Single Image Super Resolution Techniques Based on Deep Learning: Status, Applications and Future Directions

Ying Liu[1,2], Yangge Qiao[1], Yu Hao[1,2], and Fuping Wang[1,2]
[1] Center for Image and Information Processing, Xi'an University of Posts and Telecommunications, Xi'an, China
[2] Key Laboratory of Electronic Information Application Technology for Scene Investigation, Ministry of Public Security, Xi'an, China
Email: {liuying_ciip, yangge_q}@163.com, {haoyu, wangfuping}@xupt.edu.cn

Sheikh Faisal Rashid
Artificial Intelligence Research Lab, Al-Khawarazimi Institute of Computer Science (KICS), Department of Computer Science, University of Engineering & Technology (UET), Main GT Road Lahore, Pakistan
Email: shfaisal@gmail.com

***Abstract*—Single Image Super Resolution (SISR) reconstruction aims to recover high-resolution images from corresponding Low-Resolution (LR) versions, which is essentially an ill-posed inverse problem. In recent years, learning-based methods have been frequently exploited to tackle this problem, which correspond to promising calculation efficiency and performance, especially in image sharpening processing based on deep neural networks. Learning-based methods can be generally categorized as conventional methods and deep learning-based methods. This survey aims to review deep learning-based image super-resolution methods, including Convolutional Neural Networks (CNN) and Generative Adversarial Networks (GAN) based on internal network structure. Furthermore, this paper describes the applications of single-frame image super resolution in various practical fields. In addition, a few future research directions of image super resolution techniques are identified.**

***Index Terms*—single image super-resolution reconstruction, deep learning, convolutional neural networks, generative adversarial networks**

I. Introduction

Image Super Resolution (SR) technology refers to the use of existing Low Resolution (LR) imaging systems to obtain the original images, and restore corresponding High-Resolution (HR) image through related algorithms. HR images can provide more detailed information, it has been widely applied in various fields [1]-[8]. SR technology was firstly proposed by Harris [9], and the research achievements of conventional SR algorithms are summarized in [10]-[12]. SISR is a special case of SR, which can correct image quality degradations caused by different conditions such as weather and hardware equipment. SISR algorithms [10] include interpolation-based methods [13]-[15], reconstruction-based methods [16]-[18], and learning-based methods [19]-[22]. Specifically, traditional algorithms often use complex prior knowledge to constrain the solution, but they are extremely complicated to restore the lost details in higher magnification factors. In this case, the reconstruction effect of these methods can't meet the practical demands [17]. The learning-based methods obtain prior information from an external training database by learning the mapping relationship between HR and LR images, and then reconstruct high resolution images. Although this method overcomes the defects of the previous ones, they rely too much on external training sets and often results in unwanted artifacts in the reconstructed HR image [23].

In recent years, due to the powerful representation abilities of Deep Learning (DL) models, the ill-posed inverse problem of SISR has been successfully tackled to a certain extent. By adopting the end-to-end non-mapping relationship between LR and HR images, it can adaptively learn the deep features to recover the texture details, and achieve more advanced performances. Numerous deep learning SISR algorithms such as Super-Resolution based on Convolutional Neural Network (SRCNN) [24], Very Deep Convolutional Networks (VDSR) [25], and Deeply Recursive Convolutional Network (DRCN) [26] have been proposed since 2014. In addition, scholars have also reviewed various deep learning based image super resolution algorithms [27]-[30] to contribute the development of this field. This paper summarizes the single-frame deep SISR reconstruction algorithms, introduces the most popular methods in detail, and lists traditional learning-based methods. Besides, this survey focuses on exploring different deep network-based SR methods, which divides these methods into two categories for analysis and comparison, and describes the applications of SISR reconstruction in four major areas.

The remainder of this paper is organized as follows. Section II introduces three traditional learning-based methods. Section III divides the current mainstream deep

Manuscript received January 29, 2021; revised July 23, 2021.

doi: 10.18178/joig.9.3.74-86

SISR methods into two categories, including convolutional neural network models and generation adversarial network models. Section IV summarizes the benchmarking datasets and experimental evaluation criteria of SR technology. Section V describes the applications of SISR reconstruction in various specific fields. Section VI discusses the future research directions of SISR.

II. The Theoretical Background of SISR Technology

A. *Imaging Model*

Single-image SR reconstruction refers to the effective recovery of high-frequency detail information lost in a single LR image, and finally obtaining an HR image. Assuming Y is the degraded LR image and X is the HR image, the entire degradation process can be illustrated as Fig. 1. The HR image is finally degraded into LR image through motion deformation M, blurring B, down-sampling S, additional noise N and H is the transfer function of the imaging process. The degradation model is expressed as in (1).

$$Y=HX+N,\ H=MBS \tag{1}$$

The problem of SR reconstruction is an inverse process of reconstructing X from Y. Because detailed information has been lost during the image acquisition process from HR to LR, the input LR image corresponds to many possible HR solutions. To obtain a stable solution, various regularization techniques have been proposed and introduced into the conventional restore framework.

B. *Learning-Based Methods*

Traditional algorithms include neighbor embedding method [31], sparse representation method [32], local linear regression method [33], [34]. In 2004, Chang [31] proposed a Neighbor Embedding (NE) method. The main idea of this method is using a similar local structure to reconstruct the HR image block. However, the fixed K neighborhood size may cause overfitting or underfitting. To tackle this disadvantage, Gao [20] combined sparse neighborhood search with clustering histograms of oriented gradients-based subsets [35] to reduce calculation time and maintain SR quality. Based on compressed sensing theory [36], Yang [32] proposed a sparse representation method in 2008, which takes the sparsity of the image as a prior constraint and combines the sparse representation of the LR image block with the dictionary D_H to generate HR image details. However, this method generates heavy calculation burden when solving sparse coding coefficients. Yang [22] proposed a coupled dictionary learning method based on LR-HR patch feature space, which directly uses the network to regress the sparse coefficients and further reduce the computation time. In 2013, Timofte [33], [34] proposed the Anchored Neighbor Regression (ANR) algorithm, which converts L_1 norm into L_2 norm to constrain the least squares problem, and directly calculates the dictionary atoms of the sample neighborhood.

All three traditional methods obtain LR and HR data from external training sets, and they use manifold learning and sparse representation to learn the mapping relationship between LR-HR image-pairs. Compared with interpolation and reconstruction methods, learning-based methods enhance the image resolution and image quality to a certain degree, and the reconstruction results are better with a larger amplification factor. But its training process is time-consuming and relies overmuch on the similarity between external training sets and the test image, which has a poor effect on suppressing aliasing artifacts in reconstructed images [23]. Recently, the deep learning methods have shown great advantages that the deep learning SR model directly establishes the end-to-end relationship from LR and HR images using multi-layer neural networks. It broke the bottleneck problem of traditional methods with efficient computing efficiency and powerful data processing capability, and the quality of reconstructed image is improved. By learning a nonlinear network structure to represent the input LR data, deep learning methods automatically learn the abundance of input and output information directly from the data [37]. Therefore, the network's ability of automatic learning image features is crucial.

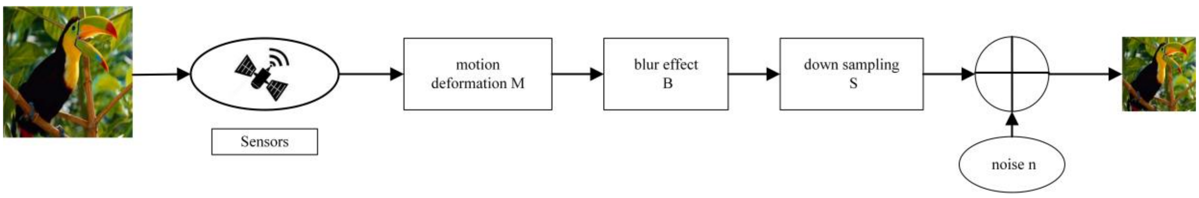

Figure 1. Degradation model of a single image.

