Super-Resolution Reconstruction Using Deep Learning: Should We Go Deeper?

Daniel Kostrzewa[1,2(✉)] , Szymon Piechaczek[1], Krzysztof Hrynczenko[1],
Paweł Benecki[1,2] , Jakub Nalepa[1,2] , and Michal Kawulok[1,2]

[1] Future Processing, Bojkowska 37A, 44-100 Gliwice, Poland
{dkostrzewa,pbenecki,jnalepa,mkawulok}@future-processing.com
[2] Institute of Informatics, Silesian University of Technology,
Akademicka 16, 44-100 Gliwice, Poland
{daniel.kostrzewa,pawel.benecki,jakub.nalepa,michal.kawulok}@polsl.pl

Abstract. Super-resolution reconstruction (SRR) is aimed at increasing image spatial resolution from multiple images presenting the same scene or from a single image based on the learned relation between low and high resolution. Emergence of deep learning allowed for improving single-image SRR significantly in the last few years, and a variety of deep convolutional neural networks of different depth and complexity were proposed for this purpose. However, although there are usually some comparisons reported in the papers introducing new deep models for SRR, such experimental studies are somehow limited. First, the networks are often trained using different training data, and/or prepared in a different way. Second, the validation is performed for artificially-degraded images, which does not correspond to the real-world conditions. In this paper, we report the results of our extensive experimental study to compare several state-of-the-art SRR techniques which exploit deep neural networks. We train all the networks using the same training setup and validate them using several datasets of different nature, including real-life scenarios. This allows us to draw interesting conclusions that may be helpful for selecting the most appropriate deep architecture for a given SRR scenario, as well as for creating new SRR solutions.

Keywords: Super-resolution reconstruction · Image processing · Convolutional neural network · Deep learning

1 Introduction

Single-image super-resolution reconstruction (SRR) is a group of methods, whose main goal is to construct a high-resolution (HR) image on the basis of a single low-resolution (LR) input image [30,40]. There are a few main approaches to resolve this task: usage of predefined mathematical formula (e.g., bilinear, bicubic, and other interpolation methods), edge based methods [12,35,38], using a dictionary with pairs of matched LR and HR image fragments [6,13,14,19,23,41], heavy-tailed gradient distribution [32], sparsity property of large gradients [23], discrete and stationary wavelet decomposition [7,8],

and universal hidden Markov tree model [28]. However, most of recent works engage convolutional neural networks (CNNs) to transform LR image to HR output.

1.1 Related Work

Super-resolution CNN (SRCNN) is the first deep learning model whose authors claimed to surpass performance of other methods [9,10]. SRCNN is a single-image super-resolution reconstruction algorithm that optimizes an end-to-end mapping from LR to HR image. This model has been improved in two ways: by increasing depth of the network, and by introducing recursive learning. In very deep super-resolution CNN (VDSR) [21] and image restoration CNN (IRCNN) [42], additional convolutional layers have been added. Kim et al. proposed recursive learning for parameter sharing in their deeply-recursive convolutional networks (DRCNs) [22]. Afterwards, Tai et al. suggested the use of recursive blocks (deep recursive residual network—DRRN [36]) and memory blocks (Memnet [37]). All mentioned networks have the same drawback—they require input images that are of the same size as the desired output image. Therefore, an additional step is needed to interpolate the LR image to the expected size.

Fast super-resolution CNN (FSRCNN) [11], the successor to the SRCNN, solves aforementioned problem by taking the original LR image as an input, and using deconvolution layer to enlarge the image to the desired resolution. Moreover, the authors improved the previous design to accelerate the processing (up to 24 frames per second). In efficient sub-pixel CNN (ESPCN) [33], a new layer type was introduced to upscale the final LR feature maps into the reconstructed output. On this basis, SRResNet [27] and enhanced deep residual network (EDSR) were proposed [29], which exploit residual learning. In recent years, generative adversarial networks (GANs) gain popularity in image generation, and they outperform are deep learning techniques in SRR [27].

