Deep learning methods in real-time image super-resolution: a survey

Xiaofang Li[2] · Yirui Wu[1] · Wen Zhang[1] · Ruichao Wang[3] · Feng Hou[4]

Abstract

Super-resolution is generally defined as a process to obtain high-resolution images form inputs of low-resolution observations, which has attracted quantity of attention from researchers of image-processing community. In this paper, we aim to analyze, compare, and contrast technical problems, methods, and the performance of super-resolution research, especially real-time super-resolution methods based on deep learning structures. Specifically, we first summarize fundamental problems, perform algorithm categorization, and analyze possible application scenarios that should be considered. Since increasing attention has been drawn in utilizing convolutional neural networks (CNN) or generative adversarial networks (GAN) to predict high-frequency details lost in low- resolution images, we provide a general overview on background technologies and pay special attention to super-resolution methods built on deep learning architectures for real-time super-resolution, which not only produce desirable reconstruction results, but also enlarge possible application scenarios of super resolution to systems like cell phones, drones, and embedding systems. Afterwards, benchmark datasets with descriptions are enumerated, and performance of most representative super-resolution approaches is provided to offer a fair and comparative view on performance of current approaches. Finally, we conclude the paper and suggest ways to improve usage of deep learning methods on real-time image super-resolution.

Keywords Image super-resolution · Real-time processing · Deep learning · Convolutional neural network · Generative adversarial network

1 Introduction

In image processing area, image generally describes more visual details with higher resolution. To better understand semantic meanings of real-world images, it is an essential task for researchers to provide high-resolution (HR) images with sharp and clear object boundary or rich visual descriptions. However, obtaining HR images by possible hardware-based approaches is difficult and expensive [74]. For example, one of the possible methods, i.e., decreasing the pixel size would decrease the amount of light achieved by sensors, results in shot noise and sensitivity to diffraction effects. Another possible way, i.e., increasing sensor size, slows down the charge transfer rate and greatly increases the cost of image systems. Therefore, it is more favorable to apply algorithmic-based methods to pursue HR images from low-resolution (LR) resource images, rather than utilizing hardware-based solutions.

Algorithmic-based methods are usually named as super-resolution (SR) methods, which intend to reconstruct a HR image from input of LR observations captured in the same

✉ Yirui Wu
  wuyirui@hhu.edu.cn

  Xiaofang Li
  lixf@czu.cn

  Wen Zhang
  zw10207199@163.com

  Ruichao Wang
  rachel@ucd.ie

  Feng Hou
  f.hou@massey.ac.nz

1   College of Computer and Information, Hohai University, Nanjing, China

2   School of Computer Science and Information Engineering, Changzhou Institute of Technology, Changzhou, China

3   School of Computer Science, University College Dublin, Dublin, Ireland

4   School of Natural and Computational Science, Massey University, Palmerston North, New Zealand

scene. It has been proved that resolution raising via SR methods can largely increase the amount of available information and thus leads to an accurate and robust vision-based machine learning system [70, 75]. Therefore, SR methods have gained great success in multiple domains, such as aerial imaging [138, 139], medical image processing [47, 118], automated mosaicking [13, 54], compressed image/video enhancement [11, 62], action recognition [73, 89], pose estimation [43, 44], face [18, 132], iris [4, 71], fingerprint [61, 101] and gait recognition [3, 140], scene text image improvement and reading [121, 122], and so on.

For the past two decades, numerous methods have been proposed to perform SR task. One of the most classical taxonomy ways to category SR methods relies on the number of LR images involved: single image or multiple images based SR methods [74]. Essentially, there exist fundamental differences on thoughts to solve SR problems in terms of number of LR images. Specifically, single-image based SR methods tend to hallucinate missing image details with relationship learned from training datasets. Multiple-image based SR methods generally utilize global/local geometric or photometric relation between multiple LR images to reconstruct HR images.

We must address that several surveys [39, 77, 107, 135] have been conducted on single-image, multi-based based SR or both, which is the basis to build up our paper. For example, Nasrollahi and Moeslund [74] provide a comprehensive overview of most of published and related papers up to 2012, which has offered plenty of reading resources to learn fundamentals and developing history for beginners in this field. Yue et al. [135] reviews SR methods and applications based on machine learning techniques up to 2016. Most recently, Hayat [41] focuses on the deep learning-based progress of three aspects in multimedia, i.e., image, video and multi-dimensions. Nguyen et al. [75] comprehensively survey the super-resolution approaches proposed for a special application domain, i.e., biometrics including fingerprint, iris, gait, face, etc. Built on but differing from these survey papers, our paper focuses on the review of deep learning-based SR methods with real-time response, which could enlarge possible application scenarios of SR methods to embedding systems, cell phone, drones and so on by effective performance and low computation cost.

In this paper, we attempt to establish a baseline for future work by providing a comprehensive literature survey of deep learning methods in real-time SR research. Incremental advances or thoughts to the state of the art thus could be made or inspired on our provided baseline. Specifically, we first summarize the fundamental problems, categorize existing methods and review the applications to the benefit of beginners. We then briefly introduce significant progress in deep learning methods. By involving deep learning structures to solve problem of SR, quantity of deep learning-based SR methods are proposed. We highlight deep learning methods for real-time SR by analyzing, comparing, and concluding their core ideas. Afterwards, performance of the most representative deep learning approaches on benchmark datasets is compared and analyzed. Finally, we conclude to bring out open questions and future trends on improving deep learning methods for real-time SR performance.

The rest of the paper is organized as follows: Sect. 2 presents background concepts on super-resolution and deep learning methods for the benefit of beginners. Section 3 highlights and analyzes deep learning-based methods for real-time SR. Benchmark datasets, evaluation methods and performance comparisons of quantity of methods are presented in Sect. 4. Finally, the paper is summarized with discussions about remaining problems and future trends in Sect. 5.

2 Super-resolution formation

To offer an overall view on deep-learning based SR methods, it is useful to provide background information about the underlying problems. In this section, we first introduce mathematic definition of SR problem by reviewing the relevant imaging model. Then, we categorize the existing methods, which helps readers not only know history of SR, but also comprehend deep learning-based SR methods by comparing them with other types of SR methods. Afterwards, we review SR applications in different domains and focus on applications requesting real-time performance, which should be considered during the design of real-time SR methods. Finally, we explain fundamental thoughts of CNN and GAN structures for readers' convenience, since this paper mainly focuses on real-time methods for SR with CNN and GAN structures.

2.1 Definition of super-resolution problem

Although SR methods give rise to many applications, the fundamental goal of SR is to collect missing detail information by reconstructing super-resolved images from LR observations. In the literature, SR can be regarded as a heavily ill-posed problem, since the informative information represented in LR images is often insufficient to complete the task of reconstruction. Therefore, SR methods need to complete three main tasks: up-sampling of LR images to increase image resolution, removing artifacts including blur and noise during SR process, and registration or fusion of multiple input LR images for a better representation of target HR image.

Based on these three tasks, we first describe the most common imaging model [51, 95] to generate LR images in

the simplest case, where this process can be modeled linearly as

$$L_i(x', y') = \frac{1}{q^2} \sum_{x=qm}^{(q+1)m-1} \sum_{y=qm}^{(q+1)n-1} R(x, y), \tag{1}$$

where L_i is an observed LR image, R is the original HR scene, q is a decimation factor or sub-sampling parameter which is assumed to be equal for both x and y directions, x and y are the coordinates of the HR image, and m and n of the LR images. The imaging model in Eq. 1 states that an LR observed image has been obtained by averaging the HR intensities over a neighborhood of q^2 pixels.

This model becomes more realistic when the other parameters involved in the imaging process are taken into account, which is shown in Fig. 1, including stepwise representations for warping, down-sampling, noise adding, and blurring. It is noted that SR methods could be generally considered as an inverse processing workflow to generate HR images based on LR images. Supposing the real-world image R is captured by n different located cameras to form multi-view images, the generated LR L_i is constructed by the following formulas:

$$\begin{cases} L_i(x', y') = (\mu_i(\delta_i(\omega_i R(x, y))) + \eta_i(x', y')) \\ L_j(x', y') = \rho_{i,j}(L_i(\tau_{i,j}(x', y'))), \end{cases} \tag{2}$$

where x, y and x', y' refer to coordinates of one pixel in real-world image R and generated LR images L_i, L_j, respectively, $\omega_i()$ is a warping function determined by locations and rotations of multiple cameras, $\delta_i()$ is a own-sampling function, $\mu_i()$ is a blurring function, $\eta_i()$ is an additive noise, $\tau_{i,j}()$ is a coordinate transformation function including horizontal

shift, vertical shift and rotation, and $\rho_{i,j}()$ is a pixel intensity transformation function. Specifically, if the LR image of L_i is displaced from the HR scene of R by a translational vector as (a, b) and a rotational angle of θ, the warping function in homogeneous coordinates could be represented as

$$\omega\left(\begin{bmatrix} x \\ y \\ 1 \end{bmatrix}\right) = \left(\begin{bmatrix} 1 & 0 & a \\ 0 & 1 & b \\ 0 & 0 & 1 \end{bmatrix} \times \begin{bmatrix} \cos\theta & \sin\theta & 0 \\ -\sin\theta & \cos\theta & 0 \\ 0 & 0 & 1 \end{bmatrix}\right)^{-1} \begin{bmatrix} m \\ n \\ 1 \end{bmatrix} \tag{3}$$

In matrix form, Eq. 2 can be written as follows:

$$\begin{cases} L_i = AR + \eta \\ L_j = \rho L_i, \end{cases} \tag{4}$$

in which A stands for the above-mentioned degradation factors. This imaging model has been used in many SR works. It is noted that the first line in Eq. 2 refers to warping, down-sampling, noise adding, and blurring on a single HR image and corresponds to inverse operations of task 1 and 2; meanwhile, the second line refers to multi-view transformation on two different LR image and corresponds to inverse operation of task 3. Moreover, we often define process of image registration [48, 80, 102] including functions $\tau_{i,j}()$ and $\rho_{i,j}()$ to handle multi-view images during SR. Essentially, the process is to geometrically align multiple images of the same scene onto a common reference plain, where images can be captured at different times and from different views, or by multiple sensors.
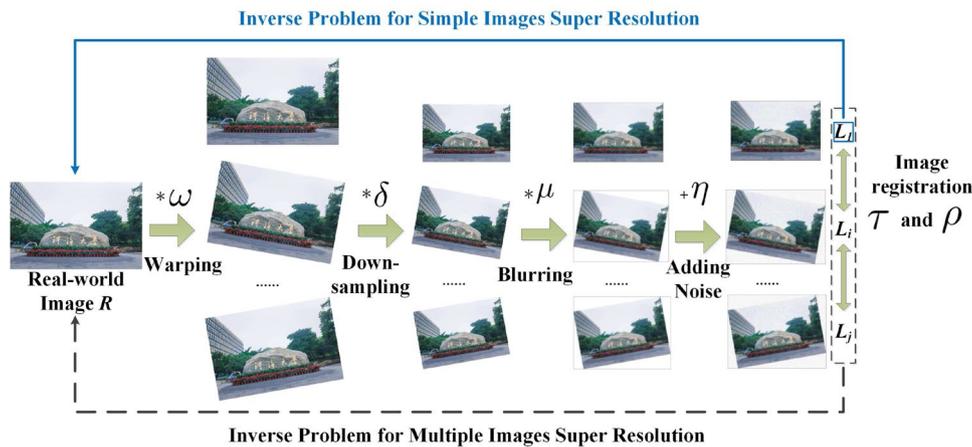

Fig. 1 Illumination of the most common used imaging model, where the forward pipeline represents the process of generating multiple LR images $\{L_i, \ldots, L_n\}$ from a real-world image R by warping (ω), down-sampling (δ), blurring (μ), and adding noise (η), and the back-ward pipeline refers to a basic method of reverse super-resolution to reconstruct the real-world or HR image from one input LR image or multiple LR images by de-blurring, up-sampling, aligning, and image registration (τ and ρ)

2.2 Categorization of super-resolution methods

Following the description of imaging model and basic categorization in [74], we further categorize the existing SR methods into single image SR (SISR) or multiple images SR (MISR) as shown in Fig. 2. In the following discussions, we would follow Fig. 2 to state the categorization of current SR methods.

