Deep learning based single image super-resolution: a survey.

HA, V.K., REN, J., XU, X., ZHAO, S. XIE, G. and VARGAS, V.M.

2018

This is a peer-reviewed, accepted author manuscript of the following conference paper:
Ha, V. K., Ren, J., Xu, X., Zhao, S., Xie, G., & Vargas, V. M. (2018).
Deep learning based single image super-resolution: a survey.
In A. Hussain, B. Luo, J. Zheng, X. Zhao, C-L. Liu, J. Ren, & H. Zhao (Eds.),
Advances in Brain Inspired Cognitive Systems - 9th International Conference,
BICS 2018, Proceedings (pp. 106-119). (Lecture Notes in Computer Vol. 10989. Cham, Switzerland:
Springer-Verlag. https://doi.org/10.1007/978-3-030-00563-4_11

Deep Learning Based Single Image Super-resolution: A Survey

Viet Khanh Ha[1], Jinchang Ren[2,1*], Xinying Xu[2], Sophia Zhao[1], Gang Xie[3], Valentin Masero Vargas[4]

[1]Department of Electronic and Electrical Engineering, University of Strathclyde, Glasgow, UK
[2]College of Electrical and Power Engineering, Taiyuan University of Technology, Taiyuan, China
[3]College of Electronic Information Engineering, Taiyuan University of Science and Technology, Taiyuan, China
[4]Dept. of Computer Systems and Telematics Engineering, Universidad de Extremadura, Badajoz, Spain

*Corresponding Author, Dr. Jinchang Ren (npurjc@gmail.com)

Abstract. Image super-resolution is a process of obtaining one or more high-resolution image from single or multiple samples of low-resolution images. Due to its wide applications, a number of different techniques have been developed recently, including interpolation-based, reconstruction-based and learning-based. The learning-based methods have recently attracted increasing great attention due to their capability in predicting the high-frequency details lost in low resolution image. This survey mainly provides an overview on most of published work for single image reconstruction using Convolutional Neural Network. Furthermore, common issues in super-resolution algorithms, such as imaging models, improvement factor and assessment criteria are also discussed.

Keywords: Image super resolution, convolutional neural network, high-resolution image

1 Introduction

Single image super-resolution (SISR) aims to obtain the visually high-resolution (HR) image from one low-resolution (LR) image. It has found practical applications in real-world problems, from remote sensing where restriction of certain bandwidth and pixel size are present, security surveillance imaging where most information regarding particular scene details, and in medical imaging where reducing irradiation is preferred. Since SISR problem usually assumes the observed LR image is to be a non-invertible low-pass filtering, down-sampled and noisy version of HR image, it is a highly ill-posed problem. There are a variety of methods has been developed recently, which can be classified into interpolation-based, reconstruction-based, and example-based methods. Interpolation-based methods typically adopt fixed-function kernels to estimate the unknown pixels in HR image. Although the interpolation-based methods are very simple and effective ways, they are produce overly smooth edges and blurring details. Reconstruction-based methods usually introduce certain image priors or constraints between the down-sampling of the reconstructed HR image and the original LR images to tackle the ill-posed problem of image super-resolution. Example-based methods, which recently achieved convincing performance, recovered missing high-frequency based on learning the map between LR patches and its HR counterparts. Example-based methods can be categorized into 5 groups: early research [1, 2, 3], sparsity methods [4, 5, 6], self-exemplar methods [7, 8], locally linear regression methods [9-15] and deep architectures [16-36], in which the CNN-based methods have drawn considerable attention due to its simple structure and excellent reconstruction quality.

In this paper, we attempt to provide a brief survey of the research on example-based methods, then focus mainly on CNN-based methods in the context of single image super-resolution. The rest of the paper is arranged as follows. Section 2 give brief review of background, followed by early approaches for super-resolution. Section 3 surveys the contemporary CNN-based approaches, mostly on the-state-of-the-art algorithm and the performance comparison among them is given in section 4. Section 5 will discuss further on multi-resolution, among them fusion methods are widely used. Section 5, 6 give an overall discussion and a conclusion, respectively.

2 Background

Learning-based algorithm aims to hallucinate missing information of the super-resolved images using relationship between LR and HR images. These algorithms contain training step in which the relationship is learn, then the learned

knowledge is then applied to unseen LR images. Although the more training database give more information to apply on unseen data, it is paradox that using larger database does not guarantee better results due to irrelevant examples misleading model to learn more information from noises. Learning-based algorithm for SISR were first introduced in [1-3] in which neighbor embedding [3] was use with idea that low-resolution patches corresponding high-resolution patches share similar local geometries highly influences the subsequent coding-based or dictionary-based methods.