1.2 Contribution

We have witnessed a breakthrough in single-image SRR, underpinned with the use of deep neural networks. Architectures of different depth and complexity are employed for this purpose, and the widely shared opinion is that the deeper models are more capable of learning the relation between low and high resolution. However, the experimental results reported in the papers introducing new SRR techniques are often limited—the test sets are commonly composed of the images of the same kind, and LR images are obtained by downscaling and degrading HR images, which serve as a reference (grount-truth, GT) for evaluation.

In this paper, we report the results of our extensive experimental study to compare a number of different deep architectures that have been proposed for SRR. We consider the magnification factor of $2\times$, as we want to compare the networks taking into account their capability of reconstructing images based on the LR information. For larger magnification factors, the networks are, in fact, trying to "guess" the high-resolution appearance of the details that are

not visible in the input image. Our main contribution lies in comparing the state-of-the-art deep SRR networks using the same training data and setup, for several test sets of different kind, namely: (i) artificially-degraded natural images, (ii) artificially-degraded satellite images, and (iii) real satellite data that are matched with HR images acquired using a different satellite of higher spatial resolution.

1.3   Paper Structure

The paper is structured as follows. Section 2 describes in details the CNNs investigated in this paper. In Sect. 3, we present the results of our experiments, which have been carried out to evaluate the implemented networks. Finally, Sect. 4 concludes the paper and shows the main goals of our ongoing research.

2   Convolutional Neural Networks for SRR

In this subsection, we discuss the implemented CNNs in detail. These networks have been experimentally validated in Sect. 3.

2.1   Super-Resolution Convolutional Neural Network

SRCNN is the first CNN that has been developed specifically for SRR (Fig. 1) [9,10]. The process of mapping LR to HR image can be separated into three steps:

1. **Patch extraction and representation**
   This part is done by the first convolutional layer. The *patch extraction* is the process of sliding kernel through the whole image with the overlaps. *Representation* is the resulting feature map which is a consequence of the aforementioned *patch extraction*.
2. **Non-linear mapping**
   The second convolving part non-linearly maps obtained feature map onto another one. This new set of features directly corresponds to the HR image.
3. **Reconstruction**
   The last part of the process gathers those HR representations and fuse them into an image which should be as similar as possible to ground-truth.

The SRCNN model is parametrized by a fairly small set of hyperparameters: the number of color channels, filter sizes in the first, second, and third layers (in this case, $k = 9$, $k = 1$, and $k = 5$, respectively), and the number of filters in all layers ($n = 64$, $n = 32$, and $n = 1$, respectively; note that the number of filters in the last layer corresponds to the number of color channels).

SRCNN requires a pre-magnification of the input image to the desired size. This design comes with one great advantage compared to other architectures, where the whole super-resolution process is done by the model itself. The advantage is that the network can be trained to restore images upscaled by any factor, including non-integer upscaling factors.
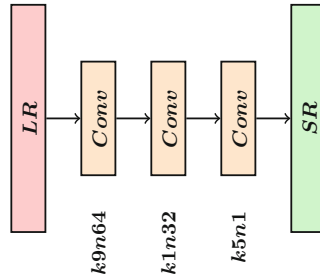
Fig. 1. A diagram of the SRCNN model

The experimental results reported in [9,10] show that SRCNN achieves better numerical results than the state of the art at the time. It can be further tuned in terms of either time efficiency or performance by changing the previously mentioned hyperparameters.

2.2   Fast Super-Resolution Convolutional Neural Network

FSRCNN is built on the basis of SRCNN. There are two main high-level differences between these networks. The first one is that FSRCNN resizing process is performed by the part of the model itself. Thus, the preprocessing step needed by the SRCNN is eliminated. SRCNN requires that for different scales, therefore the network has to be trained from scratch.