Being a highly ill-posed problem without sufficient information about the original image sets, early SISR methods tend to utilize analytical interpolation to reconstruct HR images. Several famous interpolation-based SISR methods can be listed as linear, bicubic, cubic splines interpolation methods, Lanczos up-sampling [29], New Edge Directed Interpolation (NEDI) [63] and so on. These methods are very simple and effective ways in smooth parts with real-time performance. However, simple rule in interpolation brings in overly smooth and blurring details, which harm the visual effect of image discontinuities like edges, boundaries, and corners. Hence more sophisticated insight to recover image details via reasonable SR ways is required.

Besides interpolation-based methods, researchers have proposed another two categories of SISR methods, i.e., reconstruction-based and learning-based methods. Reconstruction-based methods suppose there exist certain priors or constraints in the form of distribution, energy function or score function between HR and the original LR images. Therefore, researchers try a variety of methods to perform SR tasks by establishing reconstruction priors like sharpening of edge details [21], regularization [5] or de-convolution [97].

Learning-based methods try to restore missing high-frequency image details by establishing implicit relationship between LR patches and their corresponding HR patches via machine learning models. This category of methods has achieved more and more attention from researchers due to its promising and visually desirable reconstruction results. It is a general idea to enhance SR quality by learning relationship from large quantity of training data. However, applying over data might introduce spurious high frequencies, resulting in noise and blur details. Therefore, it is important to keep balance between size of training data and reconstruction visual effects.

With the development of machine learning technologies, researchers have tried quantity of learning models to solve SR problem. We further classify learning-based methods into five groups based on differences of their core ideas: neighbor embedding methods [7, 15], sparse coding methods [14, 28], self-exemplar methods [34, 46], locally linear regression methods [38, 129], and deep learning methods [26, 52, 67, 104]. In this paper, we focus on utilizing deep learning-based methods to solve SR problem, due to their significant HR reconstruction results. Deep learning-based methods will be comprehensively discussed in Sect. 3. In the following discussions, we will briefly introduce other four groups of learning-based methods for comparisons with deep learning methods.

Neighbor embedding (NE) methods consider that similar local geometries property is shared between LR patches and their corresponding HR patches. Due to the similar local geometry property of LR and HR feature space, patches in the HR feature domain can be computed with a form of a weighted average of local neighbors. After construction weight scheme, the whole SR computation process can share the same weights within LR feature domain. Based on this thought, Chang et al. [15] propose a SR method by applying a typical kind of manifold learning method, i.e., locally linear embedding (LLE) [87] on weight learning. Their proposed method assumes each sample and its neighbors lie on or near a locally linear patch of the manifold, the idea of which has greatly influenced the subsequent coding-based methods in early times.

Sparse coding methods consider image patches as a sparse linear combination of elements, which could be selected from a pre-constructed and sparse enough dictionary. By exploiting reasonable and sparse enough representation for each patch of low-resolution inputs, the process of generating high-resolution outputs can be

Fig. 2 Categorization of the existing super-resolution methods. Among these methods, we pay special attention to deep learning methods
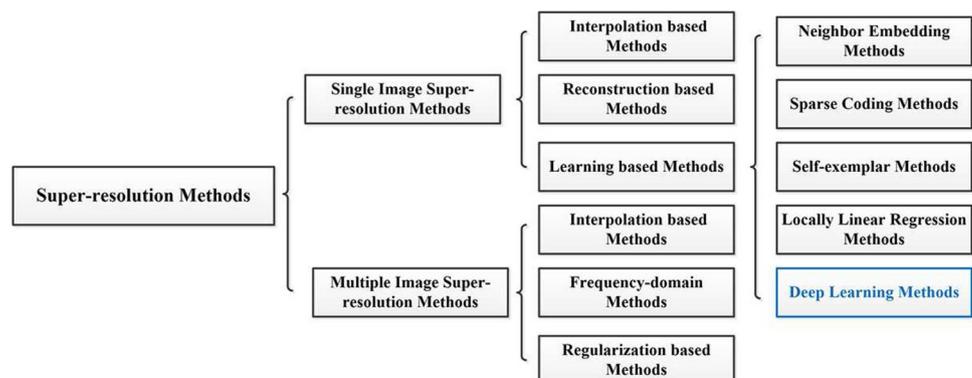
represented as coefficients computing. For example, Yang et al. [130] train a joint dictionary to find a highly sparse and over-complete coefficient matrix, which directly and clearly describes relations between LR image patches and corresponding HR image patches.

However, these sparse coding approaches are generally slow in computing speed, since sparse encoding largely increases memory usage and more elements are thus required to be processed.

Natural images generally have self-similarity property, which inspires researchers to utilize internal similar properties among LR patches to help effective and qualified SR. Based on this core idea of self-exemplar methods, Glasner et al. [34] propose a scale space pyramid, which can be trained with internal class data, i.e., local pathes. Furthermore, self similarity information represented by the trained pyramid can help explicitly describe the mapping between LR and HR pairs. However, high computation cost, especially in building scale space, has prevented the further usage of self-exemplar methods.

The thought of locally linear regression methods sources from Timofte et al. [109], which focuses on solving computational speed shortcomings of sparse coding methods by replacing complex dictionary with many and light-scale ones. Following this trend, researchers try quantity of supervised machine learning techniques to replace heavy sparse coding dictionary, which not only directly learns mapping relationship embedded between LR image domain to HR image domain, but also helps keep low values in computation cost. The most popular supervised machine learning techniques to learn mapping relationship can be listed as anchored neighborhood regression [15, 109, 110], random forest [93, 94] , manifold embedding [79] and so on.

For multiple-image SR, three main types of multi-frame methods can be listed: interpolation-based methods [8, 78], frequency-domain methods [23, 49, 145] and regularization-based methods [6, 106]. For a detailed description on category of MISR, we highly suggest the readers to read [19, 74]. Unlike interpolation-based methods for SISR, MISR consists of registration, interpolation, and deblurring, where extra image regression step reconstructs both local and global geometry based on multiple input LR images. Frequency-domain methods generally utilize DFT, DCT, DWT or other frequency transform methods to discover high-frequency details of HR images. Regularization-based methods are specially designed to perform SR task with conditions like a small number of LR images or heavy blur operators. With either deterministic or stochastic regularization strategy, these methods try to form and incorporate prior knowledge of all the unknown HR images.

2.3 Super-resolution applications

Over the past two decades, there have been numerous SR-related applications, which are the most basic concerns and requirements to improve SR methods. We thus offer some successful application examples of SR to analyze user needs for SR methods.

Medical diagnosis Doctors need various medical images to comprehend healthy messages sent by the inside human body structure. Unfortunately, resolution limitations of medical images do great harm to such precise diagnosis. By reconstructing high-resolution magnetic resonance images (MRI) [86], positron emission tomography (PET) [37] or other key medical images [25, 114], SR methods help maintain geometry structure captured by original 3-D imaging system. For instance, to collect similar images to establish a database [113], researchers apply example-based SR for single frames to create clear enough samples. Specifically, the adopted training database was established with a set of five standard images, including computed tomography (CT) and MRI images from various parts of the human body. Since medical imaging systems are originally designed to acquire multi-view and multiple medical images, the main challenge of SR methods for medical diagnosis lies in the precise reconstruction with less errors and higher robustness.

Text image enhancement and reading Text detection and recognition has been applied in many real-life applications such as iTown, Rosetta and many other smart city developments [72, 131]. These applications encounter problems when performing natural scene semantics understanding or analysis, due to the low resolutions of images captured by cell phone or CCTV cameras, large variations in text rotations or illumination embedded in complex backgrounds with buildings, trees, etc. It is needed to increase the image quality and processing speed of captured text images and by appropriate real-time SR methods, in order to fulfil requests of robust text reading and implementations on embedding systems, cell phone or CCTV cameras [81, 83].

Biometrics Biometrics is defined as a reliable method to automatically identify individual persons based on their behavioral and physiological characteristics, i.e., faces [115, 151], fingerprints [134], and iris images [30]. Even though biometrics system has achieved great success [105], it still faces several challenges. Among these challenges, short distance of image acquisition has become the major one, since it prevents further wide usage of such systems. With proper SR methods, image details like shapes and structural texture can be clearly enhanced. Meanwhile, SR could help maintain the global structure of images. All these advantages can greatly improve the recognition ability in biometrics-related applications.

Remote sensing In the past decades, researchers have applied SR techniques to improve quality of remote sensing

images. After years of efforts, there exist successful and applicable examples in remote sensing area [138]. For example, Skybox Imaging Plan utilizes SR techniques to help provide real-time remote sensing images under a submeter resolution, which shows that SR methods could help improve outputs of remote sensing applications. The main challenge for remote sensing image SR lies in two aspects: (1) how to deal with scene variations in case of temporal differences, and (2) how to modify the existing and successful SR methods to handle large amounts of remote sensing images captured by satellites every day.

Based on former discussions, we can observe the variety of application scenarios of SR methods. Among these applications, medical diagnosis requires to process with less errors and higher robustness, while remote sensing requires to overcome temporal differences and handle massive amounts of input. Besides, both categories of applications have common property that users, i.e., doctors and scientists, could buy computation resource abundant devices to handle SR for edge computing or big data services [35, 127]. In other words, they can easily access better equipments for more satisfying SR results, meanwhile keeping low computing time. For biometrics and text image reading applications, they mainly run on embedding devices and have high request for real-time responses to improve user experience. Therefore, it is a key problem for such applications to keep balance between SR quality and computation cost. In this paper, we would like to emphasize applications like biometrics and text image reading, thus reviewing SR methods for real-time computing and fulfilment of their emergency requests.

2.4 Deep learning background

The most popular structures related to SR can be roughly categorized into the two groups: convolutional neural network (CNN) and generative adversarial networks (GAN). These deep structures have high capability to represent information abundant and distinctive features by self-learning strategies. For comparison, traditional learning structures require humans to observe and design manual features to perform classification tasks. In this subsection, we focus on explanation of fundamental thoughts of CNN and GAN structures.

Convolutional neural network Inspired by the promising classification results achieved by a typical CNN, i.e., AlexNet [55], quantity of trials on structures, learning strategies and applications have been made, some of which can be listed as: VGG [100], Googlenet [103], ResNet [42], R-CNN [33] and so on. Several common types of layers are involved to construct CNN structure: convolutional layers, pooling layers, and fully connected layers.

Convolutional layers are designed to gather information of neighboring pixels. In fact, each pixel is closely associated with neighboring pixels and nearly irrelevant with pixels in long range, which is named as local receptive field. We show the comparison between local connection and full connection adopted by fully connected layer in Fig. 3. Essentially, a convolution kernel can only extract one specific feature in a local sense. Researchers thus design multiple convolution kernels to extract a variety of features from input images. With different feature maps produced by multiple convolution kernels, convolutional layers lead to better understanding of image content.

Unlike convolutional layers, pooling layers are defined without parameters. Pooling layers utilize down-sampling operation to extract features from feature map, which reduce data size without modifications on data characteristics. The resulting abstract feature not only owns generalization ability from feature maps, but also has a certain degree of variety for translation, rotation, and scaling invariance. Pooling layers thus help improve robustness and generalization performance of the whole network.

Fully connected layers locate at the end of CNN, where each neuron is connected to all neurons in the upper layer. With such processing, features locally extracted are globally involved to output the final result. Above all, CNN structure combines abilities provided by convolutional, pooling and fully connected layers, which

Fig. 3 Comparison of reception field and parameter size between locally and fully connected neurons
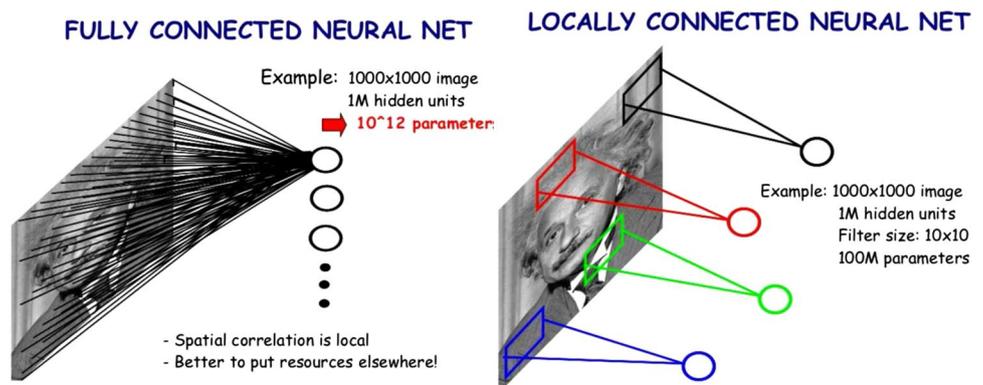
help approximate any continuous function and ensure to perform difficult classification and recognition tasks successfully.