2.1 Sparsity-based method

The sparse representation theory each atom unit is a basic unit that can be used to reconstruct larger units. Also, image patches can be well-represent as a sparse linear combination of elements from appropriately chosen over-complete dictionary. By exploiting sparse representation for each patch of low-resolution inputs, the coefficients of this representation can be applied to generate the high-resolution outputs. Let say, if dimensionality of the input image is 64 x 64 (equal 4096), the dimensionality of dictionary is N x 4096, where N is very large (N > 4096, in this case we have over-complete representation).

$$D \text{ x } \alpha = \text{X} \tag{1}$$

D is basic vector, X is input data and α is unknown. D >> dim(X) in case for super-resolution, where we want to build dictionary for most scenarios of input. To solve over-complete system, assumption that X is composed of no more than a fix number (k) of bases from *D*, then find the set of k bases that best fit the data point X. The observed low-resolution image Y is blurred and down-sampled version of the high-resolution X:

$$\boldsymbol{Y} = S.H.\boldsymbol{X} \tag{2}$$

Here, H represents a blurring filter and S the down-sampling operation. This is ill-posed problem, since for given low-resolution input, infinitely many high-resolution satisfy the constraint. Yang et al. [4] used joint dictionary training to find α coefficient. Given the sampled training patch pairs P = $\{X^h, Y^l\}$, where $X^h = \{x_1, x_2, \dots x_n\}$ are the set of sampled high-resolution image patches and $Y^l = \{y_1, y_2, \dots y_n\}$ are the corresponding low-resolution image patches. Both dictionaries are trained, so sparse representation of high-resolution patch is the same as the sparse representation of the corresponding low-resolution patch. However, limitation appears that two dictionaries are not connected by linear transform also mentioned by authors in this paper. Other works [5-6] proposed to solve two equations but still have many limitations of extracted features and mapping function, which are not adaptive or optimal for generate HR images.

2.2 Self-exemplar methods or Internal database based algorithm

Based on the observation that natural image has self-similarity property, which tends to recur many times inside the image, Glasner et al. [7] proposed a scale space pyramid of LR to match LR and HR pairs. Through the training, patches contained in internal data are more relevant than that of external data. Since internal dictionary are constructed only on given limited LR-HR patch pairs, Huang et al. [8] extended search space to both planar perspective and affine transform of patches to achieve lower errors and more accurate prediction. The complexity of computation makes this method not suitable for real time problem.

2.3 Locally linear Regression methods

An external database based super resolution methods, use external images to try and find mapping between the high-resolution and low-resolution images. The algorithms use different supervised machine learning techniques such as ridge regression [10], anchored neighborhood regression [10, 12], random forest [13], manifold embedding [15]. The database is categorized separately into clustering using k-mean, random forest dictionary to find linear regression.

3 Deep Architecture Methods

3.1 CNNs-based models

The CNNs have been developed rapidly in the last two decades. However, its application to solve SISR problem is first introduced by Dong et al. [16, 17], who described a three-layer CNN for super-resolution as Super-Resolution

Convolutional Neural Network (SRCNN). In this method, CNN has been used to learn the non-linear mapping between the LR and HR images and it significantly outperforms previous non-deep learning methods. The training objective is to find optimal model, given training set $\{x^i, y^i\}_{i=1...N}$. The best mode f then will use to accurately predicts value $\mathbf{Y} = f(\mathbf{X})$, where $\mathbf{X}$ is unobserved examples. The SRCNN [16, 17] consists of following operation, as show in Fig. 1 [16]:

1) Preprocessing: Upscale LR image to desired HR image using bicubic interpolation.
2) Feature extraction: Extracts a set of feature map from the upscale LR image.
3) Non-linear mapping: Maps the feature maps between LR and HR patch.
4) Reconstruction: Produce the HR image from HR patches.

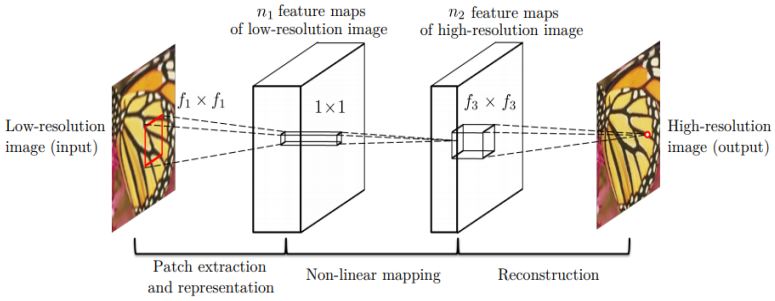

Fig. 1. SRCNN structure [16]

These networks contain only 3 convolutional layers and is improved later with 8 layers [18] impressively outperform conventional non-deep learning methods. However, this model has been mostly restricted to limited training and testing on single scale factor, do not achieve better performance due to the difficulty of deepening networks. This led to observation that whether 'the deeper the better' is or not the case in SR. Inspired of success of very deep networks including Res-Net, Kim et al. [19, 20] proposed two models named Very Deep Convolutional Networks (VDSR) [19] as show in Fig. 2 [19] and Deeply Recursive Convolutional Network [20] (DRCN), both stacking 20 convolutional layers. To speed up training in deep network, the VDSR [19] is trained with very high learning rate (10^{-1} instead of 10^{-4} in SRCNN) in order to accelerate the convergence speed and introduced gradient clipping to control explosion problem. Residual learning is used instead of predict the whole image has several advantages such as fast convergence and better accuracy to compare with SRCNN. In addition, data argumentation allows network adapt well with multiple scale factors (2x, 3x, 4x) without degrading performance. The zero padding also is introduced to avoid the size of feature map reduces quickly through layers of convolution, which appears in deep networks.

