Recent Advances in Deep Learning for Single Image Super-Resolution

Yungang Zhang(✉) and Yu Xiang

Department of Computer Science, Yunnan Normal University, Kunming 650500, China
{yungang.zhang,yu.xiang}@ynnu.edu.cn

Abstract. Image super-resolution is an important research field in image analysis. The techniques of image super-resolution has been widely used in many computer vision applications. In recent years, the success of deep learning methods in image super-resolution have attracted more and more researchers. This paper gives a brief review of recent deep learning based methods for single image super-resolution (SISR), in terms of network type, network structure, and training methods. The advantages and disadvantages of these methods are analyzed as well.

Keywords: Single image super-resolution · Deep learning
Convolutional neural networks

1 Introduction

Nowadays, the high-definition digital displays are widely used, however the super-resolution techniques are still necessary in many computer vision applications. In many areas such as video surveillance, medical imaging and remote sensing, due to many physical effects, it is common that the quality of images or videos are away from our expectation. Therefore, the users have to enhance resolutions for their original imaging results. Super-resolution (SR) techniques aim to generate a high resolution image from the original obtained low-resolution (LR) image.

The pioneering work of using deep learning in super-resolution can be found in [1], where an end-to-end mapping between the low and high-resolution images is learned, and the mapping is represented as a deep convolutional neural network. Since then, the deep learning based methods have become the mainstream in image and video super-resolution areas. The existing published results show that deep learning is a promising direction in image/video super-resolution.

In this article, a brief review of deep learning based methods in single image super-resolution is given. The rest of the article is organized as follows. Section 2 gives some background concepts in super-resolution. The literature of deep learning techniques for single image super-resolution is reviewed in Sect. 3. Section 4 concludes the article.

Supported by the National Natural Science Foundation of China (No. 61462097) and Yunnan Provincial Education Department Research Grant (No. 2018JS143).

J. Ren et al. (Eds.): BICS 2018, LNAI 10989, pp. 85–95, 2018.
https://doi.org/10.1007/978-3-030-00563-4_9

2 Background

Image Super-resolution (SR) techniques try to construct a high resolution (HR) image from one or more observed low resolution (LR) images [2]. As SR is an ill-posed problem, there may be many solutions exist. From the perspective of input LR images, the SR techniques can be divided into two main categories, namely single image super-resolution (SISR) and multiple image or multi-frame super-resolution. The SISR only uses one input LR image to produce the corresponding HR image, therefore SISR has attracted more focus from researchers as it is more close to the real scenarios in people's daily life. There are two types of SISR techniques:

1. **The learning based methods.** The techniques from machine learning are used in this type of SISR methods to estimate the HR image. The typical methods in this category are: the pixel-based methods [3] and the example-based methods [4]. The techniques such as sparse coding and neighbor embedding are also widely used [5].
2. **The reconstruction based methods.** Which needs prior knowledge to define constraints for the target HR image. The typical techniques used in this category are edge sharpening [6], regularization [7] or deconvolution [8].

In order to alleviate the ill-posed problem of SISR, recent methods try to learn prior information from LR and HR pairs, typical methods are neighbor embedding regression [9], random forest [10] and deep convolutional neural networks [1].

3 Deep Learning for Single Image Super-Resolution

In recent years, deep learning [14], especially convolutional neural networks (CNNs) [11] have become an important tool in computer vision applications. Although CNNs are not perfect [39], the excellent performances of CNNs are reported in various computer vision applications [12,13]. In this section, the recent CNNs-based and relative methods for SISR are briefly reviewed.

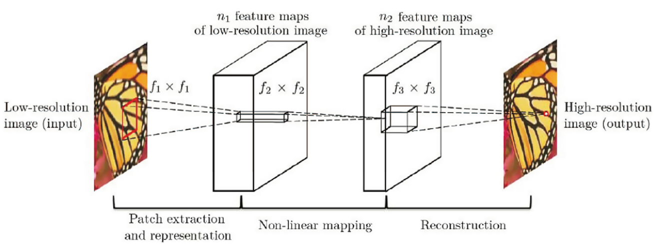

Fig. 1. SRCNN model illustration [15].