Generative adversarial network As one of the most significant improvements on the research of deep generative models, GAN [36] provides a novel way to learn depth representation of features without large labeled training data. With this power for distribution modeling, GAN is extremely suitable to for unsupervised tasks [69, 148] including image generation [24], image editing [150], and representation learning [84].

The key idea of GAN stems from a two-player game preformed by a generator and a discriminator, where we show the basic structure of a typical GAN for generating hand-written digital images in Fig. 4. Specifically, the discriminator is responsible to judge whether an input image generated by the generator appears natural or could be found with artifacts; meanwhile the generator is to create images with the goal of making discriminator believe the created image is a natural image without any artifacts. After rounds of training, Nash Equilibrium will be achieved, where the trained generator would have the ability to understand inherent and intern representation of real images, thus generating real enough images.

GAN is still developing with great leading steps on structures, training algorithms and so on. With the development of GAN structure, more related applications have been applied by researchers. However, training GAN for data augmentation is challenging, since the training process can be easily trapped into the mode collapsing problem. Essentially, mode collapsing problem is defined as that where the generator only concentrates on producing samples lying on a few modes, instead of the whole data space [16, 92]. It is noted this problem exists in SR applications with GAN model as well, which will be further explained in Sect. 3.2.

3 Deep learning methods for real-time super-resolution

Despite there exist quantity of review papers on deep learning methods for SR [39, 77, 107, 135], there is a lack of reviewing of methods for real-time SR to the benefit of researchers. In this section, we attempt to survey deep learning literature, including CNN, GAN and other deep learning methods, with the view of real-time super-resolution.

In the following discussion of each subsection, we first introduce several typical methods, which have achieved significant HR reconstruction results but failed to obtain real-time performance. Then, we highlight fast and real-time deep learning methods, where we would carefully explain their core ideas, innovations, algorithm steps and performances. Finally, we conclude current state and discuss future developing trend. Furthermore, it is noted that we fuse SISR and MISR methods in this section, since many deep learning-based SR methods have ability to perform SISR and MISR.

3.1 CNN-based methods in real-time image super-resolution

3.1.1 CNN-based methods for SR

CNN-based SR methods are quite large in amount, due to their impressive HR image reconstruction results. The first work to solve SISR problem by CNN structure is proposed by Dong et al. [26], which constructs a three-layer CNN named as Super-Resolution Convolutional Neural Network (SRCNN) to learn mapping between LR patches and corresponding HR patches. Utilizing a bicubic interpolation for pre-processing, SRCNN optimizes and learns nonlinear mapping in manifold space based on information abundant feature maps produced by convolutional layers. With the high distinguishing power of deep structure, SRCNN

Fig. 4 Structure of a typical GAN model with a generator and a discriminator
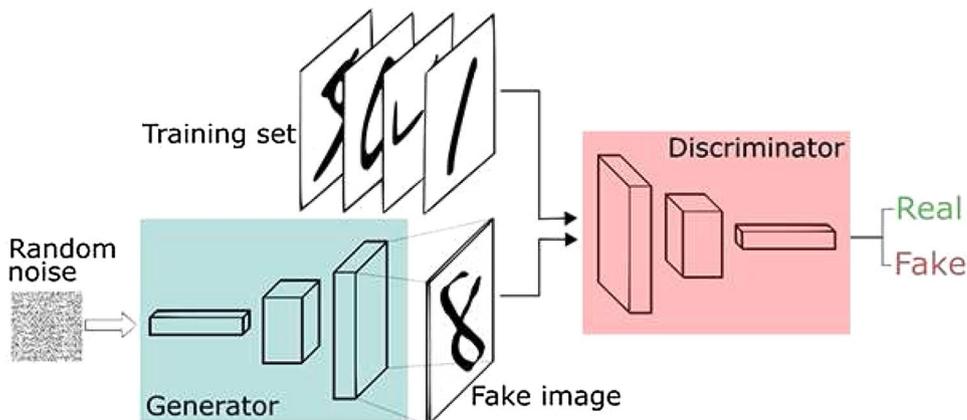
has achieved promising reconstruction results outperforming majority of former methods, such as Self-Ex [46], A+ [110] and Kernel-based learning. Although SRCNN claims efficiency with a lightweight structure, its performance is still far from real-time response due to its time-consuming pre-processing step, i.e., bicubic interpolation.

Essentially, SRCNN is an important work to offer inspirations on utilizing deep structure for SR purposes. After its publication, SRCNN appeared in many other works as a baseline method for comparisons or a basic structure to modify for new learners. More important, its structure with only convolutional layers has greatly affected later CNN-based SR methods, which successfully avoid down-sampling effects brought by pulling or subsampling layers. However, the additional convolutional layers largely increase size of parameters, resulting in more possibility to be overfit with training dataset and be harder to achieve real-time responds. How to keep a balance between desirable reconstruction performance and real-time computing speed thus becomes a major challenge in utilizing deep structures for SR.

With the remarkable success to construct very deep structures achieved by Res-Net, its core idea of residual learning is adopted by researchers to perform SR tasks. Residual learning not only offers capability to construct larger number of layers for better HR reconstruction results, but also reduce difficulty in training process with fast convergence and a small number of epoches. For example, Kim et al. [52] first try residual learning for SR with a novel Very Deep Convolutional Networks (VDSR), which imitates VGG-net structure [100] to build 20 convolutional layers as a a very deep network. Following the trend of applying residual learning on SR, Tai et al. [104] propose Deeply Recursive Residual Network (DRRN), which first utilizes global residual learning to identify branch during inference and then proposes new concept of local residual learning to optimize local residual branch.

Recently, the main focus of CNN-based SR research is to utilize proper technologies for either improved HR reconstruction results or fast computing speed. For example, Deep Back-Projection Networks (DBPN) [40] propose iterative up- and down- sampling layers to form an error feedback scheme, which help transmit projection errors among different layers. With such scheme, they can represent the process of image degradation and super-resolution by simply connecting up- and down-sampling layers, thus improving HR reconstruction results with large scaling factors.

Related to the topic of real-time SR, Lim et al. [65] develop an enhanced deep super-resolution network (EDSR) with performance exceeding current state-of-the-art SISR methods. Their proposed method performs optimization by removing unnecessary modules in conventional residual networks and expands model depth with a stable training procedure. Inspired by the development of attention models,

Zhang et al. [142] adopt an existing channel attention mechanism to construct very deep residual channel attention networks (RCAN). Their proposed residual in residual (RIR) structure is specially designed to bypass abundant low-frequency information for learning of high-frequency information. Hu et al. [45] propose a channel-wise and spatial feature modulation (CSFM) network, which connects feature-modulation memory (FMM) modules with stack connections for transforming low-resolution features to high informative features. Considering that most of CNN-based methods have not fully exploited all the features of the original low-resolution image, Shamsolmoali et al. [96] proposed an effective model based on dilated dense network operations to accelerate deep networks for image SR, which supports the exponential growth of the receptive field parallel by increasing the filter size.

3.1.2 CNN-based methods for real-time SR

CNN-based SR methods have demonstrated remarkable performance in quality of reconstructed HR images, compared with the previous non deep learning based models. However, high computation cost and large computing time

prevent its further practical usage, especially in phones or wearable devices that demand small computing burden and real-time performance. There are thus many trials to accelerate network for real-time performance, in order to enlarge possible application scenarios of CNN-based SR methods.

One of the most famous successful and inspiring trials is advised and performed by Shi et al. [99], who find that utilizing a single filter, usually bicubic interpolation, before reconstruction to up-scale input LR images is suboptimal and time-consuming. They thus prefer to avoid such pre-interpolation operation by utilizing an end-to-end and unified CNN structure named as Efficient Sub-pixel CNN (ESPCN) for SR tasks, which tries to directly learn a upscaling filter, i.e., sub-pixel convolution layer, and integrates it into the structure of CNN network. We show the structure of ESPCN in Fig. 5, where we can notice feature maps to fill up image detail information is extracted in the LR space, rather than performing in HR space by most of the SR methods. Afterwards, extracted feature map in different layers is fed into sub-pixel convolution layer for further processing.

As far as we know, ESPCN is the first CNN-based SR methods with real-time performance, which is reported to perform real-time SR tasks on 1080p videos using a K2 GPU device. Besides, authors report reconstruction results achieved by ESPCN is better than SRCNN by +0.15dB on Set14 dataset images.

Nearly the same time with ESPCN [99], Dong et al. [27] successfully accelerate SRCNN by constructing a compact hourglass-shape CNN structure, named as FSRCNN. As shown in Fig. 6, their structure modification on accelerating
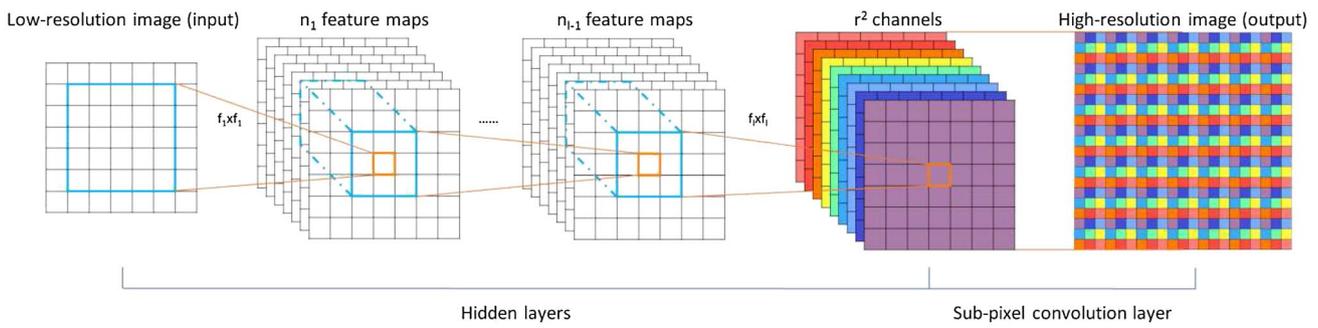
Fig. 5 Network structure of Shi et al. [99] for real-time SR, where they propose sub-pixel convolutional layers to perform up-sampling operations
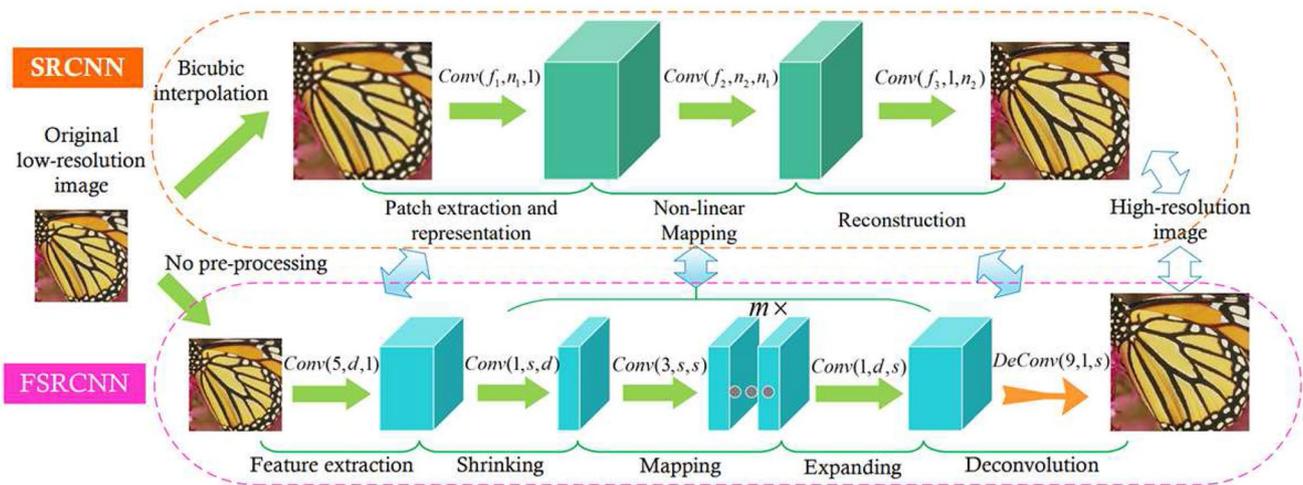

Fig. 6 Network structure of FSRCNN [27], which successfully accelerates SRCNN by several modifications

SRCNN lies in three aspects: (1) they replace bicubic interpolation operation of SRCNN with a de-convolution layer located at the end of CNN network, thus avoiding bicubic pre-interpolation with visible reconstruction artifacts and unnecessary computational cost; (2) they train four convolution layers in a joint optimization manner to complete three tasks during feature extraction, i.e., shrinking, mapping, and expanding. Essentially, structure design of placing mapping layer after shrinking layer will greatly reduce feature dimensions, leading to smaller computation cost; (3) they utilize smaller filter sizes and more mapping layers, in order to achieve desirable reconstruction results and less computing burden at the same time.