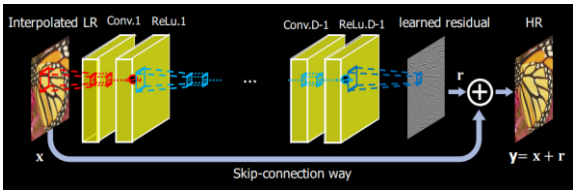

Fig. 2. VDSR model [19] contains 20 convolutional layers, global residual learning represented by skip-connection.

Similar to DRCN [20], Tai et al. [21] proposed Deeply Recursive Residual Network (DRRN), which using both global residual learning and introduces new concept of local residual learning. The global residual learning is used in the identity branch and recursive learning in local residual branch. Mao et al. [22] proposed a 30-layer convolutional auto-encoder network named very deep Residual Encoder-Decoder Network (RED30), as given in Fig.3 [22], which used multiple symmetric connection to boost training convergence. The convolutional layers work as feature extractor, encode image content, while the de-convolutional layers decode and recover image details. This single model has been testing for several tasks of image restoration such as image de-noising, JPEG de-blocking, non-blind de-blurring and image in-painting. [22]

Recent advances in CNN design such as Dense-Net, Network in Network, Residual Network enable numbers of SISR approaches [23, 24, 25] to produce better performances compare to pioneer SRCNN model. Among them, Enhance Deep Residual Networks (EDSR) [26], mostly based on Res-Net model, is convinced to be the-state-of-the-art, as shown in Table 2.

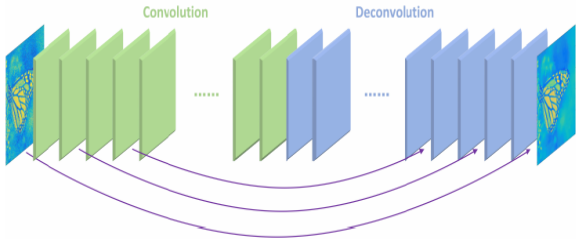

Fig. 3. RED30 structure [22] contain 30 layers. Symmetric skip connection between convolutional and de-convolutional layers

Instead of using interpolation or deconvolution [18] as up-scale method for pre-processing, Lai et al. [27] proposed Laplacian Pyramid Super-Resolution Network (Lap-SRN) to present images as a series of high-pass bands and low-pass bands. This structure enables the residuals (high-pass bands) learn in progressive ways. As shown in Fig.4 [27], at each step, numbers of convolutional layer learn the residual and one transposed convolutional layer to up-sample feature extraction.

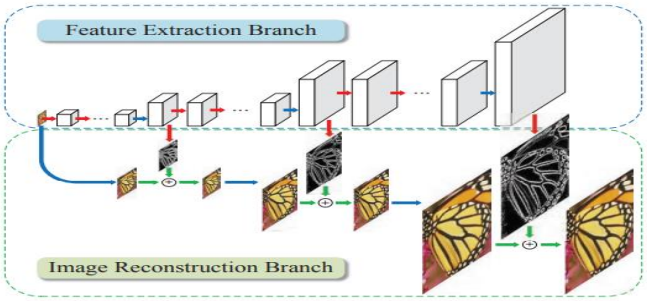

Fig. 4. Lap-SRN structure [27]

One of the drawback with most SISR approaches is that they have been restricted to limit up-scale factors to 2x, 3x, 4x. Otherwise, features available in the LR space have not sufficient to exactly reconstruct the image. To achieve higher scale factor, Wang et al. [28] proposed a fully Progressive Asymmetric Pyramidal Structure to adapt with multiple up-scale factors and up to 8x. Also, a Deep Back Projection Network [29] using mutually connected up- and down-sampling stages has been claimed for reaching such high up-scale factor. To facilitate network training, most CNN-based methods assume that low-resolution image is down-sampled from high-resolution image, they ignored the true degradation in real world application such as noises. Kai Zhang et al. [30] proposed Super-Resolution Multiple Degradation (SRMD) structure with dimensionality stretching strategy scheme to handle blur, noise, and down-sampled image. Shocher et al. [31] inspired by the observation that the natural image has strong internal data repetition, then the information for tiny object is better to be found elsewhere inside the image, other than in any external database of example. Therefore, a 'Zero Shot' SR (ZSSR) is proposed without relying on any prior images example or prior training. It exploits cross-scale internal recurrent of image-specific information, then the test image itself is trained before feed again to resulting trained network.

Although these approaches attempt to higher scale factor and deal with more degradation form of input, they still need to research further to produce persuaded results.

3.2 RNN-CNN-based models

On the view of Recurrent Neural Network (RNN), a Dual-State Recurrent Network (DSRN) [32] allows LR path and HR path captions information at different spaces and connected at every step in order to contribute jointly to learning process. At each stage, LR image are sequence inputs of HR image and vice versa, so called dual state, as given in Fig. 5 [32]. Inspired by concept of Long Short Term Memory (LSTM) block in RNN, Tai et al. [33] proposed Memory Network (MemNet), which uses recursive layers follow by memory unit to allow combination short and long-term memory for image reconstruction, as shown in Fig. 6 [33].