3.1 Convolutional Neural Networks for SISR

The first CNNs-based model for SISR is called SRCNN [15], which is illustrated in Fig. 1. The method tries to learn an end-to-end mapping between the input LR image and the corresponding HR image. The bicubic interpolation is used as its pre-processing step, then the image patches are extracted by convolution as feature vectors, which are then non-linearly mapped to find the most appropriate patches to reconstruct the HR image. SRCNN has only convolutional

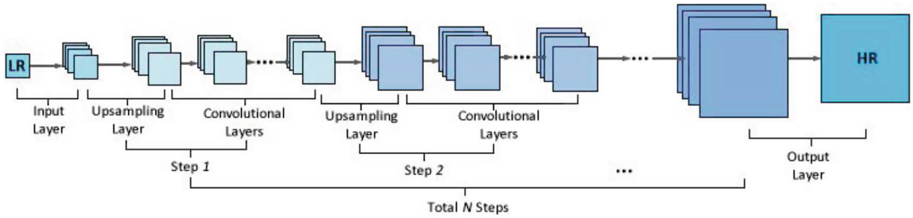

Fig. 2. The GUN model [33].

layers, which has the advantage that the input images can be of any size and the algorithm is not patch-based [16]. SRCNN outperforms many “traditional” models. Although SRCNN is efficient due to a lightweight structure is used, a fast SRCNN (FSRCNN) [17] is proposed to further improve the performance of SRCNN. The FSRCNN replaces pre-processing bicubic interpolation step of SRCNN by a post-processing step in the form of deconvolution. A 40 times improvement on time cost compared to SRCNN was reported by the authors. A deep network cascade (DNC) is introduced in [18], the model upscales an

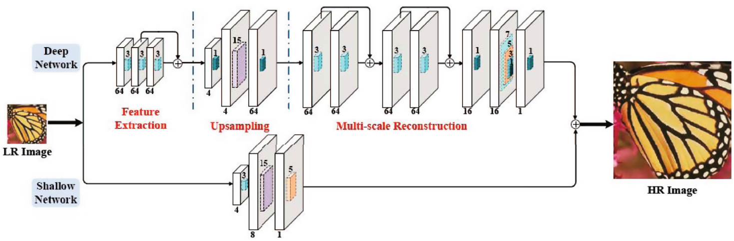

Fig. 3. Network architecture of EEDS [34].

input LR to the HR image layer by layer. The non-local self-similarities are then explored to refine the high-frequency details of the patches. The Gradual Upsampling Network (GUN [33]) gradually magnifies LR to HR. The model (Fig. 2) consists of an input convolution layer, a set of upsampling and convolutional layers, and an output layer. It is believed that the gradual upsampling strategy of adopting very small magnification factor is cost effective in terms of efficiency.

In [34], the authors replace the bicubic upsampling in the first step with feature extraction. Therefore, the LR image can be mapped into a deep feature space. Then a learning based upsampling of the deep features to the desired dimensions is carried out. For the HR reconstruction, context information is derived from the upsampled features, in a multi-scale way that incorporates both short- and long-range contextual information at the same time (Fig. 3).

Inspired by the success of VGG-net in image classification, the VDSR [19] model, as illustrated in Fig. 4, uses a very deep convolution network for SISR [20]. The success of VDSR shows that a deeper network may bring more accurate

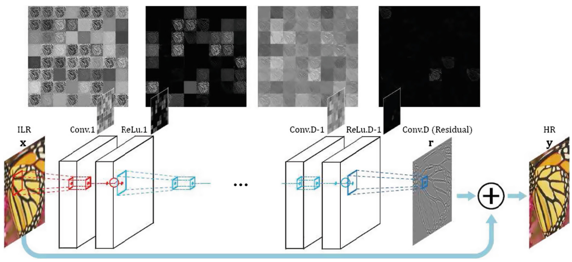

Fig. 4. The VDSR model [19].

outputs. However, deeper networks may introduce two side effects: overfitting and heavy models. Therefore, Kim et al. [21] propose a deeply-recursive convolutional network (DRCN) to apply the same convolutional layer recursively. The core idea of their model is to simultaneously have a very deep network while suppress the number of model parameters. The efficient sub-pixel convolutional neural network (ESPCN) is proposed in [32] to reduce the space and time complexities in SISR tasks. The model first downsamples the HR images to LR images, then the feature maps are extracted from LR images. Except