FSRCNN (Fig. 2) is composed of five different parts, which resembles the SRCNN structure:

1. **Feature extraction**
   Feature extraction is made at the very beginning of the processing pipeline (first layer). Comparing to the SRCNN counterpart, its kernel size has been reduced due to the fact that the input LR image is of the original size (it is not interpolated).
2. **Shrinking**
   To improve processing speed, the second convolutional layer is appended with a kernel of size 1×1. Its purpose is to reduce the number of channels and as a result it lessens the number of parameters.
3. **Non-linear mapping**
   The non-linear mapping is performed by several convolutional layers with smaller kernels instead of one layer with a bigger filter of a greater size.
4. **Expanding**
   The features are "expanded" to correspond directly to the high-resolution image from which the final image is produced by the last part.
5. **Deconvolving**
   The last part of the model aggregates the features and performs the upscaling to produce the final high-resolution image.
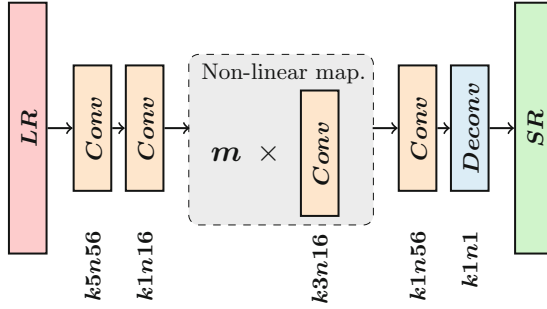
Fig. 2. A diagram of the FSRCNN model

The FSRCNN architecture improves computational speed without negatively impacting the performance. Both, the computational speed and the performance are highly adjustable with the number of non-linear mapping layers. Increasing that number results in growing complexity and mapping accuracy.

The results presented in [11] showed that the improved architecture is truly able to achieve real-time processing speed. Another effect of the changes is that the reconstruction quality also was improved. Multiple narrow layers (non-linear mapping layers) instead of one wider layer give better performance. The authors showed that including upscaling part as a deconvolution layer also positively affects the reconstruction quality. FSRCNN benefits from transfer-learning, once it is trained for one upscaling factor, only the last deconvolutional layer has to be retrained for other upscaling factor values.

2.3   Super-Resolution Residual Neural Network

Deeper networks have proven the ability to render performance impossible to obtain by the shallow networks—more layers allow for modeling mappings of a very high complexity. However, increasing the depth of the network is not as easy as adding more layers—the problems of convergence and degradation (of the network performance) can easily emerge. The vanishing/exploding gradient problems [3,15] can stop the network from converging—error gradient may bring too large or too small update for the network. Still, if the network manages to converge, its performance may deteriorate with the increase in depth [34]. The former problem has been addressed by normalized initialization [15,17,26,31] or batch normalization layers [20]; for the degradation problem, the proposed solutions encompass the highway networks [34] (utilizing information flow between layers) and residual networks [18] (with the input skipping layers).

The SRResNet [27] (Fig. 3) architecture belongs to the family of residual networks, and it was inspired by a network proposed by He et al. [18]. It deals with both vanishing/exploding gradient and degradation problems (by the use of intermediate normalization layers and residual connections). The *residual blocks* (RBs) are the groups of layers stacked together with the input of the block added
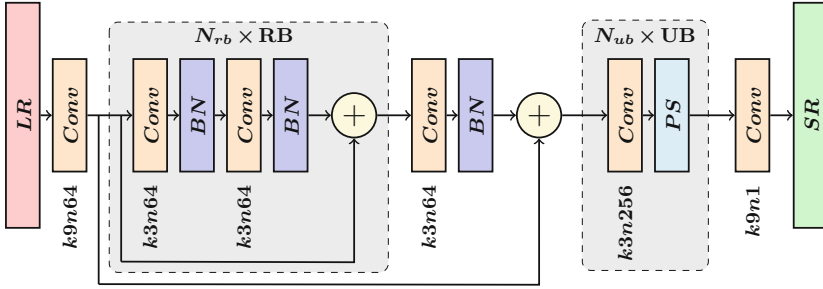
Fig. 3. A diagram of the SRResNet model

to the output of the final layer contained in this block. In SRResNet, each block encompasses two convolutional layers, each followed by a *batch normalization* (BN) layer that neutralizes the internal co-variate shift. The *upsampling blocks* (UBs) allow for image enlargement by *pixel shuffling* (PS) layers that increase the resolution of the features. The number of both RBs and UBs is variable—by increasing the number of RBs, the network may model a better mapping, whereas by changing the number of UBs, we may tune its scaling factor. However, by adding more blocks, the architecture of the network becomes increasingly complex, which makes it harder to train.