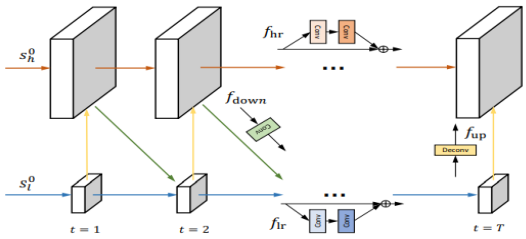

Fig. 5. DSRN structure [32]. The top branch operates on HR space, where bottom branch works on LR space. A connection from LR to HR using de-convolution operation; a delay feedback mechanism to connect previous predicted HR to LR at next stage.

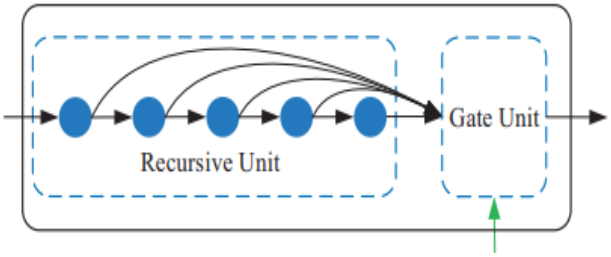

Fig. 6. Memory block in MemNet [33] includes multiple Recursive units and a Gate Unit

3.3 GAN-based model

Basically, Generative Adversarial Network (GAN) contains two models, a generative model and a discriminative model. The discriminative model has the task of determining whether a given image looks natural or looks like it has been artificial created. The task of generative model is created images so that the discriminator gets trained to produce a correct output. The interesting point is during the training, the discriminator is aware of the internal represent of data because it has been trained to understand the differences between real image and artificial created.

One major issue in measurement metric is that a performance of algorithm is commonly measured using pixel-wise such as MSE (as in equation [4]) in favor of maximizing the peak signal-to-noise ratio (PSNR). This will show poorly visual to human perception even with high PSNR due to the mean of many possible solutions. Ledig et al. [34] proposed Super-Resolution Generative Adversarial Network (SRGAN) in favor of perceptual similarity, has delivered great performance for human perception. The extension GAN model is further used in [35, 36], which improved SRGAN with fusion of pixel-wise loss, perceptual loss, texture matching loss. It is shown in Fig. 7 [34], where the optimization desire to human perception, bring reconstruction image look realistic. One of major advantage of GANs-based SR model is that GANs use a largely unsupervised training process on the real images, so it does not require label or prior condition between LR and HR image.

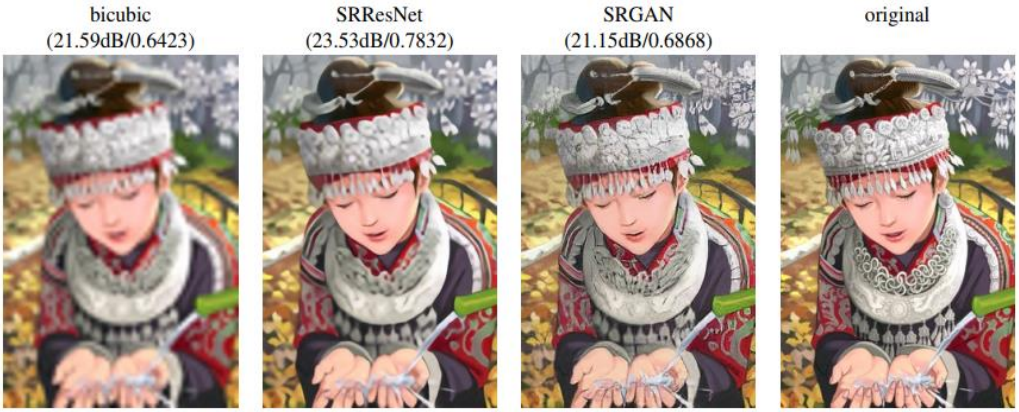

Fig. 7. [34] From left to right, image is reconstructed by bicubic interpolation, deep residual network (SRResNet) measured by MSE, SRGAN optimize more sensitive to human perception, original image. Corresponding PSNR and SSIM are provided on top.

Using MSE-based measurement can have poorly performance even the reconstructed image achieves high PSNR. To support that idea on the side of SRGAN, we also add two of reluctant reconstructed images during implement on SR field. Our model focus on global contextual and get result with low PSNR lower than that of bicubic, as given in Fig 8. It can be explained that errors on pixel at background deteriorate overall result while the pixel value is not equally important. The bicubic interpolated images has better score, though they are all blur.