III. SISR Based on Deep Learning

In 2006, Hinton [37] proposed the concept of DL, which is a new branch of machine learning algorithms. DL methods directly learn the mapping relationship between LR and HR images through an end-to-end training model and have strong automatic learning capabilities, which are widely applied in many fields such as image classification [38], object recognition [39]. Recently, many scholars have conducted extensive researches on the SISR problem, which address ill-posed problems by using various deep learning-based techniques.

This section introduces SISR techniques based on deep learning, which are mainly summarized as the CNN-based SISR method and GAN-based SISR method. Fig. 2 illustrates the overall classification of various models.

The convolutional neural network-based methods are further divided into linear structure, recursive structure, dense convolutional structure, multi-scale reconstruction network and the networks combined attention mechanism.

A. Convolutional Neural Network

1) Linear structure

The linear structure is achieved by stacking several convolutional layers together and sequentially transmitting to the output layer through the input layer. The network performs different upsampling operations for LR images, including predefined upsampling and post-upsampling.

Interpolation pre-processing methods: The predefined upsampling uses a linear network to perform upsampling on the input LR image. In other words, the size of the input LR image is enlarged similarly as the target HR image using interpolation pre-processing techniques and so on. It is worth noting that the input LR image is of arbitrary size or upscaling factor. SRCNN was firstly proposed for image SR reconstruction. Fig. 3 shows the overall structure of SRCNN, f1, f2, and f3 to respectively represent the size of the convolution kernel. The SRCNN consists of feature extraction of image blocks, feature mapping to high-dimensional feature vectors, and image reconstruction. Compared with traditional methods, SRCNN combines Sparse Coding (SC) with CNN to enhance the reconstruction speed and performance. Because the network is too shallow and convolution kernel is small, the reconstruction performance is affected and the deep features of image can't be obtained.

Compared with the simple structure of SRCNN, Kim [25] proposed a deep VDSR network. In the training phase, the network uses higher learning rate and residual learning [40] to accelerate the network convergence, and suppresses the problem of gradient explosion through adjustable gradient clipping, to improve the stability of network training. The mathematical formula of VDSR is shown in (2). Where d represents the number of convolutional layers, f denotes the convolution layer output function, x and y are input and output of VDSR respectively.

$$y = f_{Rec}\left(f_{d-1}\left(f_{d-2}\left(kf_1\left(x\right)\right)k\right)\right) + x \tag{2}$$

Based on predefined upsampling methods, the input LR image uses an interpolated image of any size and scale factor, which has noise amplification and blurring issues. Some methods have been designed to solve this problem. For example, Zhang [41] used feed-forward denoising convolutional neural network to improve image denoising performance; Zhang [42] proposed to use CNN denoiser in a model-based optimization method.

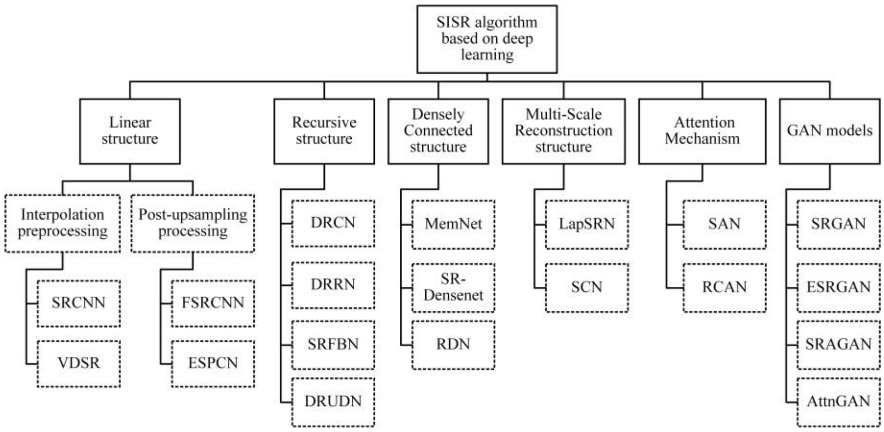

Figure 2. Overall classification of deep learning SISR methods.

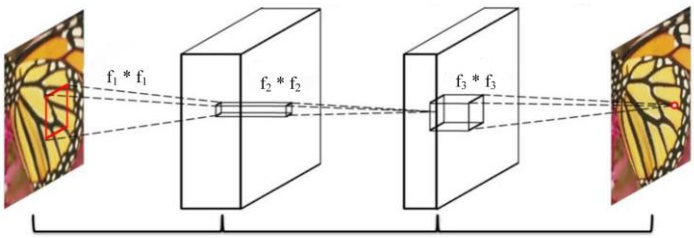

Figure 3. SRCNN network structure [24].

Post-upsampling processing methods: Although the predefined upsampling methods improve the training speed of the convolutional network, it increases the computational complexity. Researchers proposed sub-pixel convolutional or deconvolutional layers to solve the above shortcomings. The core concept of post-upsampling processing is to send low-resolution images to deep CNN for processing without increasing the resolution and to apply upsampling processing at the end of the network.

Ref. [43] proposed fast SR method based on CNN (FSRCNN), which introduces a deconvolution layer to realize the upsampling process. Shi [44] proposed a high-Efficiency Sub-pixel Convolutional Network (ESPCN).

The main idea is replacing the ordinary bicubic interpolation method with a zero-filled sub-pixel interpolation method. The original LR image is sampled as the HR image through the convolutional network, and the LR feature map is gathered through the sub-pixel convolutional layer at the end of the network. Because the feature extraction process of the ESPCN is performed in a low-dimensional space, the resolution is improved at the sub-pixel layer, this method reduces the amount of calculation and space complexity.

In the predefined upsampling method, SRCNN combines SC with CNN to generate HR images. VDSR uses a deeper structure (20 layers) to reconstruct better quality HR images. In the post-upsampling method, FSRCNN and ESPCN methods directly perform feature extraction on LR images, and use deconvolution and sub-pixel convolution for upsampling at the end of the network to improve the speed of reconstruction.

2) Recursive structure

The recursive network uses recursively connected units to deepen the convolutional layer and improve the reconstruction quality of SR [26]. While reducing the problem of more parameters caused by the deep network, the model can obtain better generalization ability [45].

Deep recursive network: Although the deep network layer can expand the receptive field and exploit more context information to describe high-frequency details, it also brings problems such as more parameters, difficulty in network training and storage. To control the model parameters, Kim [26] introduced a Deep Recursive Convolutional Network (DRCN). As the number of recursive layers increases, numerous parameters remain unchanged due to weight sharing. In the training strategy, the recursive multi-supervision [46] is used to alleviate effects of disappearing/explosive gradients. Also, the input LR image and HR image in SR are related and share the same information to a certain extent. Long [47] uses skip connections to reduce the difficulty of DRCN training convergence. The result shows that the combination of recursive network and skip connection enhances the feature learning ability of the network and is superior to SRCNN and VDSR in the reconstruction of reference images. The mathematical formula of DRCN is expressed in (3). Li [48] proposed a Deep Recursive Up-down Sampling Network (DRUDN). The main concept is performing nonlinear mapping through recursive up-down sampling blocks and applying upsampling processing at the end of the network. The reconstruction accuracy of this network is higher than DRRN. Where T represents the number of recursions in DRCN, and f represents the output function of the convolutional layer, x and y are input and output of DRCN respectively.

$$y=\sum_{t=1}^{T} W_t\left(f_{Rec}\left(f_2^{(t)}\left(f_1(x)\right)\right)\right)+x \tag{3}$$