Throughout the whole network, Parametric ReLU (PReLU) is used as an activation function. Similarly to Leaky ReLU, it introduces a small slope for negative values, and PReLU enables the network to learn the optimal value for this slope. Both Leaky and Parametric versions of ReLU decrease the time needed for network to converge, and help with the dying ReLU problem [39].

2.4 Super-Resolution Generative Adversarial Network

GANs are rather complicated structures and may be described as two competing networks, the first one (generator) is trying to produce the image indistinguishable from the real (not generated) pictures, while the second (discriminator)—focuses on identifying these "fake" images. Once the discriminator starts to distinguish well between "real" and "fake" images, the generator network needs to produce increasingly better images. At the end of the training, discriminator should not be able to differentiate original images from those produced by the generator. In such a training scheme, the role of the discriminator may be perceived as an adaptive loss function, that moderates the training of the network by gradually increasing demand for the networks' performance.

Scenarios for the usage of GANs encompass image generation from noise, image inpainting and style transfer. In the context of single image super-resolution, GANs were firstly introduced by Ledig et al. [27]. In SRGAN (Fig. 4), both generator and discriminator (Fig. 5) are very deep networks; in fact SRResNet is used as a generative part of the network. To train the generator network, two loss functions are used and weighted—content loss is evaluated as an MSE
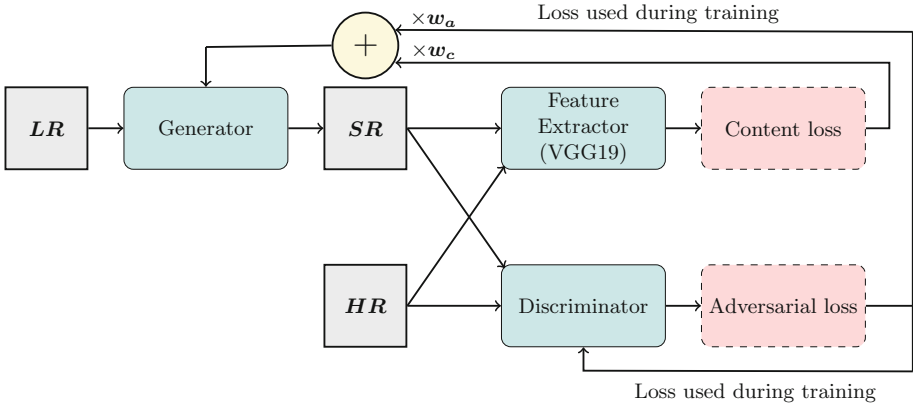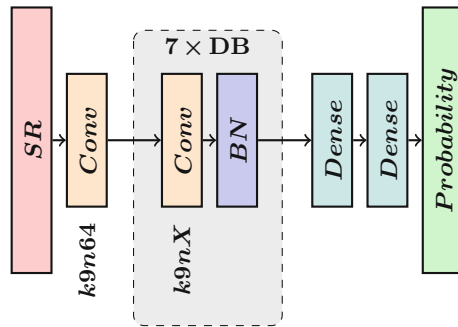
Fig. 4. A diagram of the SRGAN model

Fig. 5. A diagram of the discriminator model. Number of convolution filters (denoted by X) progressively increase, with $X \in \{64, 128, 256, 512\}$

between features of "real" and "fake" images (extracted with VGG19 network); adversarial loss favors the solutions that generate images unable to distinguish from the "real" ones. For discriminator, the loss proposed in [16] is used.