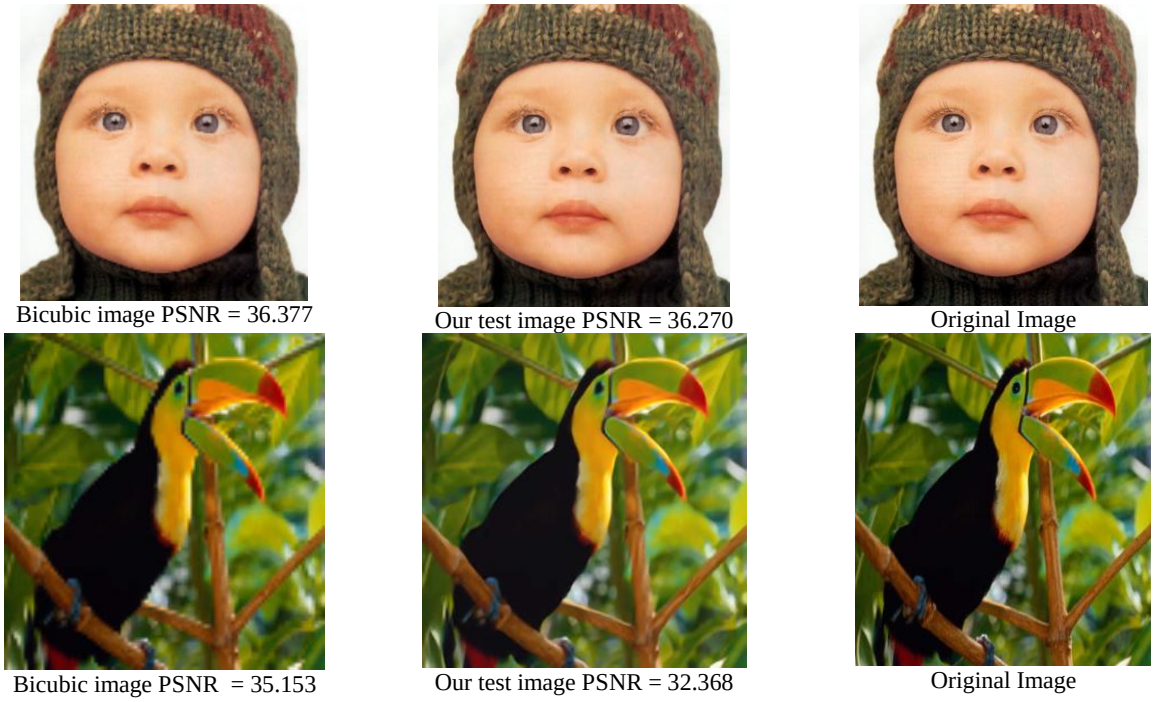

Fig. 8. From left to right, images reconstructed by bicubic interpolation, ours model and original image. PSNR at bottom of each image.

4 Comparison

There are two measures has been mostly used to compare the performance, a Peak Signal-to-Noise Ratio (PSNR) and a Structural SIMlarity (SSIM) index [37]. The higher the PSNR, the better of reconstructed image.

$$\text{PSNR} = 10*\log_{10}\frac{R^2}{\text{MSE}} \tag{3}$$

where R is maximum fluctuate in the input image datatype, MSE is Mean Squared Error between two images, has expressed as:

$$\text{MSE} = \frac{\sum_{M,N}[I_1(m,n)-[I_2(m,n)]^2}{M*N} \tag{4}$$

Here, M and N are the number of rows and columns in the input images, respectively, and SSIM is quantitative measure used to quantify the similarities of structure between two image.

We compare CNN based SR models including SRCNN [17], FSRCNN [18], VDSR [19], DRCN [20], DRRN [21], RED30 [22], MemNet [33], EDSR [26], LapSRN [27], Zeo Shot [31], Dual State [32], SRGAN [34] on algorithm in Table 1 and performance in Table 2. In Table 2, the benchmark datasets are used including SET5, SET14, B100, URBAN 100 which mostly used for comparison in SR algorithms. Scale factor used include 2x, 3x, 4x, and information that were not provided by the authors is marked by [-]. All quantitative results are duplicated from the original papers.

Table 1. Comparison of CNN based SR algorithm. Methods with direct reconstruction perform one-step up-sampling (with bicubic interpolation of transpose convolution) from LR to HR images, while progressive predict HR images in multiple step. Multiple scale factor mean training and testing image down-scaling with multiple factor at the same process, not perform on single factor separately.

Models	Input	Multiple-scale factor	Type of Network	Number of layers	Reconstruction method	Residual	Loss Function	Training time
SRCNN	LR + Interpolation	No	Supervised	3	Direct	No	L2 (MSE)	a week
FSRCNN	LR	Yes	Supervised	8	Direct	No	L2 (MSE)	few hours
VDSR	LR + Interpolation	Yes	Supervised	20	Direct	Yes	L2 (MSE)	4 hours
DRCN	LR + Interpolation	No	Supervised	20	Direct	Yes	L2 (MSE)	6 days
DRRN	LR + Interpolation	Yes	Supervised	52	Direct	Yes	L2 (MSE)	4 days
RED30	LR + Interpolation	Yes	Supervised	30	Direct	Yes	L2 (MSE)	Not given
MemNet	LR + Interpolation	Not given	Supervised	80	Direct	Yes	L2 (MSE)	5 days
EDSR	LR + Interpolation	Yes	Supervised	32	Direct	Yes	L1	8 days
LapSRN	LR	Not given	Supervised	27	Progressive	Yes	Charbonnier	3 days
Zero Shot	LR + Interpolation	Yes	Unsupervised	8	Direct	Yes	L1	days or weeks
Dual Sate	LR + Interpolation	Not given	Supervised	18	Progressive	Yes	L2 (MSE)	Not given
SRGAN	LR through generator network	Yes	Unsupervised + Supervised	54	Direct	Yes	Perceptual loss	Not given

In Table 2, we observe that EDSR outperformed other algorithms with large margin, reached to the-state-of-the-art model recently. Meanwhile, MemNet, achieved in most case the second best performance on SET5 and SET14. The application of residual learning brings benefits to SR image reconstruction, therefor it has been successfully applied in several network models.