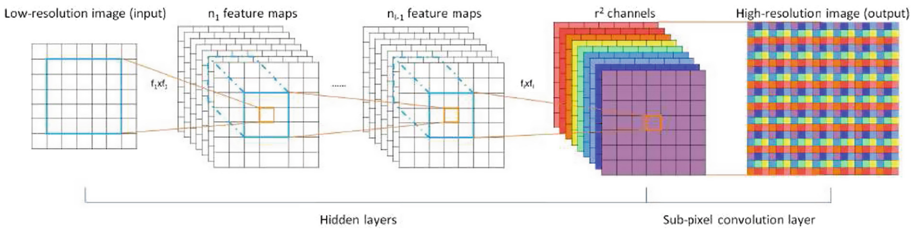

Fig. 5. The ESPCN model [32].

the last sub-pixel convolution layer, all other layers are characterized by its own upscaling filter for the feature map of the concerned layer. The “sub-pixel convolution layer” then upscales the low resolution image to a super-resolved image. The model is illustrated in Fig. 5. An enhanced deep super-resolution network

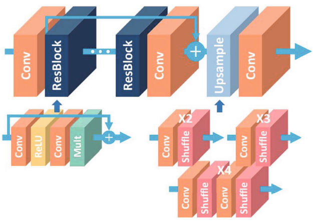

Fig. 6. The EDSR model [40].

(EDSR) is introduced in [40], as illustrated in Fig. 6. The batch normalization layers are removed from the conventional residual networks, and the performance is further improved by expanding the model size. The EDSR achieved the leading performance in CVPR Ntire 2017 challenge.

3.2 Adaptive Models

Besides to use CNN models originally from image classification tasks. Many researchers also try to build models to be more adaptive to the image contents (pixels or structures) for SISR. A Deep Projection CNN (DPN) method is introduced in [25]. DPN uses model adaptation to explore the repetitive structures in LR images. In [26], the pixel recursive super-resolution network is proposed, which comprises a conditioning network and a prior network. The conditioning network transform the LR image to logits to predict the log-likelihood of each HR pixel, and the prior network is a PixelCNN [41]. With this type of structure, the model is able to synthesize realistic details into images while enhancing their resolution. In [27], the authors propose a model named deep joint super resolution (DJSR) in order to adapt deep model for joint similarities.

More recently, Zhang et al. propose a novel adaptive residual network (ARN) [35] for high-quality image restoration. The ARN is a deep residual network, which consists of six cascaded adaptive shortcuts, convolutional layers and PReLUs. Each adaptive shortcut contains two small convolutional layers, followed by PReLU activation layers and one adaptive skip connection. The ARN model can be trained adaptively according to different applications.

3.3 Generative Adversarial Networks Based Models

Generative adversarial networks (GANs) are known for the ability to preserve texture details in images, create solutions that are close to the real image manifold and look perceptually convincing. Therefore, GANs are also can be used for SISR. The first successful application of GAN in SISR can be seen in [43], where the authors propose an adversarial loss and a content loss, a discriminator network is used to differentiate photo-realistic images from SR images created by the generator network.

In [29], the authors propose a Depixelated Super Resolution Convolutional Neural Network (DSRCNN). The model is designed for super-resolving partially pixelated images for super-resolution. It consists of an autoencoder combined with two depixelate layers through deconvolution. The autoencoder is composed of a generator and a discriminator.

GAN is applied in [30] as well, where a GAN-based architecture using densely connected convolutional neural networks (DenseNets) is proposed to super-resolve overhead imagery with a factor of up to 8×.

3.4 Sparsity-Based Models

Some researchers have revealed that combining sparse coding with CNNs can produce better performances than using CNNS only [22,31,37]. A sparse coding based network (SCN) is proposed in [23], with the sparse priors, the model is more compact and accurate compared to SRCNN. Another model called SCRNN-Pr [24] also tries to explore image priors during the training of a deep CNN. Better training time cost and super-resolution performances are reported compared to other state-of-the-art methods.

In [51], a hybrid wavelet convolution network (HWCN) (Fig. 7) is proposed. The LR image is fed into a scattering convolution network (a wavelet tree in nature) to obtain scattering feature maps, sparse codes are then extracted from those maps and used as inputs for a CNN. The model is able to use a tiny dataset to train complex deep networks with better generalization. The authors in

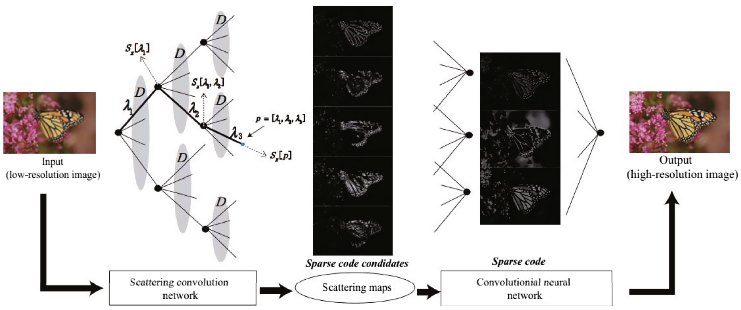

Fig. 7. The HWCN model [51].