FSRCNN is reported to achieve real-time performance (> 24 fps) on test images in all benchmark datasets, which is almost a 40-time improvement than SRCNN in computing speed. Compared with ESPCN [99] which achieves real-time performance on GPU, FSRCNN [27] could process LR images in real-time on a CPU-based platform, which largely expands its possible applicable scenarios.

Following the idea to replace time-consuming up-sampling with partly design of neural network in ESPCN [99] and FSRCNN [27], Yamanaka et al. [128] integrate network in network (NIN) structure [66], i.e., Parallelized 1×1 CNNs, into the whole network as a post processing step for efficient up-sampling operations. Specifically, they first involve deep CNN layers and skip connection layers to extract feature maps by gathering information from both local and global areas. Afterwards, they utilize NIN layers to perform up-sampling operation for reconstruction of HR images. Authors report 10 times lower computation cost achieved by their proposed work than typical deep residual network for SR tasks, thus ensuring real-time performance. However, its simple structure design results in relatively worse reconstruction results, comparing with results achieved by FSRCNN.

To achieve real-time performance, ESPCN [99], FSRCNN [27] and Yamanaka et al. [128] replace bicubic up-sampling operation with sub-pixel convolution layers, deconvolution layers, and NIN layers, respectively. However,

Lai et al. [56] argue that sub-pixel convolution, de-convolution or NIN layer adopts small size networks, which cannot guarantee to describe complicated mappings with their limited network representation capacity. They thus propose the Laplacian Pyramid Super-Resolution Network (LapSRN) to progressively reconstruct sub-band residuals for visually desirable HR images, where we show its structure in Fig. 7. During construction of each pyramid, feature extraction branch first predicts missing high-frequency residuals of HR images based on the input of coarse-resolution feature maps. Afterwards, image reconstruction branch adopts transposed convolutions to perform up-sampling operation, thus generating finer and coarser feature maps as input for next level. Furthermore, they design recursive layers as parameters sharing scheme across and within pyramid levels, which helps reduce the number of parameters in a large amount. With the careful design of pyramid structure, LapSRN claims to adaptively build model with different up-sampling scales, thus reducing computational complexity and achieving real-time computing speed on public testing datasets.

Besides pyramid structure, another solution to improve small network without enough representation ability is constructing neural networks with skip connections, which not only help go deeper of neural networks by preventing gradient loss, but also relieve computation burden with unnecessary computing steps. Tong et al. [112] thus introduce dense skip connections in a very deep neural network for SR tasks, where we show its structure in Fig. 8. In each dense block, we can notice input low-level features and generated high-level features are combined in a reasonable way to boost reconstruction performance. To properly fuse low- and high-level features, dense block structure propagates feature maps generated by each layer into all subsequent layers and designs dense skip connection to allow for deeper structure. Since deep network generally leads to large computation cost, they further integrate de-convolution layers to reduce number of parameters for boosting speed of reconstruction process. By employing algorithm on a platform with a Titan X GPU, their proposed method could achieve an average running time of 36.8ms to reconstruct a single image from B100 dataset, thus guaranteeing real-time SR effects.

Unlike former CNN-based methods for real-time SR to utilize up-sampling operation and residual learning, Johnson et al. [50] tend to consider SR as an image transformation problem between input LR and output HR images. They thus model SR as a global optimization problem under a given
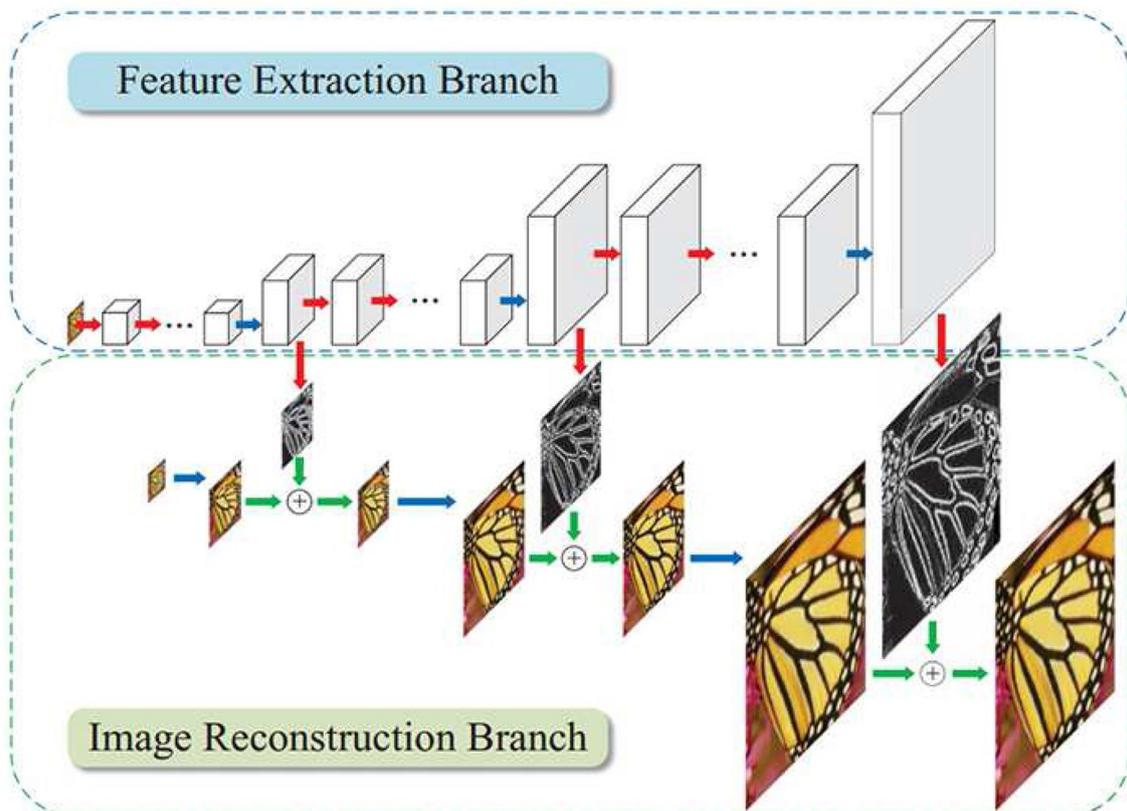

Fig. 7 Network structure of Lai et al. [56], where a careful design with pyramid structure and two working branches is adopted to progressively reconstruct sub-band residuals of HR images
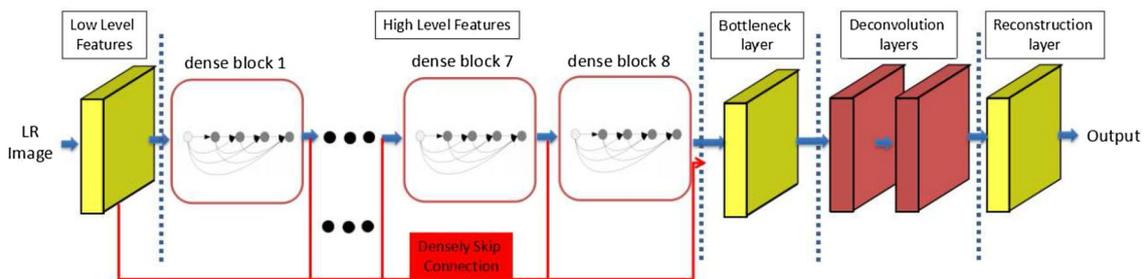
Fig. 8 Network structure of Tong et al. [112], which represents that all levels of features are combined via skip connections as input to reconstruct HR images

objective function, which is formed as the sum of per-pixel loss between HR and ground-truth images and perceptual loss based on high-level features extracted from pre-trained neural networks. Such optimization problem could be solved by Gatys et al. [32] in real-time, thus achieving similar qualitative results but with three times faster running speed than SRCNN and other comparative methods. Essentially, their usage of perceptual loss functions to train feed-forward networks for SR problems and other image transformation tasks is novel and offers inspirations to many other similar works.

Inspired by A+ [110] and ARN [1] which build mapping between input LR space and output HR space by learning proper local linear functions, Li et al. [64] construct convolutional anchored regression network (CARN) to learn representative mapping function for fast and accurate SR. Different from A+ and ARN, regression blocks inside CARN are built on the basis of automatically extracting feature map by convolutional filters, other than limited and hand-crafted features used by A+ and ARN. Besides, CARN transforms all key concepts during SR operation, such as feature extracting, anchor detection, and regressor construction, into convolution operations with different parameters, so that users can jointly optimize all steps in an end-to-end manner. Such end-to-end design relieves the burden of complicated step-wise optimization and thus decreases running time without lose of accuracy. CARN is reported to achieve 10 times lower computation cost than SRCNN, thus reaching real-time performance on most platforms.

Due to the requirement of heavy computation, deep learning methods cannot be easily applied to realworld applications. To address this issue, Ahn et al. [2] propose an accurate and lightweight deep network for image super-resolution. Specifically, they design an architecture that implements cascading connections starting from each intermediary layer to the others upon a residual network, where we show its special structure design in Fig. 9. Such connections are made on both the local and global levels, which allows for the efficient flow of information and gradient.

To explore feature correlations of intermediate layers rather than focus on wider or deeper architecture design,

Dai et al. [22] explore to enhance the representational power of CNNs for more powerful feature expression and feature correlation learning. Specifically, they propose a second-order attention network (SAN) to adaptively rescale the channel-wise features by using second-order feature statistics for more discriminative representations. Furthermore, they present a non-locally enhanced residual group (NLRG) structure, which not only incorporates non-local operations to capture long-distance spatial contextual information, but also contains repeated local-source residual attention groups (LSRAG) to learn increasingly abstract feature representations. All these improvements in structure have been shown in Fig. 10.

To solve the problem of lack of realistic training data and information loss of the model input, Xu et al. [124] propose a new pipeline to generate realistic training data by simulating the imaging process of digital cameras. To further remedy the information loss of the input, they develop a dual convolutional neural network to exploit the originally captured radiance information in raw images. They gain favorable reconstruction results both quantitatively and qualitatively, and their proposed method is declared to enable super-resolution for real captured images.

After precise description on quantity of works about CNN-based methods for real-time SR, we can conclude their general structure. Specifically, in order to widen the receptive field, increasing network depth is one way adopted by quantity of methods, which is to construct network by a convolutional layer with filter size larger than a 1×1 or a pooling layer that reduces the dimension of intermediate representation. However, such designing may have a major drawback: a convolutional layer introduces more parameters and a pooling layer typically discards some pixel-wise information.