5 Multi-resolution related approaches

Image fusion has emerged as a promising research area that aim to combine information from different sources into a single composite for interpretation. It requires the first extraction of the features contained in the various input sources, then characteristics those feature as size, shape, color, contrast, and texture. The fusion is thus enable to detect useful features with higher confidence based on those extracted features. The data fusion has been applied for broad applications in image processing such as in image detection, image registration, image reconstruction. The actual fusion can take place at different types or levels of information representation, from combinations of color and spatial features [38, 39], thermal and visible features [40], spatial and frequency features, spatial and temporal features [41]. When most proposed algorithm processed images in separate colored channel, Yan et al. [38, 39] proposed the fusion of color and spatial features, which is particularly important in retrieval of logo/trademark images. In this method, dominant colors are first extracted via color quantization and k-means clustering, then a component-based spatial descriptor is derived as local features. For detection, Yan et al. [40] proposed the combination of thermal and visible imagery for detection and tracking of pedestrians achieved better distinguishability in human visual perception and less sensitive to these noise effects such as illumination noise and shadows. The fusion of intensity and inter-component chromatic difference has been proposed by Ren et al. [42] for effective and robust color edge detection. Ren et al. also proposed multiresolution decomposition scheme, which decomposes the signal into several components. The 2-D translation is decomposed into two 1-D Fourier transform, which provides improved accuracy in sub-pixel motion estimation [43]. Also, another method [44] based on phase correlation uses linear weighting of the height of the main peak accompany with the difference between two neighboring side peak on the other. This method [44] and gradient-based method [45] effectively deals with noisy image to achieve high accuracy sub-pixel motion estimation. Chai et al. [46] use shape characteristic of a walking object, trajectory-based joint kinematics characteristic motion characteristic of body parts for effective gait recognition.

Last but not least, hyperspectral imagery provides spatial 2-D image in hundreds of different wavelengths, and as the result, gives better capability to see unseen because of its high spectral resolution. However, it requires effective method for dimensionality reduction and feature extraction [47] to reduce level of complexity.

Table 2. Quantitative evaluation of the-state-of-the-art SR algorithm. Average PSNR/SSIM for scale factor 2x, 3, 4x. Red text indicates that the best and blue text indicates the second best performance.

	Scale	Set5 PSNR/SSIM	Set 14 PSNR/SSIM	B100 PSNR/SSIM	Urban100 PSNR/SSIM
SRCNN	2	36.66/0.9542	32.45/0.9067	-	-
	3	32.75/0.9090	29.30/0.8215	-	-
	4	30.49/0.8628	27.50/0.7513	-	-
FSRCNN	2	37.00/0.9558	32.63/0.9088	-	-
	3	33.16/0.9140	29.43/0.8242	-	-
	4	30.71/0.8657	27.59/0.7535	-	-
VDSR	2	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140
	3	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279
	4	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524
DRCN	2	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133
	3	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276
	4	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510
DRRN	2	37.74/0.9591	33.23/0.9136	-	31.23/0.9188
	3	34.03/0.9244	29.96/0.8349	-	27.53/0.8378
	4	31.68/0.888	28.21/0.7720	-	25.44/0.7638
RED30	2	37.66/0.9599	32.94/0.9144	-	-
	3	33.82/0.9230	29.61/0.8341	-	-
	4	31.51/0.8869	27.86/0.7718	-	-
MemNet	2	37.78/0.9597	33.28/0.9142	-	31.31/0.9195
	3	34.09/0.9248	30.00/0.8350	-	27.56/0.8376
	4	31.74/0.8893	28.26/0.7723	-	25.50/0.7630
LapSRN	2	37.52 / 0.959	33.08 / 0.913	-	30.41 / 0.910
	4	31.54 / 0.885	28.19 / 0.772	-	25.21 / 0.756
	8	26.14 / 0.738	24.44 / 0.623	-	21.81 / 0.581
EDSR	2	38.20 / 0.9606	34.02 / 0.9204	32.37 / 0.9018	33.10 / 0.9363
	3	34.77/0.9290	30.66/0.8481	29.32/0.8104	29.02/0.8685
	4	32.62 / 0.8984	28.94 / 0.7901	27.79 / 0.7437	26.86 / 0.8080
Zero Shot	2	37.37 / 0.9570	33.00 / 0.9108	-	-
	3	33.42/0.9188	29.800.8304	-	-
	4	31.13 / 0.8796	28.01 / 0.7651	-	-
Dual Sate	2	37.66 / 0.959	33.15 / 0.913	-	30.97 / 0.916
	3	33.88/0.922	30.26/0.837	-	27.16/0.828
	4	31.40 / 0.883	28.07 / 0.770	-	25.08 / 0.747
SRGAN	2	-	-	-	-
	3	-	-	-	-
	4	29.40/0.8472	26.02/0.7397	-	-