[28] propose a very deep Residual Encoder-Decoder Network: RED-Net, which consists of a series of convolutional and deconvolutional layers aim to obtain end-to-end mappings from LR images to HR images. The use of stacked autoencoder

for image restoration can be found in [49], where a non-local stacked autoencoder is proposed. The stacked autoencoder (SAE) is used in [47] and [48] for saliency detection, nevertheless the models can be extended for SISR purpose as well.

3.5 Other Types of Methods

The combining of multiple super-resolution models and ensemble of multiple CNNs are also reported can obtain competitive performance in SISR.

Model Combination. In [42], the authors combines some contemporary state-of-the-art super-resolution methods using conditioned regression models. Their proposed "Regression conditioned" SRCNN has the idea to constructed a single training model to avoid model re-training for different images. The model can be effective for different blur kernels.

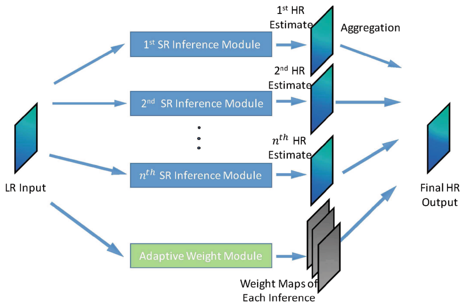

Fig. 8. MSCN [38].

Ensemble of CNNs. The use of multiple CNNs is also a promising direction in this area. In [38], the authors apply a number of SR methods to the LR image independently to get various HR estimates, which are then combined on the basis of adaptive weights to produce the final result. The method is named as MSCN-n, where n is the number of employed inference modules (Fig. 8). Similarly, another multiple CNNs model is proposed in [50], each individual CNN is trained separately with different network structure. A Context-wise Network Fusion (CNF) approach is proposed to integrate the outputs of individual networks by additional convolution layers.

A wavelets guided multiple CNNs method is proposed in [36]. The wavelet decomposition of images are used for multi-scale representations. Then multiple CNNs are trained for approximating the wavelet multi-scale representations, separately. For inference, the trained CNNs regress wavelet multi-scale representations from a low-resolution image, followed by wavelet synthesis that forms a restored high-resolution image. The illustration of the model can be seen in Fig. 9.

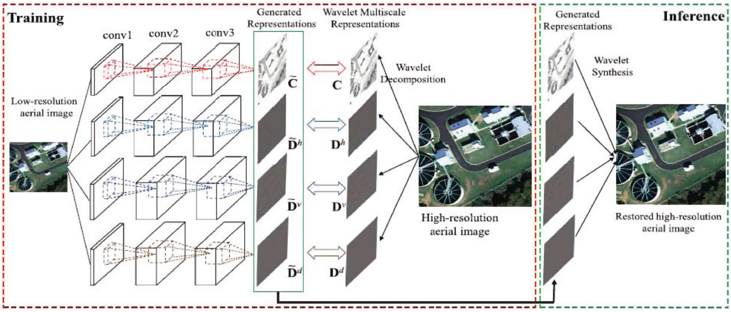

Fig. 9. The wavelet multiscale CNNs model [36].

There are also other types of deep learning methods for SISR, such as deep belief networks [44], the Laplacian Pyramid Super-Resolution Network (LapSRN [45]), and recurrent neural networks (RNNs) [46]. All obtain state-of-the-art performances.

4 Conclusion

In this article, the recent deep learning based single image super-resolution methods are briefly reviewed. The convolution neural networks and relative deep learning models have achieved better performance in SISR than other conventional methods. However, how do we recover the finer texture details when we super-resolve at large upscaling factors is still a question need to be answered. In addition, the models which can be applied in real time scenarios also need further investigations. In our future work, we will pay particular attention to the sparsity-based deep learning models, as sparsity-based models are able to reduce the complexity of the learning processes and extract local information that leads to better performance [48]. Moreover, we believe that the combination of GANs and sparse models is a promising direction in SISR.

