- [27] Yulun Zhang, Kunpeng Li, Kai Li, Lichen Wang, Bineng Zhong, and Yun Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision*, pages 286–301, 2018. [2](#)
- [28] Yulun Zhang, Yapeng Tian, Yu Kong, Bineng Zhong, and Yun Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 2472–2481, 2018. [2](#)