3   Experiments

All convolutional network models have been trained with DIV2K dataset [1]. The images were converted to 8-bit greyscale and downsampled $2\times$ using bicubic interpolation (in order to create LR patches). The set of prepared training and validation images is available online[1] under the license provided on the website [24].

Several popular benchmarks were selected for the experiments: **Set5** [5], **Set14** [41], and **BSD** [4]. In addition to these, we also created two datasets,

---

[1] https://doi.org/10.7910/DVN/DKSPJF.

based on B4MultiSR [25], consisting of satellite images. The first one, **Artificialy Degraded Satellite** dataset, is composed of images gathered during the Sentinel-2 mission. In this case, GT is the original image and LR is a bicubic downscaled counterpart. **Real Satellite** dataset is the most challenging scenario in our experiments—GTs are obtained by downscaling images from the Digital Globe WorldView-4 satellite (original ground sampling distance equals 30 cm/px), and LR are images from Sentinel-2.

To quantitatively compare the quality of the investigated models, we utilized several popular metrics: peak signal-to-noise ratio (PSNR), structural similarity index (SSIM), universal image quality index (UIQI), and visual information fidelity (VIF). Additionally, we exploit the following metrics: $PSNR_{HF}$ and KFS_{SIFT} [2]. In Table 1, we gathered all numeric results of conducted experiments, whereas Figs. 6, 7, and 8 render images for the visual comparison.

Table 1. The results of our experiments. We boldfaced the best results for each dataset, and the deep learning methods which render the results worse than Bicubic are annotated with the gray background.

Dataset name	Model	PSNR	SSIM	UIQI	VIF	$PSNR_{HF}$	KFS_{SIFT}
Set5	Bicubic	32.33	0.922	0.859	0.605	38.94	42.56
	SRCNN	34.96	0.944	0.871	0.672	45.40	44.67
	FSRCNN	**35.52**	**0.949**	**0.876**	**0.690**	46.70	44.73
	SRResNet	35.20	0.935	0.858	0.689	46.32	**44.97**
	SRGAN	33.86	0.931	0.828	0.637	**46.86**	44.57
Set14	Bicubic	28.75	0.857	0.781	0.500	34.34	42.46
	SRCNN	30.75	0.894	0.818	0.559	39.33	44.55
	FSRCNN	**31.11**	**0.900**	**0.825**	**0.574**	40.14	**44.69**
	SRResNet	30.29	0.889	0.813	0.558	39.21	44.62
	SRGAN	30.11	0.878	0.777	0.534	**40.49**	44.11
BSD	Bicubic	28.23	0.832	0.768	0.467	33.70	42.91
	SRCNN	29.88	0.876	0.814	0.518	38.45	45.06
	FSRCNN	**30.19**	**0.884**	**0.825**	**0.532**	**39.30**	**45.13**
	SRResNet	29.85	0.880	0.817	0.527	38.66	44.98
	SRGAN	29.88	0.876	0.814	0.518	38.45	45.06
Artificially Degraded Satellite	Bicubic	**30.22**	**0.899**	**0.873**	**0.517**	**41.97**	**43.54**
	SRCNN	26.84	0.829	0.800	0.440	34.76	41.85
	FSRCNN	26.11	0.812	0.783	0.425	33.43	41.49
	SRResNet	26.31	0.813	0.785	0.429	33.64	41.49
	SRGAN	24.60	0.727	0.696	0.363	30.43	39.92
Real Satellite	Bicubic	16.81	0.391	0.170	0.087	35.00	39.71
	SRCNN	16.64	0.345	0.138	0.075	35.68	40.55
	FSRCNN	16.72	0.343	0.139	0.076	35.81	40.34
	SRResNet	16.90	**0.394**	**0.179**	**0.090**	36.21	40.76
	SRGAN	**17.11**	0.391	0.173	0.088	**37.45**	**40.79**