Deep recursive residual network: Ref. [45] proposed a Deep Recursive Residual Network (DRRN) which is similar to DRCN. Without increasing any weight parameters, the network is deepened to 52 layers through residual recursive block stacking, which reduces the computational cost and improves the performance of SR. In terms of training strategy, DRRN combines global residual learning and multi-path local residual learning to reduce the difficulty of deep network training. The mathematical formula of DRRN is expressed as (4). Where B and R represent the number of remaining recursive blocks and the output function of the recursive module respectively, x and y are input and output of DRRN respectively. Lin [49] used split-concatenate-residual to reduce parameters, improve image quality and save run time in the recursive network.

$$y=f_{Rec}\left(R_B\left(R_{B-1}\left(k\left(R_1(x)\right)k\right)\right)\right)+x \tag{4}$$

Super-resolution feedback network: Most previous algorithms (SRCNN, FSRCNN, etc.) share the information of the network layer in a feed-forward manner, so that the former layer of the network can't access the useful information of the latter layer, which limits the accuracy of reconstruction. Li [50] proposed a Super-Resolution Feedback Network (SRFBN) based on a recursive structure, which realizes feedback connection through RNN, and uses high-resolution information to recursively refine low-resolution information. SRFBN uses Feedback Block (FB) as the basic module, and utilizes the information in multiple sets of up-down sampling layers through dense skip connections. It effectively handles the feedback information flow and function reusing. Besides, when there are multiple types of degradation in the LR image, the use of curriculum learning strategies can help the model to train well.

Compared with the linear network, the recursive structure has deeper network layer, which can alleviate the problem of more training parameters caused by the network depth. The above three methods have differences in network complexity, SR framework, and key strategies. Among them, DRCN mainly achieves the information transmission between layers; DRRN combines global and local residual information to reduce the difficulty of training, and the basic module is the residual unit; SRFBN directly extracts features from the original LR image, using feedback connection method to learn LR information.

3) Densely connected structure

Most CNN-based deep SR models can't fully exploit the feature information of the original LR image, which leads to relatively poor reconstruction quality. In the process of reconstructing HR images, the input LR images are expected to obtain more information, which requires a large receptive field and hierarchical feature information extracted from the network layer [51].

Memory network: Ref. [52] proposed memory network (MenNet) with 80 convolutional layers, which are currently the deepest network model. The network consists of three parts: Feature extraction network - extracts LR images features; A series of memory blocks stacked in densely connected structures - retains high frequency information; Reconstruction network - reconstruct residual images. In the network structure, the basic module of MenNet is a Memory Block (MB) (each layer has different weights), which includes a recursive unit and a gate unit. The gate unit learns the adaptive weights of different memories to achieve long-term memory. Compared to MenNet (MemNet-NL) [52] without persistent memory, MemNet uses multi-path structure dense connection between memory blocks to facilitate information transmission on the network and strengthen the recovery of medium and high frequency signals.

Super-resolution using dense skip connections: Tong presented a SR network using dense skip connections (SR-DenseNet) [53]. SR-DenseNet is a 64-layer deep network and adapts Dense Blocks (DB) as the basic module. This method learns the dense network blocks of high-level features, and merges low-level/high-level features through dense skip connections. Finally, the HR image is reconstructed. Unlike DRCN and DRRN, SR-DenseNet continuously uses deconvolution layers at the end of the network to implement the up-sampling process, which improves the efficiency of SR reconstruction. Because the dense skip connection method combines the information of low-level and high-level features, it provides richer details for reconstructing HR images and effectively avoids the problem of gradient disappearance. Compared with SRCNN, the Peak Signal to Noise Ratio (PSNR) value of this method is improved by about 1.0 dB. Compared with VDSR, the PSNR value is improved by about 0.5 dB.

Residual dense network: Ref. [54] designed a Residual Dense Network (RDN) that combines dense connection and residual learning. Compared with the previous models (MenNet, DRCN, DRRN), the advantage of RDN is to direct extraction of hierarchical features from the original LR image, which reduces the computational complexity. In addition, the network uses Residual Dense Block (RDB) as the basic module to extract rich local features, integrates the global hierarchical features of the previous layer in the LR space and improves the sharing of information at various levels by establishing a continuous memory mechanism. the experimental results show that the HR image details recovered by the algorithm are more refined.

Inspired by the feedback network, Haris [55] proposed an SR Dense Deep Back-Projection Network (D-DBPN). Using the iterative up-sampling SR framework, the network directly learns the feedback error signal between LR and HR images through iterative back-projection. Liu [56] proposed a so-called Residual Feature Aggregation Network, which effectively utilizes the hierarchical features on the residual branches and improves the performance of dense networks.

Among the three SR algorithms based on dense connections, the MenNet method exploits multi-path dense connected memory blocks to effectively process the information flowing between layers. But the original LR image needs to be interpolated pre-processing at first; SR-DenseNet directly inputs LR images in the network, and dense skip connections are used between dense blocks; the RDN method uses global feature fusion strategy to achieve the transmission of information between layers.

4) Multi-scale reconstruction structure

Both the post upsampling SR frameworks FSRCNN and ESPCN can't satisfy the requirement of multi-scale super-resolution. Therefore, the deep laplacian pyramid super-resolution network (LapSRN) proposed by Lai [57]. This network can handle multi-scale SR with a large magnification factor. The cascaded CNN gradually reconstructs high-resolution images and refines them by CNNs. This greatly reduces the learning difficulty, especially in the case of large factors, the algorithm reconstruction effect is better. Secondly, some scholars have considered combining traditional methods with deep neural networks. For example, Wang [58] proposed a Sparse Coding-Based Network (SCN) using the learned Iterative Shrinking Threshold Algorithm (LISTA) [59] to estimate sparse coding. The main advantage of this algorithm is all parts of the sparse coding are jointly trained through back-propagation, and multiple cascaded SCNs achieve SR with any amplification factor. It solves the time-consuming reasoning problem in traditional sparse coding and increases the model's flexibility.

5) Attention mechanism

The SR model of the attention mechanism can select more important activation values, and assign more weights, thereby improving the reconstruction effect. Dai [60] proposed a Second-Order Attention Network (SAN). the network uses second-order channel attention as a basic module, which adopts second-order feature statistics to adaptively adjust the channel to learn more expressing the relationship between features. From the perspective of channel attention, Residual Channel Attention Networks (RCAN) [61] uses the residual channel attention module to enable the network to assign different weights to different characteristic channels, thereby constructing a very deep network.

B. Generative Adversarial Network

In 2014, Goodfellow [62] proposed GAN. GAN is composed of generator G and discriminator D. The generator and the discriminator compete with each other until the discriminator can't distinguish between real and fake sample images. The main purpose of GAN training is to output high-probability real samples for D; So that D gives high-probability sample data for G. i.e., achieve Nash equilibrium [63]. In recent years, many scholars attempt to use GAN for image SR reconstruction tasks to generate visually high-quality HR images.

1) Improved method based on SRGAN

In 2017, Ledig [64] applied GAN to SISR for the first time, proposed an image Super Resolution Generative Adversarial Network (SRGAN). As shown in Fig. 4, the depth generator network is accumulated by residual blocks to generate HR images. the discriminator network is composed of 8 convolutional layers to optimize the generated HR image. Due to the Mean Square Error (MSE) pixel loss function is limited in capturing high-frequency details, sometimes the PSNR value is high but the quality of the reconstructed image is poor [65]. Therefore, SRGAN introduces perceptual loss to make the generated image with excellent visual effects [66]. As shown in (5), the perceptual loss consists L^{SR} of two parts: content loss L_X^{SR} and confrontation loss L_{Gen}^{SR}.

$$L^{SR} = L_X^{SR} + 10^{-3} L_{Gen}^{SR} \tag{5}$$

SRGAN can generate explicit image details, but some created textures are distorted. Wang [67] improved the SRGAN network and proposed a residual block-based encoder module, which removed the batch standard layer and added a micro-encoder network to extract key feature information. Compared with the original SRGAN, the improved network improves evaluation performance and generates clearer and more natural images. Wang [68] proposed the enhanced SRGAN (ESRGAN), which uses Residual-in-Residual Dense Block (RRDB) as the basic network unit, by removing Batch Normalization (BN) layer and merging dense blocks to improve the artifact problem in SRGAN to obtain more realistic images.