6 Discussion

One possible contribution to SR field is to propose effective models for image reconstruction. However, it is less likely adding more parameters as solutions since super-resolution aims to recover at pixel-level, which requires much more comparison than in classification. The ways to improve image resolution is how to make the neural networks to learn more about the relationship between LR and HR images. While the regular supervised CNN networks attempt to learn directly the mapping and highly depend on predetermined assumptions, GANs based networks are much more flexible with promising performance due to incorporated unsupervised training. Also, traditional measurements expose several constraints to human perception, and the integrated perceptual assessment produces better results. The fusion of unsupervised/supervised models and multi-resolution can reconstruct image with more accuracy and flexibility, yet it still requires further investigation.

7 Conclusion

This paper contains a survey on recent super-resolution techniques that underlie on learning based methods. Among them, we noticed that convolutional neural network based methods have recently achieved the best performance. There are remain challenges to bring them into real time applications since they are only applied on standard benchmark dataset and require to adapt well with differently structured images.

Although LR image is assumed to be a down-sampled version of the HR image, most CNN-based super resolution models fail to work on large scaling down-sampled factors with the exception of noise. The evaluation metrics also have to consider in different perspective of applications. These will also form the base for our future work.

Acknowledgement

The authors would like to thank the support from the Shanxi Hundred People Plan of China and colleagues from the Image Processing group in Strathclyde University for their valuable suggestions.

References

1. Freeman, W.T., Pasztor, E.C. and Carmichael, O.T.: Learning low-level vision. *Int. J. Computer Vision*, *40*(1), pp.25-47 (2000).
2. Freeman, W.T., et al: Example-based super-resolution. *IEEE Computer Graphics and Applications*, *22*(2), pp.56-65 (2002).
3. Chang, H., Yeung, D.Y. and Xiong, Y.: Super-resolution through neighbor embedding. In Proc. *CVPR*, Vol. 1, pp. I-I, (2004).
4. Zeyde, R., Elad, M. and Protter, M.: On single image scale-up using sparse-representations. In Proc. *Int. Conf. on curves and surfaces* (pp. 711-730), Springer, (2010).
5. Dong, W., Zhang, L., Shi, G. and Wu, X.: Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. *IEEE Trans. on Image Processing*, *20*(7), pp.1838-1857 (2011).
6. Peleg, T. and Elad, M.: A statistical prediction model based on sparse representations for single image super-resolution. *IEEE Trans. on Image Processing*, *23*(6), pp.2569-2582 (2014).
7. Glasner, D., Bagon, S. and Irani, M.: Super-resolution from a single image. In Proc. ICCV, pp. 349-356, (2009).
8. Huang, J.B., et al: Single image super-resolution from transformed self-exemplars. In Proc. CVPR, (pp. 5197-5206) (2015).
9. Gu, S., Sang, N. and Ma, F.: Fast image super resolution via local regression. In Proc. ICPR, (pp. 3128-3131), (2012).
10. Timofte, R., De, V. and Van Gool.: Anchored neighborhood regression for fast example-based super-resolution. In Proc. ICCV, (pp. 1920-1927), (2013).
11. Yang, C.Y. and Yang, M.H.: Fast direct super-resolution by simple functions. In Proc. ICCV, (pp. 561-568), (2013).
12. Timofte, R., De Smet, V. and Van Gool, L: A+.: Adjusted anchored neighborhood regression for fast super-resolution. In Proc. ACCV, (pp. 111-126), Springer, Cham (2014).
13. Schulter, S., Leistner, C. and Bischof, H.: Fast and accurate image upscaling with super-resolution forests. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 3791-3799) (2015).
14. Salvador, J. and Perez-Pellitero, E.: Naive bayes super-resolution forest. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 325-333) (2015).
15. Pérez-Pellitero, E., Salvador, J., Ruiz-Hidalgo, J. and Rosenhahn, B. : Psyco: Manifold span reduction for super resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (pp. 1837-1845) (2016).