For Set5 (Fig. 6), Set14 and BSD benchmarks, being most widely used datasets for SRR testing, FSRCNN renders best results both qualitatively and
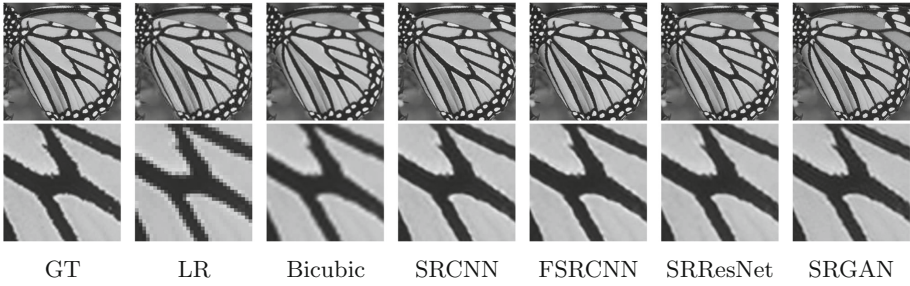
Fig. 6. Example of the reconstructed image from the Set5 dataset. A part of the image is zoomed for clarity (second row).

quantitatively. On the other hand, deeper networks (SRResNet and SRGAN), introduced the halo effect around the edges, decreasing the numerical results as well. It is also worth mentioning that these benchmarks are composed of images in the `jpeg` format (which introduces artifacts)—this may degrade the results too (in fact, deeper networks seem to magnify such artifacts).

The results for the Artificially Degraded Satellite images "favor" the outcomes produced by the bicubic interpolation. However, the images obtained by the deep networks seem to be sharper than the high-resolution version of the image (Fig. 7). This may be the reason for the observed lower quantitative scores.
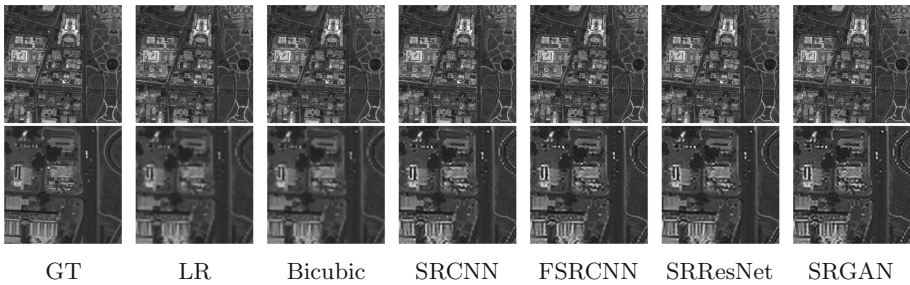

Fig. 7. Example of the reconstructed image from the Artificially Degraded Satellite dataset. A part of the image is zoomed for clarity (second row).

Finally, for the images from the Real Satellite benchmark (Fig. 8), deeper networks obtained the highest scores. Still, in this case, compared images (GT and SR versions of the image) present the same area, however they do not share similar pixel values (e.g., due to the variable lighting conditions). This is the reason why some metrics (see e.g., PSNR) have such low values.
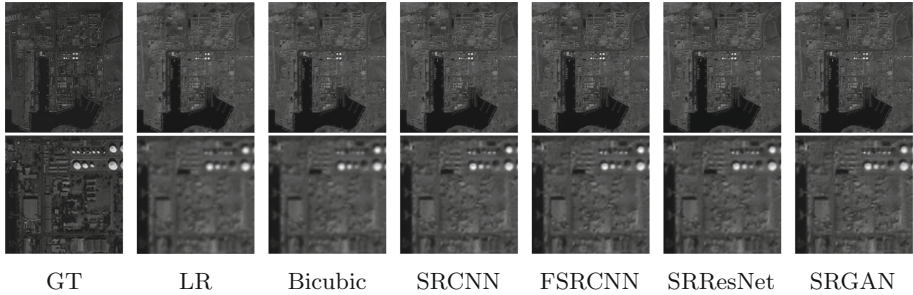
GT   LR   Bicubic   SRCNN   FSRCNN   SRResNet   SRGAN

Fig. 8. Example of the reconstructed image from the Real Satellite dataset. A part of the image is zoomed for clarity (second row).