2) Attention generation network

The generation network utilizes the attention mechanism model to screen important activation values, and assigns more weights to improve the quality of the reconstructed image. Wang [69] proposed an SR reconstruction based on self-attention GAN (SRAGAN), which combines the self-attention mechanism and residual module to form a deep generator. The global feature information of the self-attention layer is used to re-construct the HR image. SRAGAN is superior to current algorithms in objective evaluation indicators, improves the richness of the HR image, and makes the generated image more realistic. Xu [70] proposed the Attentional Generative Adversarial Networks (AttnGAN) for fine grained text to image synthesis. The model consists of two parts including AttnGAN and Deep Attentional Multimodal Similarity Model (DAMSM). Among them, DAMSM provides additional fine-grained image-text matching loss for the training generator, and the initial score on the COCO dataset is increased by 170.25%. It proves the effectiveness of the attention mechanism and the performance of the GAN model is improved.

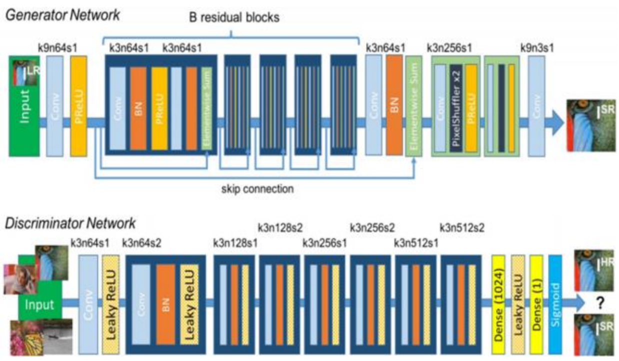

Figure 4. SRGAN structure [64].

C. Other Networks

Additionally, researchers have also proposed Restricted Boltzmann Machine (RBM) and Deep Belief Nets (DBNs). other methods such as RBM proposed by Gao [71] regards the sparsity of the image as a prior constraint and synthesizes HR images by learning dictionary pairs. The hidden layer of RBM is used to calculate the sparsity coefficient, and the HR-LR image blocks are obtained through the observation layer. Experiments show that the algorithm has strong robustness to different initialization parameters. Nakashika [72] used the SR method based on DBNs including the training and recovery phases. In the training stage, it uses the two-dimensional discrete cosine transform coefficients to train the DBNs so that the interpolated low score image is the same size as the image block of the training data. In the recovery stage, the trained DBN is used to restore the missing high frequency components after the image is transformed into the frequency domain.

D. Analysis and Performance Comparison

Most super-resolution algorithms mainly report PSNR and SSIM (structural similarity image measurement) values on the benchmark data sets for performance evaluation. Table I compares the performance of existing SISR methods with the experimental data taken from the original literature. Compared with the traditional techniques based on low-level features, the SR algorithms based on deep learning directly establish an end-to-end relationship between the LR and HR images using a multi-layer neural network, and achieve better reconstruction performance.

Comparing shallow and deep structures of deep learning algorithms, the PSNR and SSIM values of certain shallow networks including SRCNN (3 layers), FSRCNN (8 layers), and ESPCN (3 layers) are slightly lower than other deep networks. The depth of the network layer also has significant impact on the reconstruction effect. When the network layer number decreases, the receptive field will be smaller, and its ability to acquire deep features and the reconstruction effect become more limited. In the deep network, the implementation of residual learning, recursive learning, dense connection and attention mechanism strategies plays an important role in improving the quality of image super-resolution. VDSR (20 layers) uses global residual learning to increase the network’s stability. DRCN (20 layers) exploits an inter-layer link information sharing structure while combining global residuals and recursive learning, which can improve network’s performance. SR-DenseNet and RDN adapt a dense connection method to better integrate the information transmission between network layers, and achieves test results with higher ranks. Especially, the SAN and RCAN models obtain the best reconstruction effect, which fully demonstrates the effectiveness of the channel attention mechanism.

TABLE I. EXPERIMENTAL RESULTS OF DIFFERENT SR ALGORITHMS IN FOUR TEST DATASETS (WITH UPSCALING FACTOR ×2, ×3, ×4)

Method	Scale	Set5	Set14	BSD100	Urban100
		PSNR/SSIM			
Bicubic	×2	33.66/0.9299	30.24/0.8688	29.56/0.8431	26.88/0.8403
	×3	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349
	×4	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577
SRCNN [24]	×2	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946
	×3	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989
	×4	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221
FSRCNN [43]	×2	36.98/0.9556	32.62/0.9087	31.50/0.8904	29.85/0.9009
	×3	33.16/0.9140	29.42/0.8242	28.52/0.7893	26.41/0.8064
	×4	30.70/0.8657	27.59/0.7535	26.96/0.7128	24.60/0.7258
ESPCN [44]	×2	-	-	-	-
	×2	33.13/	29.49/	-	-
	×2	30.90/	27.73/	-	-
VDSR [25]	×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 [26]	×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 [45]	×2	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188
	×3	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378
	×4	31.68/0.8888	28.21/0.7721	27.38/0.7284	25.44/0.7638
MemNet [52]	×2	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195
	×3	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376
	×4	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630
RDN [54]	×2	38.24/0.9614	34.01/0.9212	32.34/0.9017	32.89/0.9353
	×3	34.71/0.9296	30.57/0.8468	29.26/0.8093	28.80/0.8653
	×4	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028
SR DenseNet [53]	×2	-	-	-	-
	×3	-	-	-	-
	×4	32.02/0.8934	28.50/0.7782	27.53/0.7337	26.05/0.7819
SAN [60]	×2	38.35/0.9619	34.44/0.9244	32.50/0.9038	33.73/0.9416
	×3	34.89/0.9306	30.77/0.8498	29.38/0.8121	29.29/0.8730
	×4	32.70/0.9013	29.05/0.7921	27.86/0.7457	27.23/0.8169
RCAN [61]	×2	38.27/0.9614	34.12/0.9216	32.41/0.9027	33.34 0.9384
	×3	34.74/0.9299	30.65/0.8482	29.32/0.8111	29.09/0.8702
	×4	32.63/0.9002	28.87/0.7889	27.77/0.7436	26.82/0.8087

IV. PUBLIC DATASETS AND EVALUATION CRITERIA

A. *Public Datasets*

Many image SR datasets are available differing in image quantity, quality, resolution, and diversity. The current datasets of image super-resolution mainly include public datasets and datasets for specific fields. Common datasets refer to natural image datasets for academic research, and specific datasets refer to datasets for specific research objects, such as face image datasets, medical image datasets, small objects image datasets, etc. Wang [30] reviewed some public available benchmark datasets and evaluation indicators. In the past year, many researchers have also published SR image datasets obtained from real scenes, such as City100 [73], SR-RAW [74], and RealSR [75]. The SISR data sets commonly used by scholars and the specific description of its year, number of HR images, image format, and other related introductions are shown in Table II. The representative sample images of the commonly used datasets are shown in Fig. 5.

B. *Evaluation Criteria*

Generally, the performance of the SR algorithm is evaluated through the following two aspects:

Subjective evaluation: This method mainly based on the subjective evaluation of image quality by physical eyes. But different people have different perceptions of the same image. This evaluation method contains many subjective factors and individual differences, such commonly named as the Mean Opinion of Score (MOS). In most cases, objective evaluation is required to compare different reconstruction algorithms.