For the first issue of too many convolutional layers, we can see that each convolutional layer represents a new weight layer so that deep structure with quantity of convolutional layers bring disadvantages of too many parameters. This problem might lead to over-fitting concern, difficulty to achieve real-time performance, and huge size of trained
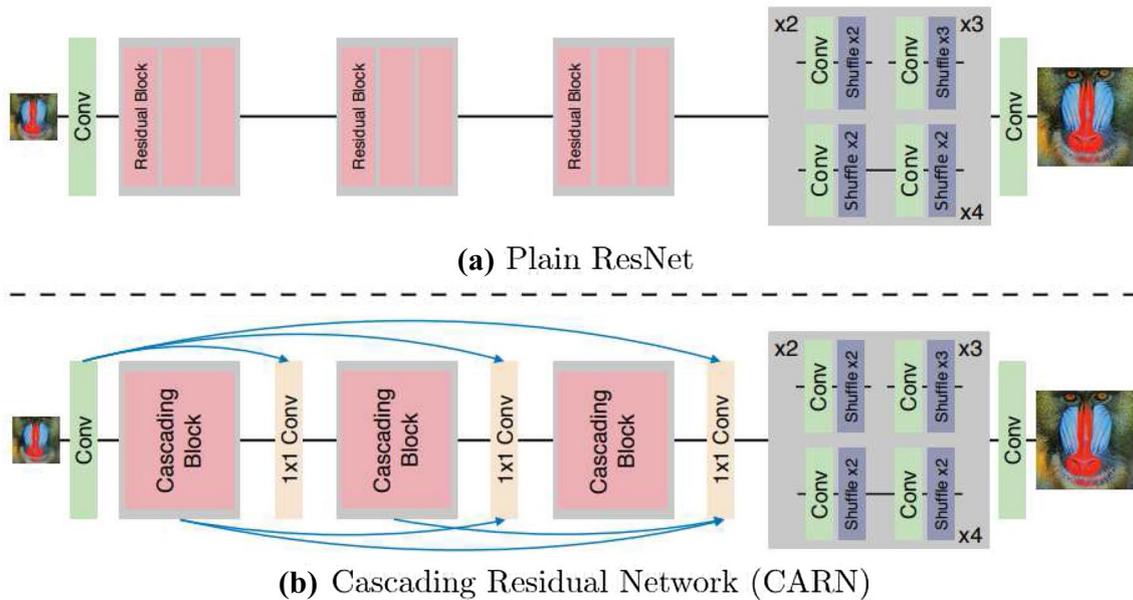
(a) Plain ResNet

(b) Cascading Residual Network (CARN)

Fig. 9 Network structure of Ahn et al. [2], where **a**, **b** represent plain ResNet and CARN structure, respectively. In the CARN model, each residual block is changed to a cascading block
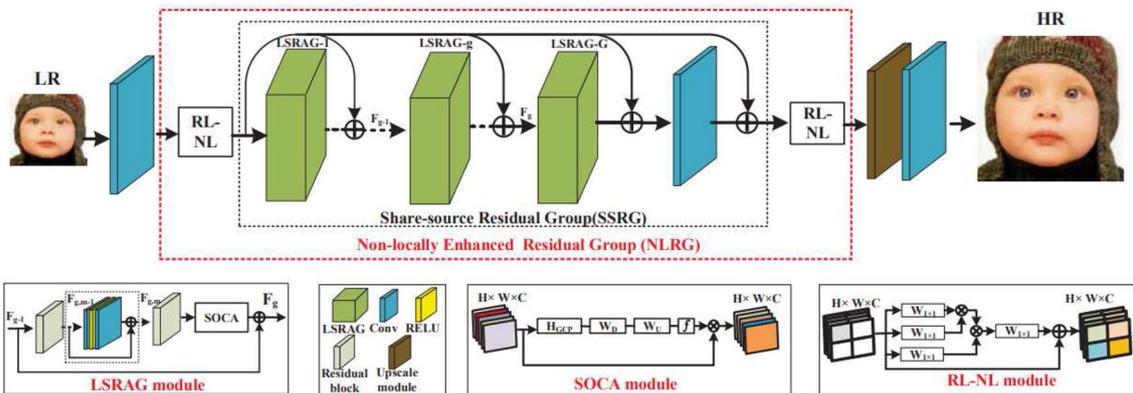

Fig. 10 Framework of second-order attention network (SAN) [22] and its sub-modules

model to store. Therefore, light structure of CNN without additional and time-consuming designs is required and pursued by researchers, which is the main trend of CNN-based methods for real-time SR.

Regarding image restoration problems such as super-resolution and denoising, image details are very important. Therefore, most deep-learning approaches for such problems do not use pooling or sub-sampling layers, so that important image details can be saved during process of SR tasks. For example, DRCN [53] repeatedly applies the same convolutional layer as many times as desired and does not apply pooling layers in their network architecture. The number of parameters does not increase while more recursions are performed, which offers a promising idea on network structure.

3.2 GAN and other deep learning-based super-resolution methods

Due to the unsupervised training property of GAN [90], GAN-base SR methods could use a large dataset of unlabeled images and work without any prior knowledge between inputting LR and HR image, which is essentially the main feature of GAN-based SR methods. Since GAN is originally designed to generate images, GAN-based SR methods could achieve super performance in generating photo-realistic SR images. However, we find no real-time methods are proposed for SR at current time. In fact, the most common use of GAN is to regard its generator part as a CNN network to perform low-to-high SR task. Without special design like statements

in last subsection, it is hard for such GAN-based methods to achieve real-time performance. In our thought, researchers on GAN are still focusing on solving several most important problems of GAN structure, like mode collapsing and hard to train.

At last but not least, we introduce other deep learning methods for SR, including deep auto-encoder and deep reinforcement learning.

Ledig et al. [60] find that most commonly used measurements to evaluate SR performance, such as MSE and PSNR, are designed with pixel-wise property. Since human perception often relies on evaluation in a global sense, some SR methods with high PSNR and MSE values would result in poor visual effects during SR process. Inspired by the significant property of GAN, they thus proposed super-resolution generative adversarial network (SRGAN) with a novel measurement, named as perceptual similarity. We show the structure of SRGAN in Fig. 11. Specifically, perceptual similarity is measured by a perceptual loss function, which serves as a sum form of an adversarial loss and a content loss. It is noted that adversarial loss ensures SRGAN could generate high-quality and photo-realistic SR images with help of a discriminator network, which is trained to classify the generated SR images are whether super-resolved images or original natural scene images.

We show sampling images of SRGAN and other comparison samples in Fig. 12. From Fig. 12, we could notice that SRGAN, designed with perceptual loss function, could recover photo-realistic textures from low-resolution images

sampled from public benchmarks. Meanwhile, SR methods designed with MSE-based measurement generate poorly visual HR images, even achieving high PSNR values at the same time. Therefore, we could conclude that SRGAN focuses on global contextual information to generate HR images, thus achieving lower pix-wise PSNR values than comparative methods as shown in Fig. 12. However, human perception is visual effect in a global sense; we thus observe photo-realistic textures from samples generated by SRGAN.

With the first and successful trial in SRGAN, Johnson et al. [50] further modify SRGAN on design of loss function, which could be improved as a sum form of pixel-wise loss, perceptual loss, and texture matching loss. In the context of combining GAN and CNN, Sajjadi et al. [91] propose EnhanceNe for automated texture synthesis, which utilizes feed-forward fully convolutional neural networks in an adversarial training setting. Their proposed network successfully create realistic textures, rather than optimizing for a pixel-accurate reproduction of ground truth images during training.

By leveraging the extension of the basic GAN framework [149], Yuan et al. [133] propose an unsupervised SR algorithm with a Cycle-in-Cycle network structure, which is inspired by the recent successful image-to-image translation applications. They further expand the algorithm to a modified version, i.e., MCinCGAN [143], which utilizes a multiple Cycle-in-Cycle network structure to deal with the more general case of SR tasks, using multiple generative adversarial networks (GAN) as the basis components.
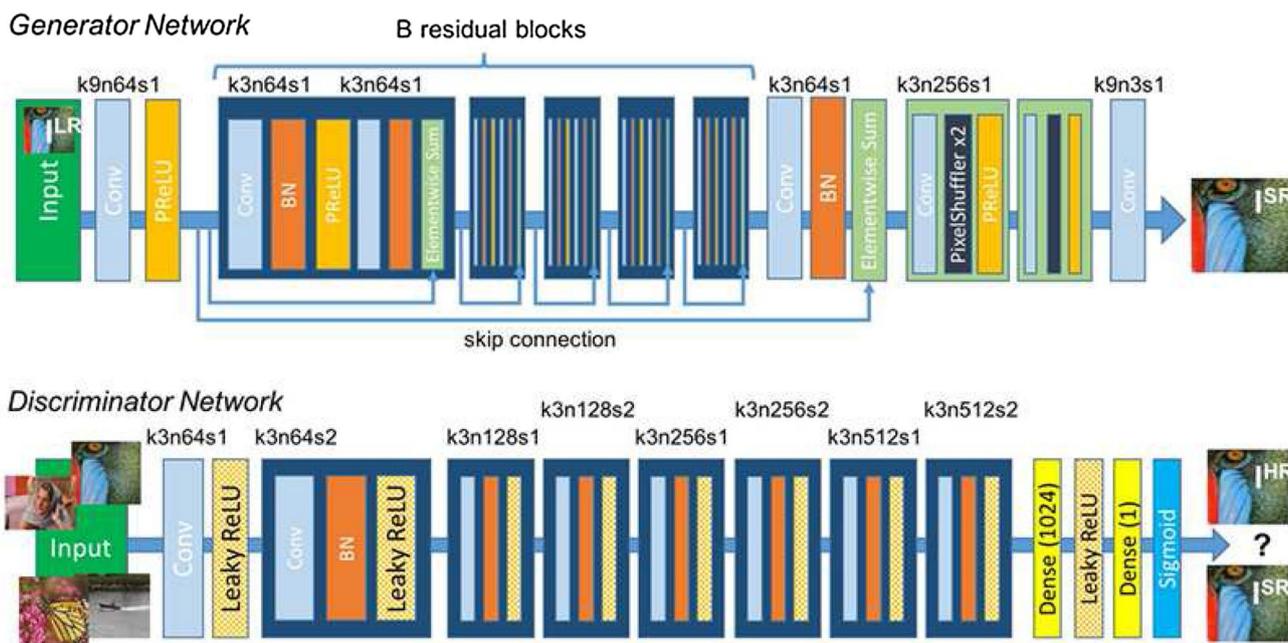

Fig. 11 Architecture of super-resolution generative adversarial network [60]

Fig. 12 Comparisons on SR quality, PSNR, and SSIM of different SR methods. It is noted that four images refer to HR images generated by bicubic interpolation, deep residual network optimized by MSE measurement, deep residual generative adversarial network designed with human perception loss [60], and original HR image, respectively

More precisely, their proposed first network cycle aims at mapping the noisy and blurry LR input to a noise-free LR space. On the basis of first network cycle, a new cycle with a well-trained ×2 network model is orderly introduced to super-resolve the intermediate output of the former cycle.

In this way, the number of total cycles depends on the different up-sampling factors (×2, ×4, ×8), which is presented in Fig. 13. Finally, users could get the desired HR images with different scale factors by training all modules in an end-to-end manner .
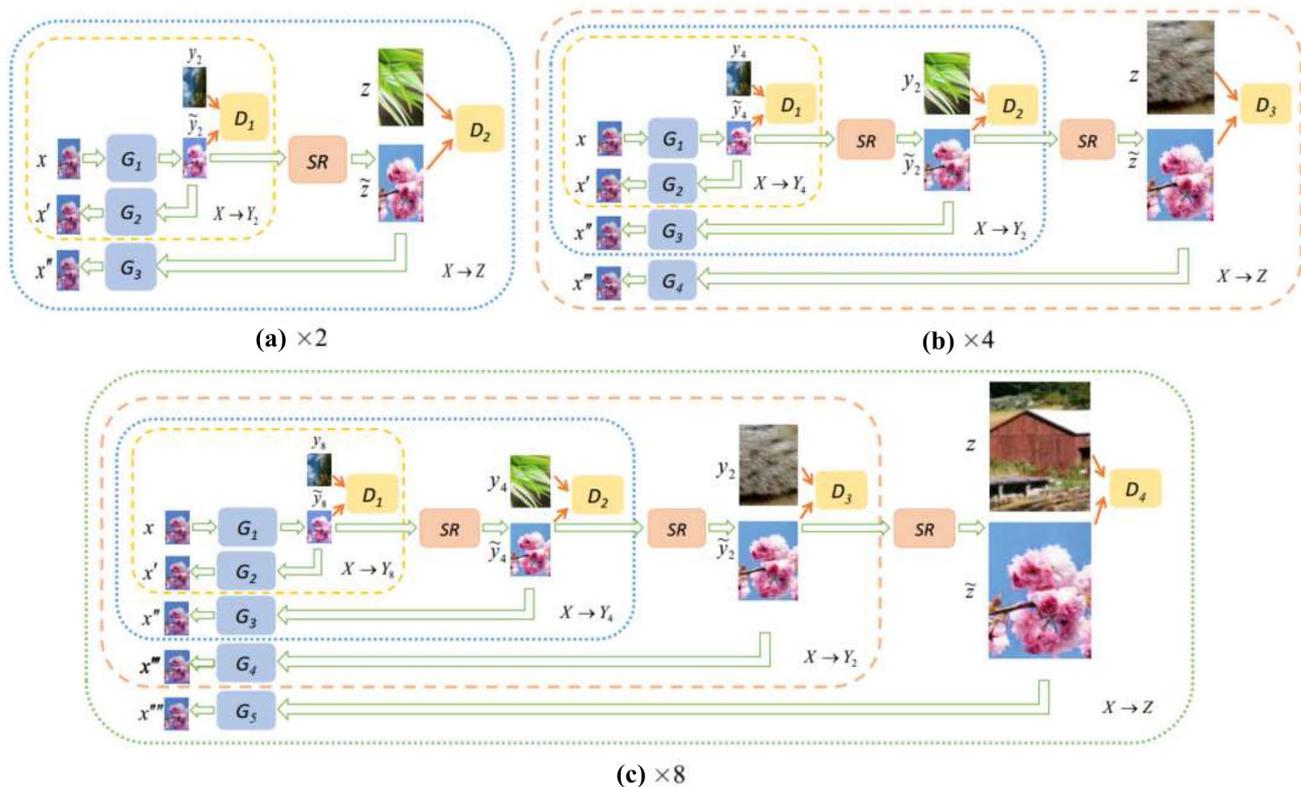