16. Dong, C., Loy, C.C., He, K. and Tang, X: Learning a deep convolutional network for image super-resolution. In Proc. ECCV, (pp. 184-199). Springer, Cham (2014).
17. Dong, C., Loy, C.C., He, K. and Tang, X: Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Analysis and Machine Intelligence (TPAMI)*, *38*(2), pp.295-307 (2016).
18. Dong, C., et al: Accelerating the super-resolution convolutional neural network. In Proc. ECCV, pp. 391-407, Springer, (2016).
19. Kim, J., et al: Accurate image super-resolution using very deep convolutional networks. In *Proc. CVPR*, pp. 1646-1654, (2016).
20. Kim, J., et al: Deeply-recursive convolutional network for image super-resolution. In *Proc. CVPR*, (pp. 1637-1645) (2016).
21. Tai, Y., Yang, J. and Liu, X: Image super-resolution via deep recursive residual network. In Proc. CVPR, (Vol. 1, No. 4) (2017).
22. Mao, X.J., Shen, C. and Yang, Y.B: Image restoration using convolutional auto-encoders with symmetric skip connections. arXiv preprint. *arXiv preprint arXiv:1606.08921*, *2* (2016).
23. Yamanaka, J., Kuwashima, S. and Kurita, T: Fast and Accurate Image Super Resolution by Deep CNN with Skip Connection and Network in Network. In Proc. NIP, (pp. 217-225). Springer, Cham (2017).
24. Tong, T., Li, et al: Image Super-Resolution Using Dense Skip Connections. In Proc. ICCV, pp. 4809-4817, (2017).
25. Zhang, Y., et al: Residual Dense Network for Image Super-Resolution. *arXiv preprint arXiv:1802.08797* (2018).
26. Lim, B., et al: Enhanced deep residual networks for single image super-resolution. In Proc. CVPR, (Vol. 1, No. 2, p. 3) (2017).
27. Lai, W.S., et al: Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proc. CVPR*, (pp. 624-632) (2017).
28. Wang, Y., et al: A Fully Progressive Approach to Single-Image Super-Resolution. *arXiv preprint arXiv:1804.02900* (2017).
29. Haris, M., et al, N: Deep Back-Projection Networks for Super-Resolution. *arXiv preprint arXiv:1803.02735* (2018).
30. Zhang, K., et al: Learning a Single Convolutional Super-Resolution Network for Multiple Degradations. *arXiv preprint arXiv:1712.06116* (2017).
31. Shocher, et al: " Zero-Shot" Super-Resolution using Deep Internal Learning. *arXiv preprint arXiv:1712.06087* (2017).
32. Han, W., et al. Image super-resolution via dual-state recurrent networks. *arXiv preprint arXiv:1805.02704* (2018).
33. Tai, Y., et al: A persistent memory network for image restoration. In *Proc. CVPR*, (pp. 4539-4547) (2017).
34. Ledig, C., et al: Photo-realistic single image super-resolution using a generative adversarial network. *arXiv preprint* (2016).
35. Sajjadi, M.S., Schölkopf, B. and Hirsch, M: Enhancenet: Single image super-resolution through automated texture synthesis. In Proc. ICCV, (pp. 4501-4510), (2017).
36. Johnson, J., Alahi, A. and Fei-Fei, L: Perceptual losses for real-time style transfer and super-resolution. In Proc. ECCV, (pp. 694-711). Springer, Cham (2016).
37. Ren, J., Zabalza, J., Marshall, S. and Zheng: Effective feature extraction and data reduction in remote sensing using hyperspectral imaging [applications corner]. *IEEE Signal Processing Magazine*, *31*(4), pp.149-154 (2014).
38. Yan, Y., Ren, J., Li, Y., Windmill, J. and Ijomah, W: Fusion of dominant colour and spatial layout features for effective image retrieval of coloured logos and trademarks. In Proc. *IEEE Int. Conf. on Multimedia Big Data* (pp. 306-311). IEEE (2015).
39. Yan, Y., Ren, J., Li, et al: Adaptive fusion of color and spatial features for noise-robust retrieval of colored logo and trademark images. In *Multidimensional Systems and Signal Processing*, *27*(4), pp.945-968 (2016).
40. Yan, Y., Ren, J., Zhao, H., Sun, G., Wang, Z., Zheng, J., Marshall, S. and Soraghan, J: Cognitive Fusion of Thermal and Visible Imagery for Effective Detection and Tracking of Pedestrians in Videos. In *Cognitive Computation*, pp.1-11 (2017).
41. Wang, Z., Ren, J., Zhang, D., Sun, M. and Jiang, J: A deep-learning based feature hybrid framework for spatiotemporal saliency detection inside videos. In *Neurocomputing*, *287*, pp.68-83 (2018).
42. Ren, J., Jiang, J., Wang, D. and Ipson, S.S: Fusion of intensity and inter-component chromatic difference for effective and robust colour edge detection. In *IET image processing*, *4*(4), pp.294-301 (2010).
43. Ren, J., Vlachos, T. and Jiang, J: Subspace extension to phase correlation approach for fast image registration. In Proc. ICIP, (Vol. 1, pp. I-481). IEEE (2007).
44. Ren, J., Jiang, J. and Vlachos, T: High-accuracy sub-pixel motion estimation from noisy images in Fourier domain. In *IEEE Transactions on Image Processing*, *19*(5), pp.1379-1384 (2010).
45. Ren, J., Vlachos, T., Zhang, Y., Zheng, J. and Jiang, J: Gradient-based subspace phase correlation for fast and effective image alignment. *Journal of Visual Communication and Image Representation*, *25*(7), pp.1558-1565 (2014).
46. Chai, Y., Ren, J., Zhao, H., Li, Y., Ren, J. and Murray, P: Hierarchical and multi-featured fusion for effective gait recognition under variable scenarios. *Pattern Analysis and Applications*, *19*(4), pp.905-917 (2016).
47. Zabalza, J., Ren, J., Zheng, J., Zhao, H., Qing, C., Yang, Z., Du, P. and Marshall, S: Novel segmented stacked autoencoder for effective dimensionality reduction and feature extraction in hyperspectral imaging. In *Neurocomputing*, *185*, pp.1-10 (2016)