Objective evaluation: The input LR image is usually obtained by the degraded model of HR image in the SR algorithm. The objective evaluation method refers to the implementation of calculation methods to decide the similarity between the original image and the reconstructed HR image. Then, it evaluates the quality of the reconstructed image by a certain algorithm. Objective evaluation methods are still the current mainstream method. Two most frequently applied objective standards are PSNR and SSIM.

$$\mathrm{MSE} = \frac{\sum_{i=1}^{M}\sum_{j=1}^{N}\left(X_{ij}-Y_{ij}\right)^2}{M \times N} \tag{6}$$

$$\mathrm{PSNR} = 10\log_{10}\frac{255\times 255}{\mathrm{MSE}} \tag{7}$$

$$\mathrm{SSIM}(X,Y) = \frac{\left(2\mu_X\mu_Y + C_1\right)\left(2\delta_{XY} + C_2\right)}{\left(\mu_X^2+\mu_Y^2+C_1\right)\left(\delta_X^2+\delta_Y^2+C_2\right)} \tag{8}$$

where MSE is the mean square error of X and Y. X is the original HR image, Y is the reconstructed HR image, and M and N represent the image size. The smaller the error between X and Y is, the higher the PSNR value will be. SSIM compares the similarity between X and Y based on the brightness, contrast, and structure. The larger this value is, the higher the similarity and the better the reconstruction effect will be. The unit of PSNR is the decibel (dB), and the value of SSIM ranges between 0-1.

V. Doman-Specific Application of SISR

The problem of poorly captured image quality brings great difficulties to subsequent image processing tasks in real life, such as remote sensing [7], target detection [2]. In this case, scholars are committed to improve the spatial resolution of images through software technology and provide promising research directions in the future.

TABLE II. Common Public Datasets

Name	Years	Number of pictures	Image format	Related introduction
ImageNet [76]	2009	400,000	JPEG	ILSVRC detection dataset
Set14 [77]	2010	14	PNG	Test data set, images include humans, animals, insects, flowers, vegetables, comics, slides, etc.
Set5 [78]	2012	5	PNG	Test data set, images contain child, bird, butterfly, head, woman
BSD300 [79]	2001	300	JPG	200 images for training, 100 for testing
Urban100 [80]	2015	100	PNG	100 high-resolution pictures of buildings
MS-COCO [81]	2014	328,000	JPG	Contains 91 easily identifiable object classes
DIV2K [82]	2017	1000	PNG	800 training images, 100 verification images and 100 test images, including people, handmade products and environment (city, village), natural scenery, etc.
T91 [21]	2010	91	PNG	Training set, including cars, flowers, fruits, human faces, etc.

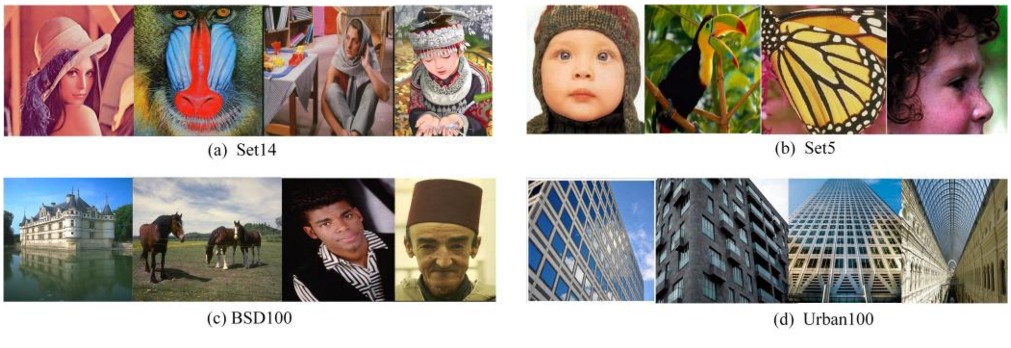

Figure 5. Sample images of part of the datasets.

A. Remote Sensing Image Super-Resolution

The SISR algorithms for remote sensing images are mainly divided into self-learning methods and example learning methods. For the self-learning methods, Vishnukumar [83] utilized the self-similarity in images, and linearly combined the dictionary of HR image blocks with the sparse representation coefficients of LR image blocks to obtain the final HR image, which doesn't require an external database. For the example-based methods, Pan [84] presented the Residual Dense Back Projection Network (RDBPN), which uses residual back projection block structure to improve the resolution of large-scale factor remote sensing images. In the real world, LR remote sensing images may not be generated by a specific down-sampling method, and hence the trained supervised learning algorithm is usually not effective in real applications. Wang [85] proposed an unsupervised learning network based on cyclic CNN, which consists of two cyclic generation networks. the unsupervised method has stronger robustness and exhibits better results in reconstructing the GaoFen-2 satellite image, compared to the supervised algorithms. In present, the main challenge for remote sensing image SR is how to improve the existing SR method to process numerous remote sensing images taken from satellites.

B. Face Image Super-Resolution

Face SR images are captured from surveillance cameras or other imaging systems, which provide important information for human visual perception and criminal investigation cases. Due to the limitation of imaging conditions, LR face image loses the structural details and hence can't be accurately restored in many cases. Talab [86] proposed a face recognition technique, which combines ESPCN and CNN. Specifically, ESPCN network converts the LR image into an original HR image while CNN network obtains the final HR face image. Liu [87] proposed to combine dense connections with attention mechanisms to improve the accuracy of low-resolution face recognition. Zhong [88] proposed high quality face images based on GAN. Specifically, the basic unit of the generator is the residual dense block without the BN layer, which is combined with the inception framework to depict LR images better. The combination of deep learning and image SR reconstruction can restore important details of the face, and the recovery effect is poor in practical applications because of the uncertain factors of the imaging systems.

C. Medical Imaging Super-Resolution

CT imaging is a clinical diagnosis technique. Due to hardware limitations, the resolution of CT images is limited and hence doctors can't determine the smaller lesion areas accurately, which would affect the treatment of patients. In this case, SR technology is applied to improve the spatial resolution of CT images. Jiang [89] proposed an improved SRGAN, which enhances image quality through adapting dilated convolution in the generator module and deleting BN layer in the Residual Block (RB). Li [90] studied the method of combining compressed sensing and similarity constraints, which uses the non-local similarity to search for similar blocks in the entire image, and trains different block dictionaries according to textures. The algorithm has been tested on the Brainweb dataset and has achieved good results, which improves the quality of the restored brain MRI image. The main challenge of the SR algorithm in medical diagnosis is how to ensure the algorithm reconstruction with high accuracy, limited errors, and strong robustness.

D. Visual High-Level Features

SISR reconstruction is a typical low-level feature task, and now many researchers use it to improve high-level vision. Dai [91] proved that SR algorithm can improve different visual tasks, including semantic image segmentation, scene recognition, etc. Bai [2] applied the GAN model to the detection of small-size faces, and the experimental results verified the effectiveness of the image SR method in recovering blurred small faces. Considering the structural correlation between objects of different sizes in the feature space, Li [92] used perceptual GAN to reduce the difference between object sizes to improve small object detection.

The SISR algorithms can break through the limitations of hardware devices and obtain high-resolution images, and is an inexpensive way to improve the resolution of image. These methods rely on paired LR-HR training data, that is, HR images with the LR counterparts obtained via degradation. However, in practical applications, such paired LR-HR training data are often not available. In addition, the distribution of the real-world data don't necessarily be the same as the LR images obtained using a specific degradation method [93]. Therefore, the SISR model trained under ideal conditions may not find satisfied results in practical scenarios. In 2020, Guo [93] has proved that the dual regression network is valid for the SR tasks of mismatched LR-HR real-world data. Therefore, how to actively explore and improve the relevant network structures in the future to meet the unpaired data in real-world becomes an important challenge.

VI. Future Research Trends

The SISR method only needs a single degraded image to restore the HR image, which is actually an ill-posed problem. Deep learning methods can adaptively learn deep features from numerous training sets. With the rapid development of machine learning and artificial intelligence, the deep neural network based image SR reconstruction methods have made great progress, but there still exists some problems to be addressed regarding the difficulties of SISR. This section points out some future research directions of SR techniques.