Fig. 13 The framework of the proposed MCinCGAN [143], where G1, G2, G3, G4, and G5 are generators and D1, D2, D3, and D4 are discriminators. It is noted that **a–c** show the frameworks for ×2, ×4 and ×8, respectively

Bulat et al. [10] propose a two-stage process involving the idea of using a GAN to learn how to perform image degradation at first and then learn image super-resolution with trained GAN, where we show its network structure in Fig. 14. Specifically, they train a High-to-Low GAN to learn degradation and down-sampling operations on HR images during the first stage. Once training process of High-to-Low GAN is finished, they utilize pairs of low- and high-resolution images computed by High-to-Low GAN as training samples for a new Low-to-High GAN. After training based on enough and variant training pairs generated by High-to-Low GAN, the resulting Low-to-High GAN could output desirable HR images The most interesting point for Bulat et al. [10] lies in the fact that their proposed method only requires unpaired image data for training so that annoying work of pairing low and high-resolution images can be avoided. By applying this two-stage process, the proposed unsupervised model effectively increases the quality of super-resolving real-world LR images and obtains large improvement over previous state-of-the-art works. Although Bulat et al. [10] can simulate more complex degradation, there is no guarantee that such simulated degradation can approximate the authentic degradation in practical scenarios which is usually very complicated. Therefore, Zhao et al. [146] improves it by exploring the relationship between reconstruction and degradation with bi-cycle structure, which jointly stabilizes the training of SR reconstruction and degradation networks. Most importantly, their degradation model is trained in an unsupervised way, i.e., without using paired images. since there are no pairs of LR-HR images in practice,

Bulat and Tzimiropoulos [9] propose Super-FAN to complete two tasks simultaneously, i.e., improves resolution of face images and detects facial landmarks inside improved face images. Essentially, super-FAN constructs two sub-networks to first optimize loss function for constructing of a convinced heat map and then perform face alignment through heat map regression. By jointly training both sub-networks, they report desirable HR and detection results based on not only input LR images, but also real-world images affected by variant factors.

To further enhance the visual quality, Wang et al. [116] propose enhanced Super-Resolution Generative Adversarial Network (ESRGAN) to generate realistic textures, which introduces the Residual-in-Residual Dense Block (RRDB) without batch normalization as the basic network building unit. Furthermore, they successfully modify network structure with relativistic GAN and improve the perceptual loss with features before activation. Benefiting from these improvements, ESRGAN achieves consistently better visual quality with more realistic and natural textures and wins the first place in the PIRM2018-SR Challenge.

Deep Reinforcement Learning (DRL) has also been introduced recently. Following the thought of reinforce learning, DRL generally designs a learning policy to guide spatial attention. In other words, they utilize reward scheme of reinforce learning to navigate up-scaling regions, which results
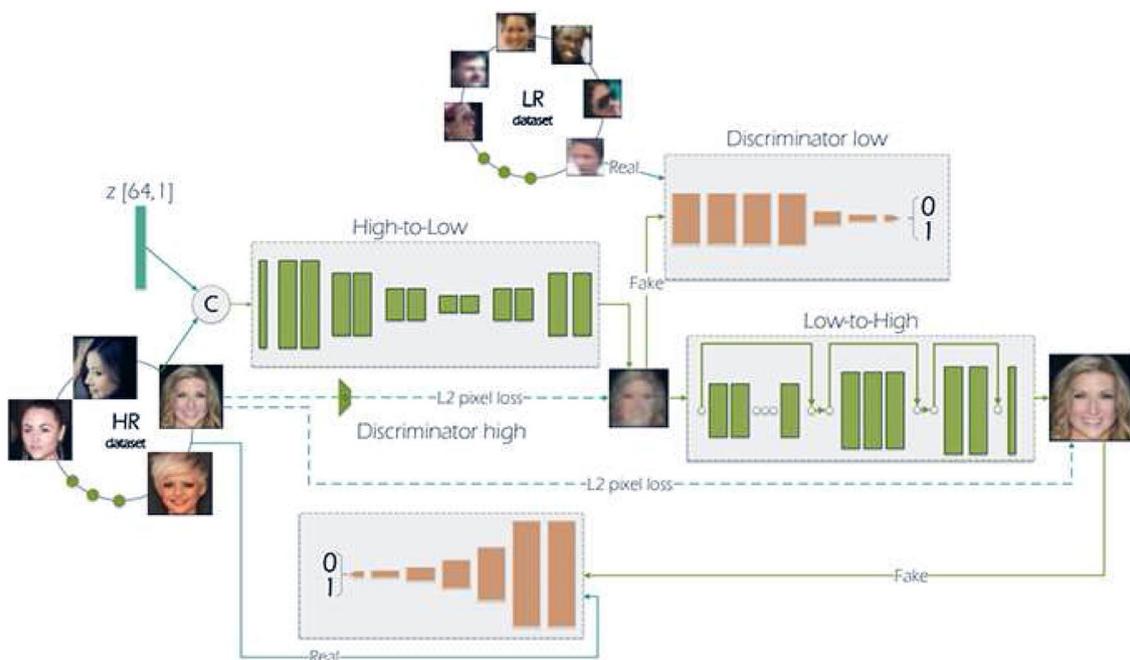

Fig. 14 Architecture design of Bulat et al. [10]. It is noted that LR and HR images are not paired in the training dataset

in an adaptively optimizing way for SR based on the characteristics of inputting images. For instance, Cao et al. [12] propose a novel attention-aware Face Hallucination framework, which follows principles of DRL to sequentially discover patches required to up-scale at first. Afterwards, they follow the resulting optimization sequence to perform facial patch enhancement by exploiting and involving global characteristics of the inputting facial image. Work of Cao et al. [12] is quite new in concept and provides a novel thought on how to adaptively process high-resolution images. However, the computation of [12] is much larger than SRCNN, reporting four times larger than SRCNN, due to the high computation cost of its deep structure.

4 Comparisons

4.1 Datasets and measurements

Many image datasets are popular to be adopted to prove and compare effectiveness of different super resolution methods in SR community. We list most of them with cites and descriptions in Table 1, where we can notice their original usages are different such as segmentation, classification, etc. Therefore, there are no actually construction rules to organize a specific dataset on SR topic. Among all these dataset, four datasets, i.e., SET5, SET14, B100, and URBAN100, are mostly commonly used for comparison in SR community. Our following performance will be performed on these four datasets.

There are two standard quality measures, i.e., peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) [37], which have been mostly used for measuring the quality of super resolution methods. Specially, the definition of PSNR relies on MSE; we thus define these three measures as follows:

$$\text{MSE} = \frac{m * n}{\sum_{i=1}^{m} \sum_{j=1}^{n} (I(i,j) - P(i,j))^2} \tag{5}$$

$$\text{PSNR} = 10 \times \log\left(\frac{255^2}{MSE}\right) \tag{6}$$

Table 1 Image dataset for super resolution and their corresponding descriptions

Index	Dataset name	Descriptions
1	SET5 [7]	It contains five images named as baby, bird, butterfly, head, and woman. Scale factor used includes 2×, 3× and 4×
2	SET14 [137]	It includes 14 commonly used images for super-resolution evaluation. Compared with Set5, more diversity are introduced by SET14 including bridge, comic, poster and so on
3	B100 [68]	Full name is Berkeley Segmentation Dataset and Benchmark, which was originally designed for natural scene image segmentation. Due to its high quality to cover variant attributes of natural scene images, it has been widely adopted to evaluate performance of SR approaches on natural scene images. It is noted that BSD300 uses 100 testing images (named here 'B100') and 200 training images. The latest version is named as BSD500 including 200 more fresh test images
4	URBAN100 [46]	It has 100 HR images diverse in real-world scenes and is widely used by examining with its self-similarities
5	ImageNet [88]	ImageNet is a large visual database in visual object recognition and classification research. Latest, more than 14 million images with over 20,000 categories have been hand-annotated. Since 2010, A famous challenge named as the ImageNet Large Scale Visual Recognition Challenge (ILSVRC) is performed on ImageNet to compete on object or scene classification and detection
6	super texture dataset [20]	The main concern of this dataset lies on SR operation on texture. It provides 136 texture images in total
7	Timeofte dataset [110]	It is one of the most widely used datasets by SISR researchers, which consists of 91 training images, and testing images in BSD500, Set5, Set14 and super texture dataset
8	DIV2K [108]	The DIVerse 2K resolution image dataset served as a benchmark for NTIRE 2017 Challenge. It is designed to evaluate SR performance on high-resolution images. It consists of 800 training, 100 validation, and 100 test images
9	Mnist [59]	The MNIST database contains 60,000 training and 10,000 testing images. It is a large database of handwritten digits to evaluate performance of SR on hand-written images
10	Manga109 [31]	It consists of numerous comic sketches from 109 Japanese comic books and often adopted as basic dataset to evaluate performance of SR on drawing images
11	LIVE [98]	All images are originally captured by another project for the purpose of generic shape matching and recognition. Now it serves as a dataset of images with visual distortion effects to examine the extreme handling performance of SR methods
12	L20 [111]	It focuses on images with large size. The common size in L20 is range from 3m to 29m pixels, while most of the other datasets own images below 0.5 m pixels

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}, \tag{7}$$

where m and n refer to the width and the height of the image, I and P represent output image after operation of super resolution and the input original image respectively, μ_x and μ_y represent the means of x and y, σ_x^2 and σ_y^2 represent the variances of x and y, σ_{xy} is the covariance of x and y, and c_1 and c_2 are two preset variables.

PSNR is often adopted to measure the quality of reconstruction during super resolution by calculating power ratio of noisy signal introduced by super resolution. Higher values of PSNR represents better quality of reconstructed image. Meanwhile, SSIM is quantitative measure used to quantify the similarities of structure between original and HR image. High SSIM value can reflect that the process of super resolution does not affect image basic structure, thus proving better reconstruction quality.

4.2 Performance analysis

Table 2 reports SR performance with five usually adopted dataset. All these statics are obtained by variety of approaches like sparse coding, CNN-based, GAN-based, and so on. We observe that SAN outperforms other algorithms in most datasets with scale factor 2×, 3×, 4×, and 8×, meanwhile DualGAN achieves best results with scale factor 8× and B100 dataset. However, it is noted that we only collect a few performance examples with smaller scale factors for most GAN-based methods, since most GAN-based methods perform experiments on large dataset and with higher scale factor. Therefore, it is not fair to say SAN is better at SR with smaller scale factor. Essentially, SAN could be regarded as an improved version of EDSR, which utilizes attention architecture to generate more task-specified feature map for SR task. Meanwhile, DualGAN modifies original structure of GAN to two generators and discriminators to learn a mixture of many distributions from prior to the complex distribution. Both methods have shown enough progress on improvement of CNN or GAN based methods.

Based on the success of DRRN and other residual-based learning methods for SR like DRRN, EDSR, LapSRN, etc. we could conclude that residual learning brings many benefits to SR image reconstruction, since it captures and models the most important characteristics of SR. For example, Ms-LapSRN is an improved version of LapSRN by reconstructing the sub-band residuals of HR images at multiple pyramid levels. DSRN exploits both lowre-solution (LR) and high-resolution (HR) signals jointly, where recurrent signals are exchanged between these states in both directions (both LR to HR and HR to LR) via delayed feedback. Based on the high performance of these residual learning network, we can conclude that CNN-based methods are still mainstream to achieve high performance on SR.