References

1. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8692, pp. 184–199. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10593-2_13
2. Yue, L., et al.: Image super-resolution: the techniques, applications, and future. Sig. Process. **128**, 389–408 (2016)
3. Zhang, K., Gao, X., Tao, D., Li, X.: Single image super-resolution with non-local means and steering kernel regression. IEEE Trans. Image Process. **21**, 4544–4556 (2012)
4. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. IEEE Comput. Graph. Appl. **22**, 56–65 (2002)

5. Gao, X., Zhang, K., Tao, D., Li, X.: Image super-resolution with sparse neighbor embedding. IEEE Trans. Image Process. **21**, 3194–3205 (2012)
6. Dai, S., et al.: Soft edge smoothness prior for alpha channel super resolution. In: Proceedings of CVPR 2007, pp. 1–8. IEEE (2007)
7. Aly, H.A., Dubois, E.: Image up-sampling using total-variation regularization with a new observation model. IEEE Trans. Image Process. **14**, 1647–1659 (2005)
8. Shan, Q., et al.: Fast image/video upsampling. ACM Trans. Graph. **27**, 153:1–153:7 (2008)
9. Timofte, R., De Smet, V., Van Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: Proceedings of ICCV 2013, pp. 1920–1927. IEEE (2013)
10. Schulter, S., Leistner, C., Bischof, H.: Fast and accurate image upscaling with super-resolution forests. In: Proceedings of CVPR 2015, pp. 3791–3799. IEEE (2015)
11. Krizhevsky, A., Sutskever, I., Hinton, G.E.: ImageNet classification with deep convolutional neural networks. In: Proceedings of NIPS 2012, vol. 1, pp. 1097–1105. Curran Associates Inc. (2012)
12. Zeiler, M.D., Fergus, R.: Visualizing and understanding convolutional networks. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8689, pp. 818–833. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10590-1_53
13. Szegedy, C., et al.: Going deeper with convolutions. In: Proceedings of CVPR 2015, pp. 1–9. IEEE (2015)
14. LeCun, Y., Bengio, Y., Hinton, G.: Deep learning. Nature **521**(7553), 436–444 (2015)
15. Dong, C., et al.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**, 295–307 (2016)
16. Kappeler, A., et al.: Video super-resolution with convolutional neural networks. IEEE Trans. Comput. Imaging **2**, 109–122 (2016)
17. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Leibe, B., Matas, J., Sebe, N., Welling, M. (eds.) ECCV 2016. LNCS, vol. 9906, pp. 391–407. Springer, Cham (2016). https://doi.org/10.1007/978-3-319-46475-6_25
18. Cui, Z., Chang, H., Shan, S., Zhong, B., Chen, X.: Deep network cascade for image super-resolution. In: Fleet, D., Pajdla, T., Schiele, B., Tuytelaars, T. (eds.) ECCV 2014. LNCS, vol. 8693, pp. 49–64. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-10602-1_4
19. J. Kim, J. K. Lee, K. M. Lee: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of CVPR 2016, pp. 1646–1654. IEEE (2016)
20. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. CoRR, vol abs/1409.1556. http://arxiv.org/abs/1409.1556. (2014)
21. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of CVPR 2016, pp. 1637–1645. IEEE (2016)
22. Gregor, K., Lecun, Y.: Learning fast approximations of sparse coding. In: Proceedings of ICML 2010, pp. 399–406. Omnipress (2010)
23. Liu, D., et al.: Robust single image super-resolution via deep networks with sparse prior. IEEE Trans. Image Process. **25**, 3194–3207 (2016)
24. Liang, Y., et al.: Incorporating image priors with deep convolutional neural networks for image super-resolution. Neurocomputing **194**, 340–347 (2016)