Reasonable evaluation criteria of SISR: In some CNN-based methods, we find that images with higher PSNR and SSIM values are too smooth and have lower perceived quality [94]. However, the PSNR and SSIM of SRGAN are slightly lower than other algorithms, but the visual perception of the reconstructed image is promising. In this case, there is no unified evaluation standard for actual image super-resolution research [65], we need to explore more accurate evaluation criteria.

Combination of specific tasks and SR framework: High-resolution images provide more detailed information, and many researchers also used them to improve the performance other high-level visual tasks, such as face recognition [3], small target detection [2]. It is therefore a very promising direction to study how to better combine SR techniques with other tasks to deal with specific problems.

Blind SR reconstruction: In a real-life scene, HR images are obtained from a series of LR images, but the LR images are degraded due to unknown blur, noise and down-sampling. Since most of the existing researches use fixed degradation model, the degradation method in actual application scenarios is unknown. Most existing super-resolution algorithms have different perceived quality when the degradation type is unknown. Thus, it is important to know how to apply SR algorithms to achieve better blind super-resolution reconstruction results. In 2019, CVPR organized the challenge of real image super-resolution, which promoted the development of this field [95].

SISR with lightweight architecture: Although the deep SISR models can achieve high reconstruction accuracy, huge amounts of parameters and heavy computational burdens make it difficult to apply in actual scenarios [54]. To solve this problem, we need to simplify the existing SISR deep models to remain the performance along with less computational load. Therefore, it is necessary to further study how to simplify the deep model while speed up the SISR process.

VII. Conclusion

This paper provides an intensive review of the existing research outcome of deep learning based SISR techniques, and summarizes the state-of-the-art methods into two categories including convolutional neural networks and generative adversarial networks based on their internal structure. Based on the experimental data and the characteristics of each network, it can be concluded that the accuracy of deep learning based SISR techniques can be improved by optimizing the network structure and integrating plug-in modules. In addition, this paper further describes the applications of deep learning based SISR techniques in various practical fields. The information and insights provided in this paper will serve as useful reference for relevant researchers.

Conflict of Interest

The authors declare no conflict of interest.

Author Contributions

Ying Liu and Yangge Qiao conducted the research; Ying Liu, Yangge Qiao, Yu Hao, and Sheikh Faisal Rashid revised and edited the paper together; Ying Liu and Fuping Wang acquired funding; all authors had approved the final version.

Acknowledgment

This work was supported by the National Natural Science Foundation of China project (No. 61802305).

References

[1] C. You, G. Li, Y. Zhang, X. Zhang, H. Shan, M. Li, *et al.*, "CT super-resolution GAN constrained by the identical, residual, and Cycle learning ensemble (GAN-CIRCLE)," *IEEE Trans. Med. Imaging*, vol. 39, pp. 188-203, Jan. 2020.

[2] Y. Bai, Y. Zhang, M. Ding, and B. Ghanem, "Finding tiny faces in the wild with generative adversarial network," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018, pp. 21-30.

[3] J. Jiang, C. Chen, J. Ma, Z. Wang, Z. Wang, and R. Hu, "SRLSP: A face image super-resolution algorithm using smooth regression with local structure prior," *IEEE Trans. Multimed*, vol. 19, pp. 27-40, Jan. 2017.

[4] Y. Liu, D. H. Sun, F. P. Wang, K. P. Lim, T. K. Chiew, and Y. Lai, "Learning wavelet coefficients for face super-resolution," *Vis. Comput.*, pp. 1-10, July 2020.

[5] Y. Wang, R. Fevig, and R. R. Schultz, "Super-Resolution mosaicking of UAV surveillance video," in *Proc. 15th IEEE International Conference on Image Processing*, San Diego, 2008, pp. 345-348.

[6] W. Yu and M. Zhang, "A mixed particle swarm optimization algorithm's application in image/video super-resolution reconstruction," in *Proc. 2nd International Conference on Image, Vision and Computing*, Chengdu, 2017, pp. 526-530.

[7] F. Kong, W. Guo, Q. Shen, and D. Wang, "Recursive dictionary-based simultaneous orthogonal matching pursuit for sparse unmixing of hyperspectral data," *Trans. Nanjing Univ. Aeronaut*, vol. 34, pp. 456-466, April 2017.

[8] L. Zhang, J. Nie, W. Wei, Y. Zhang, S. Liao, and L. Shao, "Unsupervised adaptation learning for hyperspectral imagery super-resolution," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Seattle, 2020, pp. 3070-3079.

[9] J. L. Harris, "Diffraction and resolving power," *J. Opt. Soc. Am.*, vol. 54, pp. 931-936, July 1964.

[10] K. Zeng and S. F. Ding, "Research progress of image super-resolution reconstruction," *Computer Engineering and Applications*, vol. 53, pp. 29-35, Aug. 2017.

[11] H. Su, J. Zhou, and Z. H. Zhang, "Survey of super-resolution image reconstruction methods," *Zidonghua Xuebao/Acta Autom. Sin.*, vol. 39, pp. 1202-1213, Aug. 2013.

[12] D. Thapa, K. Raahemifar, W. R. Bobier, and V. Lakshminarayanan, "A performance comparison among different super-resolution techniques," *Comput. Electr. Eng.*, vol. 54, pp. 313-329, 2016.

[13] R. G. Keys, "Cubic convolution interpolation for digital image processing," *IEEE Trans. Acoust.*, vol. 29, pp. 1153-1160, Dec. 1981.

[14] Y. J. Lee and J. Yoon, "Nonlinear image upsampling method based on radial basis function interpolation," *IEEE Trans. Image Process*, vol. 19, pp. 2682-2692, Oct. 2010.

[15] Y. Romano, M. Protter, and M. Elad, "Single image interpolation via adaptive nonlocal sparsity-based modeling," *IEEE Trans. Image Process*, vol. 23, pp. 3085-3098, July 2014.

[16] H. Stark and P. Oskoui, "High-Resolution image recovery from image-plane arrays, using convex projections," *J. Opt. Soc. Am.*, vol. 6, pp. 1715-1726, Nov. 1989.

[17] S. Dai, M. Han, Y. Wu, and Y. Gong, "Bilateral back-projection for single image super resolution," in *Proc. IEEE International Conference on Multimedia and Expo*, Beijing, 2007, pp. 1039-1042.

[18] R. Fattal, "Image upsampling via imposed edge statistics," *ACM Trans. Graph*, vol. 26, p. 95, 2007.

[19] X. Gao, K. Zhang, D. Tao, and X. Li, "Image super-resolution with sparse neighbor embedding," *IEEE Trans. Image Process*, vol. 21, pp. 3194-3205, July 2012.

[20] X. Gao, K. Zhang, D. Tao, and X. Li, "Joint learning for single-image super-resolution via a coupled constraint," *IEEE Trans. Image Process*, vol. 21, pp. 469-480, Feb. 2012.

[21] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Trans. Image Process*, vol. 19, pp. 2861-2873, Nov. 2010.

[22] J. Yang, Z. Wang, Z. Lin, S. Cohen, and T. Huang, "Coupled dictionary training for image super-resolution," *IEEE Trans. Image Process*, vol. 21, pp. 3467-3478, Aug. 2012.

[23] K. Zhang, D. Tao, X. Gao, X. Li, and J. Li, "Coarse-to-Fine learning for single-image super-resolution," *IEEE Trans. Neural Networks Learn*, vol. 28, pp. 1109-1122, May 2017.

[24] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a deep convolutional network for image super-resolution," in *Proc. the European Conference on Computer Vision*, Zurich, 2014, pp. 184-199.

[25] J. Kim, J. K. Lee, and K. M. Lee, "Accurate image super-resolution using very deep convolutional networks," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016, pp. 1646-1654.

[26] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-Recursive convolutional network for image super-resolution," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016, pp. 1637-1645.