In Table 3, we compare average frame rate (FPS) of different methods on SET5, SET14, BSDS100, URBAN100, and MANGA109 with the scale factors 2×, 4×, and 8×, respectively. It is noted that since the codes of SRCNN and FSRCNN for testing are based on CPU implementations, we reconstruct these models in MatConvNet with the same network weights to measure the run time on GPU. By evaluating the execution time of each algorithm on a machine with 3.4 GHz Intel i7 CPU (64G RAM) and Nvidia Titan X GPU (12G Memory), we upscale input images by scale factors 2×, 4×, and 8× in experiments, respectively.

From Table 3, we can notice that FSRCNN is a fast enough algorithm, since it applies several convolution operations on LR images and has fewer network parameters. LapSRN achieves large FPS value, since it continually upscales images in a pyramid structure and adaptively applies different number of convolutional layers. D2GAN could achieve better runtime performance with smaller scale factors, while it achieves much lower FPS value than FSRCNN with larger scale factor. This fact indicates GAN is not stable in running time and requires further development to obtain consistent performance. Besides, runtime performance of SRCNN, VDSR, RFL, SCN, DRCN, A+, and LapSRN all depend on the size of output images, while the speed of FSRCNN is almost constant. FSRCNN, LapSRN, ProGAN, and D2GAN achieve real-time speed (i.e., 24 frames per second) on most datasets. Runtime performance of SRCNN could be commented as fast, while FPS values of other methods are quite low, even implemented with a powerful GPU card. Therefore, there is still much work on adapting deep networks to achieve real-time SR.

4.3 Comparisons between CNN and GAN-based SR methods

Based on results achieved by CNN and GAN-based SR methods, we aim to generally compare differences and future developments of these two categories. Essentially, CNN and GAN-based SR methods can be regarded as typical supervised and unsupervised learning algorithms respectively, which is the main reason for their huge differences. More precisely, supervised CNN methods attempt to learn directly the mapping between LR and HR images and highly depend on predetermined assumptions. Meanwhile, GAN-based networks are much more flexible with promising performance due to incorporated unsupervised training.

How to appropriately collect training data with sufficient enough information to support SR tasks has become a major problem in CNN-based SR methods. For majority researchers, they collect image data for SR by first down-sampling

Table 2 Quantitative evaluation of a quantity of SR algorithm, where average PSNR/SSIM with scale factor ×2, ×3, ×4 is listed, information not provided by original authors is marked with [–]

Methods	Scale	Set5	Set14	B100	Urban100	Manga109
×2	Bicubic	33.66/0.9299	30.24/0.87688	29.56/0.8431	26.88/0.8403	30.80/0.9339
	A+ [110]	36.54/0.9544	32.28/0.9056	31.21/0.8863	29.20/0.8938	–/–
	SRCNN [26]	36.66/0.9542	32.45/0.9067	31.36/0.8879	29.50/0.8946	35.60/0.9663
	VDSR [52]	37.53/0.9590	33.05/0.9130	31.90/0.8960	30.77/0.9140	37.22/0.9750
	EDSR [65]	38.11/0.9602	33.92/0.9195	32.32/0.9013	32.93/0.9351	39.10/0.9773
	LapSRN [58]	37.52/0.9591	33.08/0.9130	31.08/0.8950	30.41/0.9101	37.27/0.9740
	GuideAE [17]	37.52/0.9591	33.08/0.9130	31.08/0.8950	30.41/0.9101	37.27/0.9740
	SRGAN [60]	37.22/0.9263	32.14/0.8862	31.89/0.8761	31.02/0.8955	–/–
	ESRGAN [116]	37.81/0.9531	33.62/0.9152	31.99/0.8873	32.01/0.9131	–/–
	SRMDNF [141]	37.79/0.9601	33.32/0.9154	32.05/0.8984	31.33/0.9204	38.07/0.976
	IDN [147]	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	38.02/0.9749
	CSFM [45]	38.26/0.9615	34.07/0.9213	32.37/0.9021	33.12/0.9366	39.40/0.9785
	CARN [2]	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.51/0.9312	–/–
	RDN [144]	38.24/0.9614	34.01/0.9212	32.34/0.9017	32.89/0.9353	39.18/0.9780
	SAN [22]	**38.35/0.9619**	**34.44/0.9244**	**32.50/0.9038**	**33.73/0.9416**	**39.72/0.9797**
×3	Bicubic	30.39/0.8682	27.55/0.7742	27.21/0.7385	24.46/0.7349	26.95/0.8556
	A+ [110]	32.58/0.9088	29.13/0.8188	28.29/0.7835	26.03/0.7973	–/–
	SRCNN [26]	32.75/0.9090	29.30/0.8215	28.41/0.7863	26.24/0.7989	30.48/0.9117
	VDSR [52]	33.67/0.9210	29.78/0.8320	28.83/0.7990	27.14/0.8290	32.01/0.9340
	EDSR [65]	34.65/0.9280	30.52/0.8462	29.25/0.8093	28.80/0.8653	34.17/0.9476
	LapSRN [58]	33.82/0.9227	29.87/0.8320	28.82/0.7980	27.07/0.8280	32.21/0.9350
	GuideAE [17]	33.82/0.9227	29.87/0.8320	28.82/0.7980	27.07/0.8280	32.21/0.9350
	SRMDNF [141]	34.12/0.9254	30.04/0.8371	28.97/0.8025	27.57/0.8398	33.00/0.9403
	IDN [147]	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	32.69/0.9378
	CSFM [45]	34.76/0.9301	30.63/0.8477	29.30/0.8105	28.98/0.8681	34.52/0.9502
	CARN [2]	34.29/0.9255	30.29/0.8407	29.06/0.8034	27.38/0.8404	–/–
	RDN [144]	34.71/0.9296	30.57/0.8468	29.26/0.8093	28.80/0.8653	34.13/0.9484
	SAN [22]	**34.89/0.9306**	**30.77/0.8498**	**29.38/0.8121**	**29.29/0.8730**	**34.74/0.9512**
×4	Bicubic	28.42/0.8104	26.00/0.7027	25.96/0.6675	23.14/0.6577	24.89/0.7866
	A+ [110]	30.28/0.8603	27.32/0.7491	26.82/0.7087	24.32/0.7183	–/–
	SRCNN [26]	30.48/0.8628	27.50/0.7513	26.90/0.7101	24.52/0.7221	27.58/0.8555
	VDSR [52]	31.35/0.8830	28.02/0.7680	27.29/0.0726	25.18/0.7540	28.83/0.8870
	EDSR [65]	32.46/0.8968	28.80/0.7876	27.71/0.7420	26.64/0.8033	31.02/0.9148
	LapSRN [58]	31.54/0.8850	28.19/0.7720	27.32/0.7270	25.21/0.7560	29.09/0.8900
	GuideAE [17]	31.54/0.8850	28.19/0.7720	27.32/0.7270	25.21/0.7560	29.09/0.8900
	SRMDNF [141]	31.96/0.8925	28.35/0.7772	27.49/0.7337	25.68/0.7731	30.09/0.9024
	IDN [147]	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	29.41/0.8936
	CSFM [45]	32.61/0.9000	28.87/0.7886	27.76/0.7432	26.78/0.8065	31.32/0.9183
	RDN [144]	32.47/0.8990	28.81/0.7871	27.72/0.7419	26.61/0.8028	31.00/0.9151
	CARN [2]	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	–/–
	SRGAN [60]	29.40/0.8472	26.64/0.7101	25.16/0.6682	25.11/0.7253	–/–
	ESRGAN [116]	31.40/0.8713	27.98/0.7624	27.21/0.7123	31.99/0.8874	–/–
	SAN [22]	**32.70/0.9013**	**29.05/0.7921**	**27.86/0.7457**	**27.23/0.8169**	**31.66/0.9222**

Table 2 (continued)

Methods	Scale	Set5	Set14	B100	Urban100	Manga109
×8	Bicubic	24.40/0.6580	23.10/0.5660	23.67/0.5480	20.74/0.5160	21.47/0.6500
	SRCNN [26]	25.33/0.6900	23.76/0.5910	24.13/0.5660	21.29/0.5440	22.46/0.6950
	VDSR [52]	25.93/0.7240	24.26/0.6140	24.49/0.5830	21.70/0.5710	23.16/0.7250
	LapSRN [58]	26.15/0.7380	24.35/0.6200	24.54/0.5860	21.81/0.5810	23.39/0.7350
	MSLapSRN [57]	26.34/0.7558	24.57/0.6273	24.65/0.5895	22.06/0.5963	23.90/0.7564
	DualGAN [136]	–/–	–/–	**27.85/0.8911**	–/–	–/–
	EDSR [65]	26.96/0.7762	24.91/0.6420	24.81/0.5985	22.51/0.6221	24.69/0.7481
	DCGAN [85]	–/–	–/–	24.09/0.8619	–/–	–/–
	GP-GAN [120]	–/–	–/–	27.06/0.8595	–/–	–/–
	D2GAN [76]	–/–	–/–	27.17/0.8067	–/–	–/–
	SAN [22]	**27.30/0.7849**	**25.23/0.6493**	24.97/0.6031	**22.91/0.6369**	**25.17/0.7964**

Bold text indicates the best performance among all comparative methods

Table 3 Comparisons of the FPS (frames per seconds) on 5 benchmark dataset with scale factors 2×, 4×, and 8×

Methods	Scale	Set5	Set14	B100	Urban100	Manga109
×2	A+ [110]	1.12	0.52	0.74	0.15	0.10
	SRCNN [26]	24.70	22.92	39.50	9.03	6.53
	FSRCNN [27]	31.04	53.86	98.20	**47.23**	34.48
	RFL [94]	0.65	0.45	0.52	0.13	0.15
	SCN [119]	1.19	0.85	1.19	0.24	0.17
	VDSR [52]	11.01	6.46	10.00	2.12	1.71
	DRCN [53]	0.70	0.37	0.59	0.10	0.08
	LapSRN [56]	30.20	40.00	97.36	16.81	85.32
	ProGAN [117]	25.88	34.80	71.59	14.97	73.48
	D2GAN [76]	**46.21**	**61.72**	**141.34**	24.86	**132.12**
×4	A+ [110]	2.86	1.62	2.43	0.49	0.41
	SRCNN [26]	21.74	22.27	40.13	9.95	7.13
	FSRCNN [27]	31.61	**56.58**	**101.54**	**53.95**	55.23
	RFL [94]	1.97	1.21	1.64	0.42	0.34
	SCN [119]	1.38	0.87	1.19	0.31	0.25
	VDSR [52]	10.71	6.59	9.91	2.15	1.76
	DRCN [53]	0.80	0.37	0.59	0.10	0.08
	LapSRN [56]	25.49	25.46	54.35	12.40	47.63
	ProGAN [117]	20.67	20.18	43.05	9.92	37.29
	D2GAN [76]	**38.33**	38.28	87.95	16.08	**75.71**
×8	A+ [110]	5.79	2.84	4.31	0.80	0.64
	SRCNN [26]	20.92	17.69	40.13	9.81	7.17
	FSRCNN [27]	**34.10**	**63.28**	**104.46**	**71.67**	**71.95**
	RFL [94]	2.54	1.61	2.25	0.47	0.33
	SCN [119]	0.79	0.53	0.68	0.21	0.19
	VDSR [52]	10.58	6.50	10.13	2.15	1.77
	LapSRN [56]	24.02	23.40	50.44	10.54	33.09
	ProGAN [117]	17.97	17.48	38.76	7.66	25.11
	D2GAN [76]	23.70	23.08	50.97	10.28	34.41

Bold indicates the best performance among all comparative methods

original HR to get LR images, and then using LR images as input to pursue HR results. However, such collecting process is not exactly the same as what happened in the application scenario, since patterns to induce quality loss can be various, such as image transmission, different compression algorithms and so on. Simply utilizing down-sampling to generate LR images would do great harm to flexibility of CNN-based SR methods with respect to different scenarios. On the contrary, GAN-based methods require less image data for SR tasks, since they use a largely unsupervised training process on the real images. Therefore, they do not require label or prior condition between LR and HR image.