4   Conclusions and Future Work

In this paper, we compared the performance of four different deep network architectures for single image super-resolution reconstruction. These networks were evaluated over five different benchmarks, including three standard ones (Set5, Set14, and BSD100) and two introduced in this work (Artificially Degraded Satellite and Real Satellite). The experiments showed that it is notably easier to reconstruct artificially degraded images (therefore, shallower networks can effectively cope with this task). In this case, deeper networks start to enhance the `jpeg` artifacts. On the other hand, reconstructing an image and comparing it to the one obtained by another sensor is much more difficult. As a result, deeper networks (with higher capacities) outperformed the others.

It is also worth mentioning, that deeper networks are much more complex structures, hence computational power needed to perform the training and prediction increase. In this article, we did not focus on efficiency of the deep networks, however our study shows that introducing more complexity to the model is not always worth it.

Currently, we are investigating the influence of the scale on the performance of deep networks (and how deeper networks deal with scales higher than 2). Also, we are focused on the comparison between the classical computer vision methods and deep networks for both single-image and multi-frame super-resolution.

References

1. Agustsson, E., Timofte, R.: 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
2. Benecki, P., Kawulok, M., Kostrzewa, D., Skonieczny, L.: Evaluating super-resolution reconstruction of satellite images. Acta Astronaut. **153**, 15–25 (2018)
3. Bengio, Y., Simard, P., Frasconi, P.: Learning long-term dependencies with gradient descent is difficult. IEEE Trans. Neural Netw. **5**(2), 157–166 (1994)
4. Berkeley Segmentation Dataset: 01 December 2019. https://www2.eecs.berkeley.edu/Research/Projects/CS/vision/bsds/
5. 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, pp. 135.1–135.10. BMVA Press (2012)
6. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. In: Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004, CVPR 2004, pp. 275–282. IEEE (2004)
7. Chavez-Roman, H., Ponomaryov, V.: Super resolution image generation using wavelet domain interpolation with edge extraction via a sparse representation. IEEE Geosci. Remote Sens. Lett. **11**(10), 1777–1781 (2014)
8. Demirel, H., Anbarjafari, G.: Image resolution enhancement by using discrete and stationary wavelet decomposition. IEEE Trans. Image Process. **20**(5), 1458–1460 (2011)
9. 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
10. Dong, C., Loy, C.C., He, K., Tang, X.: Image super-resolution using deep convolutional networks. IEEE Trans. Pattern Anal. Mach. Intell. **38**(2), 295–307 (2016)
11. 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
12. Fattal, R.: Image upsampling via imposed edge statistics. ACM Trans. Graph. (TOG) **26**, 95 (2007)
13. Freeman, W.T., Jones, T.R., Pasztor, E.C.: Example-based super-resolution. IEEE Comput. Graph. Appl. **22**(2), 56–65 (2002)
14. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: 2009 IEEE 12th International Conference on Computer Vision, pp. 349–356. IEEE (2009)
15. Glorot, X., Bengio, Y.: Understanding the difficulty of training deep feedforward neural networks. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics, pp. 249–256 (2010)
16. Goodfellow, I., et al.: Generative adversarial nets. In: Advances in Neural Information Processing Systems, pp. 2672–2680 (2014)
17. He, K., Zhang, X., Ren, S., Sun, J.: Delving deep into rectifiers: surpassing human-level performance on imagenet classification. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1026–1034 (2015)

18. 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, pp. 770–778 (2016)

19. Huang, J.B., Singh, A., Ahuja, N.: Single image super-resolution from transformed self-exemplars. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 5197–5206 (2015)

20. Ioffe, S., Szegedy, C.: Batch normalization: accelerating deep network training by reducing internal covariate shift. arXiv preprint arXiv:1502.03167 (2015)