25. Liang, Y., et al.: Single image super resolution - when model adaptation matters. CoRR, vol abs/1703.10889. https://arxiv.org/abs/1703.10889. (2017)
26. Dahl, R., Norouzi, M., Shlens, J.: Pixel recursive super resolution. CoRR, vol abs/1702.00783. http://arxiv.org/abs/1702.00783. (2017)
27. Wang, Z., et al.: Self-tuned deep super resolution. In: Proceedings of CVPR Workshop 2015, pp. 1–8. IEEE (2015)
28. Mao, X., Shen, C., Yang, Y.: Image denoising using very deep fully convolutional encoder-decoder networks with symmetric skip connections. CoRR, vol abs/1603.09056. http://arxiv.org/abs/1603.09056. (2016)
29. Mao, H., et al.: Super resolution of the partial pixelated images with deep convolutional neural network. In: Proceedings of 2016 ACM on Multimedia Conference, pp. 322–326. ACM (2016)
30. Bosch, M., et al.: Super-resolution for overhead imagery using denseNets and adversarial learning. CoRR, vol abs/1711.10312v1. http://arxiv.org/abs/1711.10312v1. (2017)
31. Shi, Y., et al.: Local- and holistic- structure preserving image super resolution via deep joint component learning. CoRR, vol abs/1607.07220. http://arxiv.org/abs/1607.07220. (2016)
32. Shi, W., et al.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of CVPR 2016, pp. 1874–1883. IEEE (2016)
33. Zhao, Y., et al.: Gun: gradual upsampling network for single image superresolution. CoRR, vol abs/1703.0424. http://arxiv.org/abs/1703.0424. (2017)
34. Wang, Y., et al.: End-to-end image super-resolution via deep and shallow convolutional networks. CoRR, vol abs/1607.07680. http://arxiv.org/abs/1607.07680. (2016)
35. Zhang, Y., et al.: Adaptive residual networks for high-quality image restoration. IEEE Trans. Image Process. **27**, 3150–3163 (2018)
36. Wang, T., et al.: Aerial image super resolution via wavelet multiscale convolutional neural networks. IEEE Geosci. Remote. Sens. Lett. **15**, 769–773 (2018)
37. Osendorfer, C., Soyer, H., van der Smagt, P.: Image super-resolution with fast approximate convolutional sparse coding. In: Loo, C.K., Yap, K.S., Wong, K.W., Beng Jin, A.T., Huang, K. (eds.) ICONIP 2014. LNCS, vol. 8836, pp. 250–257. Springer, Cham (2014). https://doi.org/10.1007/978-3-319-12643-2_31
38. Liu, D., et al.: Learning a mixture of deep networks for single image super-resolution. CoRR, vol abs/1701.00823. https://arxiv.org/abs/1701.00823. (2017)
39. Shalev-Shwartz, S., Shamir, O., Shammah, S.: Failures of deep learning. CoRR, vol abs/1703.07950v1. https://arxiv.org/pdf/1703.07950v1.pdf. (2017)
40. Lim, B., et al.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of CVPR Workshop 2017, pp. 1132–1140. IEEE (2017)
41. van den Oord, A., et al.: Conditional image generation with PixelCNN decoders. CoRR, vol abs/1606.05328. http://arxiv.org/abs/1606.05328. (2016)
42. Riegler, G., et al.: Conditioned regression models for non-blind single image super-resolution. In: Proceedings of ICCV 2015, pp. 522–530. IEEE (2015)
43. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of CVPR 2017, pp. 105–114. IEEE (2017)
44. Brosch, T., Tam, R.: Efficient training of convolutional deep belief networks in the frequency domain for application to highresolution 2D and 3D images. Neural Comput. **27**, 211–227 (2015)
45. Lai, W., et al.: Deep Laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of CVPR 2017, pp. 5835–5843. IEEE (2017)

46. Han, W., et al.: Image super-resolution via dual-state recurrent networks. CoRR, vol abs/1805.02704. https://arxiv.org/pdf/1805.02704.pdf (2018)
47. Han, J., et al.: Background prior-based salient object detection via deep reconstruction residual. IEEE Trans. Circuits Syst. Video Technol. **25**, 1309–1321 (2015)
48. Zabalza, J., et al.: Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. Neurocomputing **185**, 1–10 (2016)
49. Wang, R., Tao, D.: Non-local auto-encoder with collaborative stabilization for image restoration. IEEE Trans. Image Process. **25**, 2117–2229 (2016)
50. Ren, H., Elkhamy, M., Lee, J.: Image super resolution based on fusing multiple convolution neural networks. In: Proceedings of CVPR 2017, pp. 1050–1057. IEEE (2017)
51. Gao, X., Xiong, H.: A hybrid wavelet convolution network with sparse-coding for image super-resolution. In: Proceedings of ICIP 2016, pp. 1439–1443. IEEE (2016)