[27] W. Wang, T. Zhang, and X. Wang, "Survey of single image super-resolution based on deep learning," *Journal of Chinese Computer Systems*, vol. 40, pp. 1891-1896, Sept. 2019.

[28] W. Yang, X. Zhang, Y. Tian, W. Wang, J. H. Xue, and Q. Liao, "Deep learning for single image super-resolution: A brief review," *IEEE Trans. Multimed*, vol. 21, pp. 3106-3121, Dec. 2019.

[29] S. Kaji and S. Kida, "Overview of image-to-image translation by use of deep neural networks: denoising, super-resolution, modality conversion, and reconstruction in medical imaging," *Radiol. Phys. Technol.*, vol. 12, pp. 235-248, April 2019.

[30] Z. Wang, J. Chen, and S. C. Hoi, "Deep learning for image super-resolution: A survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, pp. 1-1, 2020.

[31] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-Resolution through neighbor embedding," in *Proc. the IEEE Conference on Computer Vision and Pattern Recognition*, Washington, 2004, pp. I-I.

[32] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *Proc. 26th IEEE Conference on Computer Vision and Pattern Recognition*, Anchorage, 2008, pp. 1-8.

[33] R. Timofte, V. De, and L. V. Gool, "Anchored neighborhood regression for fast example-based super-resolution," in *Proc. IEEE International Conference on Computer Vision*, Sydney, 2013, pp. 1920-1927.

[34] R. Timofte, V. De, and L. V. Gool, "A+: Adjusted anchored neighborhood regression for fast super-resolution," in *Proc. Asia Conference on Computer Vision*, Singapore, 2014, pp. 111-126.

[35] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, San Diego, 2005, pp. 886-893.

[36] D. L. Donoho, "Compressed sensing," *IEEE Trans. Inf. Theory*, vol. 52, pp. 1289-1306, April 2006.

[37] G. E. Hinton, S. Osindero, and Y. W. Teh, "A fast learning algorithm for deep belief nets," *Neural Comput.*, vol. 18, pp. 1527-1554, July 2006.

[38] W. Wang and Y. Yang, "Development of convolutional neural network and its application in image classification: A survey," *Opt. Eng.*, vol. 58, p. 1, April 2019.

[39] W. Wang, C. Tang, X. Wang, Y. Lou, and J. Li, "Image object recognition via deep feature-based adaptive joint sparse representation," *Comput. Intell. Neurosci.*, pp. 1-9, Feb. 2019.

[40] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016, pp. 770-778.

[41] K. Zhang, W. Zuo, Y. Chen, D. Meng, and L. Zhang, "Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising," *IEEE Trans. Image Process,* vol. 26, pp. 3142-3155, July 2017.

[42] K. Zhang, W. Zuo, S. Gu, and L. Zhang, "Learning deep CNN denoiser prior for image restoration," in *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017, pp. 2808-2817.

[43] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *Proc. the European Conference on Computer Vision*, Amsterdam, 2016, pp. 391-407.

[44] W. Shi, J. Caballero, F. Huszar, J. Totz, A. P. Aitken, R. Bishop, *et al*., "Real-Time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Las Vegas, 2016, pp. 1874-1883.

[45] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017, pp. 2790-2798.

[46] C. Y. Lee, S. Xie, P. W. Gallagher, Z. Zhang, and Z. Tu, "Deeply-Supervised nets," *Proceedings of Machine Learning Research*, vol. 38, pp. 562-570, 2015.

[47] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Boston, 2015, pp. 3431-3440.

[48] Z. Li, Q. Li, W. Wu, J. Yang, Z. Li, and X. Yang, "Deep recursive up-down sampling networks for single image super-resolution," *Neurocomputing*, vol. 398, pp. 377-388, 2019.

[49] D. Lin, G. Xu, Q. Xu, Y. Wang, X. Sun, and K. Fu, "SCRSR: An efficient recursive convolutional neural network for fast and accurate image super-resolution," *Neurocomputing*, pp. 399-407, 2019.

[50] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu, "Feedback network for image super-resolution," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019, pp. 3862-3871.

[51] G. Huang, Z. Liu, L. V. D. Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017, pp. 2261-2269.

[52] Y. Tai, J. Yang, X. Liu, and C. Xu, "MemNet: A persistent memory network for image restoration," in *Proc. the IEEE International Conference on Computer Vision*, Venice, 2017, pp. 4549-4557.

[53] T. Tong, G. Li, X. Liu, and Q. Gao, "Image super-resolution using dense skip connections," in *Proc. the IEEE International Conference on Computer Vision*, Venice, 2017, pp. 4809-4817.

[54] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu, "Residual dense network for image super-resolution," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018, pp. 2472-2481.

[55] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018, pp. 1664-1673.

[56] J. Liu, W. Zhang, Y. Tang, J. Tang, and G. Wu, "Residual feature aggregation network for image super-resolution," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, 2020, pp. 2356-2365.

[57] W. S. Lai, J. Huang, N. Ahuja, and M. H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017, pp. 5835-5843.

[58] Z. Wang, D. Liu, J. Yang, W. Han, and T. Huang, "Deep networks for image super-resolution with sparse prior," in *Proc. the IEEE International Conference on Computer Vision*, Santiago, 2015, pp. 370-378.

[59] K. Gregor and Y. Lecun, "Learning fast approximations of sparse coding," in *Proc. 27th International Conference on Machine Learning*, Haifa, 2010, pp. 399-406.

[60] T. Dai, J. Cai, Y. Zhang, S. T. Xia, and L. Zhang, "Second-Order attention network for single image super-resolution," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019, pp. 11057-11066.

[61] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu, "Image super-resolution using very deep residual channel attention networks," in *Proc. the European Conference on Computer Vision*, Munich, 2018, pp. 294-310.

[62] I. J. Goodfellow, J. Pougetabadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, *et al*., "Generative adversarial nets," in *Proc. the Advances in Neural Information Processing Systems*, Montreal, 2014, pp. 2672-2680.

[63] L. J. Ratliff, S. A. Burden, and S. S. Sastry, "Characterization and computation of local Nash equilibria in continuous games," in *Proc. 51st Annual Allerton Conference on Communication, Control, and Computing*, Monticello, 2013, pp. 917-924.

[64] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, *et al*., "Photo-Realistic single image super-resolution using a generative adversarial network," in *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017, pp. 105-114.

[65] J. Bruna, P. Sprechmann, and Y. LeCun, "Super-Resolution with deep convolutional sufficient statistics," in *Proc. 4th International Conference on Learning Representations*, San Juan, 2016.

[66] J. Johnson, A. Alahi, and F. F. Li, "Perceptual losses for real-time style transfer and super-resolution," in *Proc. the European Conference on Computer Vision, Amsterdam*, 2016, pp. 694-711.

[67] H. Wang, W. Wu, Y. Su, Y. Duan, and P. Wang, "Image super-resolution using a improved generative adversarial network," in *Proc. IEEE 9th International Conference on Electronics Information and Emergency Communication*, Beijing, 2019, pp. 312-315.

[68] X. Wang, K. Yu, S. Wu, J. Gu, Y. Liu, C. Dong, *et al*., "ESRGAN: Enhanced super-resolution generative adversarial networks," in *Proc. the European Conference on Computer Vision*, Munich, 2019, pp. 63-79.

[69] X. S. Wang, J. Chao, and Y. H. Cheng, "Image super-resolution reconstruction based on self-attention GAN," *Control and Decision*, pp. 1-9, 2020.

[70] T. Xu, P. Zhang, Q. Huang, H. Zhang, Z. Gan, X. Huang, *et al.*, "AttnGAN: Fine-Grained text to image generation with attentional generative adversarial networks," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Salt Lake City, 2018, pp. 1316-1324.

[71] J. Gao, Y. Guo, and M. Yin, "Restricted Boltzmann machine approach to couple dictionary training for image super-resolution," in *Proc. IEEE International Conference on Image Processing*, Salt Lake City, 2018, pp. 1316-1324.