21. Kim, J., Kwon Lee, J., Mu Lee, K.: Accurate image super-resolution using very deep convolutional networks. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1646–1654 (2016)

22. Kim, J., Kwon Lee, J., Mu Lee, K.: Deeply-recursive convolutional network for image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1637–1645 (2016)

23. Kim, K.I., Kwon, Y.: Single-image super-resolution using sparse regression and natural image prior. IEEE Trans. Pattern Anal. Mach. Intell. **6**, 1127–1133 (2010)

24. Kostrzewa, D., Piechaczek, S., Hrynczenko, K., Benecki, P., Nalepa, J., Kawulok, M.: Replication data for: super-resolution reconstruction using deep learning: should we go deeper? (2019). https://doi.org/10.7910/DVN/DKSPJF

25. Kostrzewa, D., Skonieczny, Ł., Benecki, P., Kawulok, M.: B4MultiSR: a benchmark for multiple-image super-resolution reconstruction. Commun. Comput. Inf. Sci. **928**, 361–375 (2018)

26. LeCun, Y.A., Bottou, L., Orr, G.B., Müller, K.-R.: Efficient BackProp. In: Montavon, G., Orr, G.B., Müller, K.-R. (eds.) Neural Networks: Tricks of the Trade. LNCS, vol. 7700, pp. 9–48. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-35289-8_3

27. Ledig, C., et al.: Photo-realistic single image super-resolution using a generative adversarial network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 105–114. IEEE (2017)

28. Li, F., Jia, X., Fraser, D.: Universal HMT based super resolution for remote sensing images. In: 15th IEEE International Conference on Image Processing 2008, ICIP 2008, pp. 333–336. IEEE (2008)

29. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops, vol. 1, pp. 136–144 (2017)

30. Nasrollahi, K., Moeslund, T.B.: Super-resolution: a comprehensive survey. Mach. Vis. Appl. **25**(6), 1423–1468 (2014)

31. Saxe, A.M., McClelland, J.L., Ganguli, S.: Exact solutions to the nonlinear dynamics of learning in deep linear neural networks. arXiv preprint arXiv:1312.6120 (2013)

32. Shan, Q., Li, Z., Jia, J., Tang, C.K.: Fast image/video upsampling. ACM Trans. Graph. (TOG) **27**(5), 153 (2008)

33. Shi, W., et al.: 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, pp. 1874–1883 (2016)

34. Srivastava, R.K., Greff, K., Schmidhuber, J.: Highway networks. arXiv preprint arXiv:1505.00387 (2015)

35. Sun, J., Xu, Z., Shum, H.Y.: Image super-resolution using gradient profile prior. In: IEEE Conference on Computer Vision and Pattern Recognition 2008, CVPR 2008, pp. 1–8. IEEE (2008)

36. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), pp. 2790–2798. IEEE (2017)
37. Tai, Y., Yang, J., Liu, X., Xu, C.: MemNet: a persistent memory network for image restoration. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 4539–4547 (2017)
38. Tai, Y.W., Liu, S., Brown, M.S., Lin, S.: Super resolution using edge prior and single image detail synthesis. In: 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, pp. 2400–2407. IEEE (2010)
39. Xu, B., Wang, N., Chen, T., Li, M.: Empirical evaluation of rectified activations in convolutional network. arXiv preprint arXiv:1505.00853 (2015)
40. Yue, L., Shen, H., Li, J., Yuan, Q., Zhang, H., Zhang, L.: Image super-resolution: the techniques, applications, and future. Signal Process. **128**, 389–408 (2016)
41. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Boissonnat, J.D., et al. (eds.) Curves and Surfaces 2010. LNCS, vol. 6920, pp. 711–730. Springer, Heidelberg (2012). https://doi.org/10.1007/978-3-642-27413-8_47
42. Zhang, K., Zuo, W., Gu, S., Zhang, L.: Learning deep CNN denoiser prior for image restoration. In: IEEE Conference on Computer Vision and Pattern Recognition, vol. 2, pp. 3929–3938 (2017)