[72] T. Nakashika, T. Takiguchi, and Y. Ariki, "High-Frequency restoration using deep belief nets for super-resolution," in *Proc. International Conference on Signal-Image Technology and Internet-Based Systems*, Kyoto, 2013, pp. 38-42.

[73] C. Chen, Z. Xiong, X. Tian, Z. J. Zha, and F. Wu, "Camera lens super-resolution," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019, pp. 1652-1660.

[74] X. Zhang, Q. Chen, R. Ng, and V Koltun, "Zoom to learn, learn to zoom," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Long Beach, 2019, pp. 3757-3765.

[75] J. Cai, H. Zeng, H. Yong, Z. Cao, and L. Zhang, "Toward real-world single image super-resolution: A new benchmark and a new model," in *Proc. the IEEE International Conference on Computer Vision*, Seoul, 2019, pp. 3086-3095.

[76] J. Deng, W. Dong, R. Socher, L. Li, K. Li, and F. Li, "ImageNet: A large-scale hierarchical image database," in *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, Miami, 2009, pp. 248-255.

[77] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *Proc. International Conference on Curves and Surfaces*, Avignon, 2010, pp. 711-730.

[78] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. A. Morel, "Low complexity single image super resolution based on nonnegative neighbor embedding," in *Proc. the 23rd British Machine Vision Conference*, 2012.

[79] D. Martin, C. Fowlkes, D. Tal, and J. Malik, "A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics," in *Proc. Eighth IEEE International Conference on Computer Vision*, Vancouver, 2001, pp. 416-423.

[80] J. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Boston, 2015, pp. 5197-5206.

[81] T. Y. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, *et al*., "Microsoft COCO: Common objects in context," in *Proc. European Conference on Computer Vision*, Zurich, 2014, pp. 740-755.

[82] E. Agustsson and R. Timofte, "NTIRE 2017 challenge on single image super-resolution: Dataset and study," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, 2017, pp. 1122-1131.

[83] S. Vishnukumar and M. Wilscy, "Super-Resolution for remote sensing images using content adaptive detail enhanced self examples," in *Proc. the IEEE International Conference on Circuit, Power and Computing Technologies*, Nagercoil, 2016, pp. 1-5.

[84] Z. Pan, W. Ma, J. Guo, and B. Lei, "Super-Resolution of single remote sensing image based on residual dense back projection networks," *IEEE Trans. Geosci. Remote Sens*, vol. 57, pp. 7918-7933, Oct. 2019.

[85] P. Wang, H. Zhang, F. Zhou, and Z. Jiang, "Unsupervised remote sensing image super-resolution using cycle CNN," in *Proc. the International Geoscience and Remote Sensing Symposium*, Yokohama, 2019, pp. 3117-3120.

[86] M. A. Talab, S. Awang, and S. A. M. Najim, "Super-Low resolution face recognition using integrated Efficient Sub-Pixel Convolutional Neural Network (ESPCN) and Convolutional Neural Network (CNN)," in *Proc. IEEE International Conference on Automatic Control and Intelligent Systems*, Selangor, 2019, pp. 331-335.

[87] Y. Liu, Z. L. Dong, T. K. Chiew, K. P. Lim, and N. Ling, "A densely connected face super-resolution network based on attention mechanism," in *Proc. 15th IEEE Conference on Industrial Electronics and Applications*, Kristiansand, 2020, pp. 148-152.

[88] X. Zhong, X. Qu, and C. Chen, "High-Quality face image super-resolution based on generative adversarial networks," in *Proc. IEEE 4th Advanced Information Technology, Electronic and Automation Control Conference*, Chengdu, 2019, pp. 1178-1182.

[89] X. Jiang, Y. Xu, P. Wei, and Z. Zhou, "CT image super resolution based on improved SRGAN," in *Proc. 5th International Conference on Computer and Communication Systems*, Shanghai, 2020, pp. 363-367.

[90] Y. Li, B. Song, J. Guo, X. Du, and M. Guizani, "Super-Resolution of brain MRI images using overcomplete dictionaries and nonlocal similarity," *IEEE Access*, vol. 7, pp. 25897-25907, 2019.

[91] D. Dai, Y. Wang, Y. Chen, and L. V. Gool, "Is image super-resolution helpful for other vision tasks?" in *Proc. IEEE Winter Conference on Applications of Computer Vision*, Lake Placid, 2016, pp. 1-9.

[92] J. Li, X. Liang, Y. Wei, T. Xu, J. Feng, and S. Yan, "Perceptual generative adversarial networks for small object detection," in *Proc. 30th IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, 2017, pp. 1951-1959.

[93] Y. Guo, J. Chen, J. Wang, Q. Chen, J. Cao, Z. Deng, *et al*., "Closed-Loop matters: Dual regression networks for single image super-resolution," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Seattle, 2020, pp. 5406-5415.

[94] Y. Blau, R. Mechrez, R. Timofte, T. Michaeli, and L. Zelnik-Manor, "The 2018 PIRM challenge on perceptual image super-resolution," in *Proc. European Conference on Computer Vision*, Munich, 2018, pp. 334-355.

[95] J. Cai, S. Gu, R. Timofte, L. Zhang, X. Liu, Y. Ding, *et al*., "NTIRE 2019 challenge on real image super-resolution: Methods and results," in *Proc. the IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops*, Long Beach, 2019, pp. 2211-2223.

Ying Liu received her B.Sc. and M.Sc. degree in Information Engineering from Xidian University, China, in 1993 and 1996 respectively. After 2 years' serving as associate lecturer in the same Dept, she joined National University of Singapore and received M.Eng. degree in 2000. She then worked as a Research Engineer in Nanyang Technological University, Singapore. In 2007, she obtained her Ph.D. degree from Monash University, Australia. From 2007 to 2010, she worked as a lecturer in Republic Polytechnic in Singapore. Since 2011, she joined Xi'an University of Posts and Telecommunications (XUPT) in China. She is currently a full professor in the School of Communication and Information Engineering in XUPT. She serves as the Director of Center for Image and Information Processing in XUPT Her research activities focus on pattern recognition and image processing.

Yangge Qiao was born in 1997. she is studying for her Master degree in the School of Communication and Information Engineering, Xi'an University of Posts and Telecommunications. Her main research interests are image super-resolution and deep learning.

Yu Hao received Ph.D. degree in computing and engineering from the University of Huddersfield, UK in 2019 and the M.Sc. degree in computer science from the Wichita State University, USA in 2011. Currently, he is a lecturer in School of Communications and Information Engineering, Xi'an University of Posts and Telecommunications, China. His research focuses on intelligent video processing.

Fuping Wang received the B.Eng. degree and the Ph.D. degree in signal and information processing from Xidian University, Xi'an, China, in 2011 and 2017, respectively. He is currently a Lecturer with the Xi'an University of Posts and Telecommunications. His research interests include pattern recognition and image processing.

Sheikh Faisal Rashid is an assistant professor at the Department of Computer Science, University of Engineering and Technology (UET) Lahore. He obtained Doctor of Engineering (Dr.-Ing.) from the University of Kaiserslautern Germany. During his PhD, he worked at Siemens Mobility, Logistics and Postal Automation division Konstanz, Germany and at the German Research Center for Artificial Intelligence (DFKI) Kaiserslautern, Germany. In addition, he is a guest research scientist at DFKI Kaiserslautern, consultant and director of Artificial Intelligence Research Lab (AIRL) at Al-Khwarizmi Institute of Computer Science (KICS) UET Lahore. He is also serving as a vice president of Pakistan Pattern Recognition Society (PPRS), which is officially affiliated with the International Association of Pattern Recognition (IAPR). His area of interest includes: artificial intelligence, deep learning, thermal imaging, document image analysis, OCR, natural language processing and information extraction.