Major difficulty in constructing GAN-based SR methods lies in their designs of architecture and loss function. It is noted the most common used measurement, i.e., MSE, is adopted in favor of maximizing PSNR. However, hallucinated details of generated HR images are often accompanied with unpleasant artifacts, even achieving high PSNR values. In other words, traditional measurements expose several constraints to human perception. However, GAN has ability to produce better results only with the integrated perceptual assessment other than single and simple measurement. In order to further improve visual quality, it is thus required to improve key components of GAN models for SR, i.e., network architecture and loss function, which are now progressively explored and developed by researchers. All these truths imply GAN is still developing and is promising to achieve better reconstruction results than CNN-based methods. Moreover, GAN is hard to train due to mode collapsing problem, which leads to early stopping by concentrating on only a few modes instead of the whole data space. On the contrary, CNN-based methods have been developed for several generations with visually desirable reconstruction results. With back-propagation training methods and well-developed CNN construction softwares, it is much easier to implement and train a CNN-based SR method than performing with GAN model.

Above all, we can conclude that GAN-based SR methods have high potential to achieve better reconstruction results than CNN-based SR methods, due to its property to describe variety of SR patterns with less labeled data. However, GAN model needs to be further developed to be easily trained and implemented.

5 Summary

Super-resolution is a hot and important research topic in computer vision and image processing community. By applying SR technologies, users can not only improve the resolution and visual appearance of inputting images, but also help improve accuracy and effectiveness of vision-based machine learning systems which generally regard high-resolution images as input. Inspired by the significant performance of deep learning methods, this paper focuses on reviewing current deep learning methods for real-time image super-resolution. As the first comprehensive survey on such topic, it has analyzed recent approaches, classified them according to as many as criteria, and illustrated performance for the most representative approaches.

In the past decade, research in this field has progressed as improved methods emerge. However, the small number of deep learning- based SR methods achieving real-time performance shows that ample room remains for future research. Essentially, the main reason for less methods on real-time SR lies in the fact that deep learning methods are hard to achieve fast processing speed without high computation resource. However, applying deep learning methods on SR is a reasonable way to achieve high performance with the current large datasets, named as "big data". We thus present main challenges of deep learning for real-time SR: how to adapt deep learning-based SR methods with acceleration strategies to deal with "big data" situation. Furthermore, low-resolution images captured in extreme imaging conditions require robust enough algorithms to deal with such real-life complexities.

It is thus essential to develop novel methods, which are not only effective and efficient for "big data" processing, but also robust enough to handle extremely processing. Although abundant optimization methods have been proposed for real-time SR with deep learning structures in Sect. 3, high efficiency, effectiveness, and robustness for SR on specific application area are still highly required by industry and require further improvement. Despite designing alternative and high effective neural network structures, cloud computing is another simple and efficient solution to improve effectiveness of SR. By providing enough computing and storage services over the Internet [82, 123, 125, 126] for local SR tasks, a powerful computing platform with easy access and high-scalability could be utilized locally and help users accomplish their SR goals. There exist other novel methods to help improve SR from different aspects. For example, deep compression methods could help prune neural networks for SR, resulting in less storage and computation consume.

Based on all these analyses, we believe that this review is useful for developers who are willing to improve performance of their SR solutions in both running time and accuracy. Our review will serve as a guidance and dictionary for further research activities in this area, especially in the deployment of real-time super-resolution with deep learning methods.

References

1. Agustsson, E., Timofte, R., Van Gool, L.: Anchored regression networks applied to age estimation and super resolution. In: Proceedings of International Conference on Computer Vision, pp. 1652–1661 (2017)

2. Ahn, N., Kang, B., Sohn, K.: Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of European Conference on Computer Vision, pp. 256–272 (2018)

3. Akae, N., Makihara, Y., Yagi, Y.: Gait recognition using periodic temporal super resolution for low frame-rate videos. In: Proceedings of International Joint Conference on Biometrics, pp. 1–7 (2011)

4. Alonso-Fernandez, F., Farrugia, R.A., Bigun, J., Fierrez, J., Gonzalez-Sosa, E.: A survey of super-resolution in iris biometrics with evaluation of dictionary-learning. IEEE Access (2018)

5. Aly, H.A., Dubois, E.: Image up-sampling using total-variation regularization with a new observation model. IEEE Trans. Image Process. **14**(10), 1647–1659 (2005)

6. Belekos, S.P., Galatsanos, N.P., Katsaggelos, A.K.: Maximum a posteriori video super-resolution using a new multichannel image prior. IEEE Trans. Image Process. **19**(6), 1451–1464 (2010)

7. Bevilacqua, M., Roumy, A., Guillemot, C., Alberi-Morel, M.L.: Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In: Proceedings of British Machine Vision Conference (2012)

8. Bose, N.K., Ahuja, N.A.: Superresolution and noise filtering using moving least squares. IEEE Trans. Image Process. **15**(8), 2239–2248 (2006)

9. Bulat, A., Tzimiropoulos, G.: Super-fan: Integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans. In: Proceedings of Computer Vision and Pattern Recognition, pp. 109–117 (2018)

10. Bulat, A., Yang, J., Tzimiropoulos, G.: To learn image super-resolution, use a gan to learn how to do image degradation first. In: Proceedings of the European Conference on Computer Vision, pp. 185–200 (2018)

11. Caballero, J., Ledig, C., Aitken, A., Acosta, A., Totz, J., Wang, Z., Shi, W.: Real-time video super-resolution with spatio-temporal networks and motion compensation. In: Proceedings of Computer Vision and Pattern Recognition, pp. 2848–2857 (2017)

12. Cao, Q., Lin, L., Shi, Y., Liang, X., Li, G.: Attention-aware face hallucination via deep reinforcement learning. CoRR **abs/1708.03132** (2017)

13. Capel, D., Zisserman, A.: Automated mosaicing with super-resolution zoom. In: Proceedings of Computer Vision and Pattern Recognition, pp. 885–891 (1998)

14. Chai, Y., Ren, J., Zhao, H., Li, Y., Ren, J., Murray, P.: Hierarchical and multi-featured fusion for effective gait recognition under variable scenarios. Pattern Anal. Appl. **19**(4), 905–917 (2016)

15. Chang, H., Yeung, D.Y., Xiong, Y.: Super-resolution through neighbor embedding. Proc. Comput. Vis. Pattern Recognit. **1**, 275–282 (2004)

16. Che, T., Li, Y., Jacob, A.P., Bengio, Y., Li, W.: Mode regularized generative adversarial networks. CoRR arxiv:abs/1612.02136 (2016)

17. Chen, R., Qu, Y., Li, C., Zeng, K., Xie, Y., Li, C.: Single-image super-resolution via joint statistical models-guided deep autoencoder network. Neural Comput. Appl. pp. 1–11 (2019)

18. Chen, Y., Tai, Y., Liu, X., Shen, C., Yang, J.: Fsrnet: End-to-end learning face super-resolution with facial priors. In: Proceedings of Computer Vision and Pattern Recognition, pp. 2492–2501 (2018)

19. Cristóbal, G., Gil, E., Šroubek, F., Flusser, J., Miravet, C., Rodríguez, F.d.B.: Superresolution imaging: a survey of current techniques. In: Advanced signal processing algorithms, architectures, and implementations XVIII, vol. **7074**, p. 70740C (2008)

20. Dai, D., Timofte, R., Van Gool, L.: Jointly optimized regressors for image super-resolution. Comput. Graph. Forum **34**(2), 95–104 (2015)

21. Dai, S., Han, M., Xu, W., Wu, Y., Gong, Y.: Soft edge smoothness prior for alpha channel super resolution. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–8 (2007)

22. Dai, T., Cai, J., Zhang, Y., Xia, S.T., Zhang, L.: Second-order attention network for single image super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 11065–11074 (2019)

23. Demirel, H., Anbarjafari, G.: Image resolution enhancement by using discrete and stationary wavelet decomposition. IEEE Trans. Image Process. **20**(5), 1458–1460 (2011)

24. Denton, E.L., Chintala, S., Fergus, R., et al.: Deep generative image models using a laplacian pyramid of adversarial networks. In: Proceedings of Advances in neural information processing systems, pp. 1486–1494 (2015)

25. Dey, N., Li, S., Bermond, K., Heintzmann, R., Curcio, C.A., Ach, T., Gerig, G.: Multi-modal image fusion for multispectral super-resolution in microscopy. In: Proceedings of Medical Imaging 2019: Image Processing, p. 109490D (2019)

26. Dong, C., Loy, C.C., He, K., Tang, X.: Learning a deep convolutional network for image super-resolution. In: Proceedings of European Conference on Computer Vision, pp. 184–199 (2014)

27. Dong, C., Loy, C.C., Tang, X.: Accelerating the super-resolution convolutional neural network. In: Proceedings of European Conference on Computer Vision, pp. 391–407 (2016)

28. Dong, W., Zhang, L., Shi, G., Wu, X.: Image deblurring and super-resolution by adaptive sparse domain selection and adaptive regularization. IEEE Trans. Image Process. **20**(7), 1838–1857 (2011)

29. Duchon, C.E.: Lanczos filtering in one and two dimensions. J. Appl. Meteorol. **18**(8), 1016–1022 (1979)

30. Fahmy, G.: Super-resolution construction of iris images from a visual low resolution face video. In: National Radio Science Conference, pp. 1–6 (2007)

31. Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., Aizawa, K.: Manga109 dataset and creation of metadata. In: Proceedings of International Workshop on Comics Analysis, Processing and Understanding, pp. 2–3 (2016)

32. Gatys, L.A., Ecker, A.S., Bethge, M.: A neural algorithm of artistic style. CoRR arxiv:abs/1508.06576 (2015)

33. Girshick, R., Donahue, J., Darrell, T., Malik, J.: Rich feature hierarchies for accurate object detection and semantic segmentation. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 580–587 (2014)

34. Glasner, D., Bagon, S., Irani, M.: Super-resolution from a single image. In: Proceedings of IEEE International Conference on Computer Vision, pp. 349–356 (2009)

35. Gong, W., Qi, L., Xu, Y.: Privacy-aware multidimensional mobile service quality prediction and recommendation in distributed fog environment. Wireless Communications and Mobile Computing **2018**, (2018)

36. Goodfellow, I.J., Abadie, J.P., Mirza, M., Xu, B., Farley, D.W., Ozair, S., Courville, A.C., Bengio, Y.: Generative adversarial nets. In: Proceedings of Neural Information Processing Systems, pp. 2672–2680 (2014)

37. Greenspan, H.: Super-resolution in medical imaging. Comput. J. **52**(1), 43–63 (2009)

38. Gu, S., Sang, N., Ma, F.: Fast image super resolution via local regression. In: Proceedings of International Conference on Pattern Recognition, pp. 3128–3131 (2012)

39. Ha, V.K., Ren, J., Xu, X., Zhao, S., Xie, G., Vargas, V.M.: Deep learning based single image super-resolution: A survey. In: International Conference on Brain Inspired Cognitive Systems, pp. 106–119 (2018)

40. Haris, M., Shakhnarovich, G., Ukita, N.: Deep back-projection networks for super-resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1664–1673 (2018)

41. Hayat, K.: Multimedia super-resolution via deep learning: a survey. Digital Signal Processing (2018)

42. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 770–778 (2016)

43. Hong, C., Yu, J., Tao, D., Wang, M.: Image-based three-dimensional human pose recovery by multiview locality-sensitive sparse retrieval. IEEE Trans. Ind. Electr. **62**(6), 3742–3751 (2015)

44. Hong, C., Yu, J., Wan, J., Tao, D., Wang, M.: Multimodal deep autoencoder for human pose recovery. IEEE Trans. Image Process. **24**(12), 5659–5670 (2015)

45. Hu, Y., Li, J., Huang, Y., Gao, X.: Channel-wise and spatial feature modulation network for single image super-resolution. CoRR arXiv:1809.11130 (2018)

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

47. Huang, Y., Shao, L., Frangi, A.F.: Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. arXiv preprint arXiv:1705.02596 (2017)

48. Hung, K.W., Siu, W.C.: New motion compensation model via frequency classification for fast video super-resolution. In: Proceedings of IEEE International Conference on Image Processing, pp. 1193–1196 (2009)

49. Ji, H., Fermüller, C.: Robust wavelet-based super-resolution reconstruction: theory and algorithm. IEEE Trans. Pattern Anal. Mach. Intell. **31**(4), 649–660 (2009)

50. Johnson, J., Alahi, A., Fei-Fei, L.: Perceptual losses for real-time style transfer and super-resolution. In: Proceedings of European Conference on Computer Vision, pp. 694–711 (2016)

51. Joshi, M.V., Chaudhuri, S., Panuganti, R.: Super-resolution imaging: use of zoom as a cue. Image Vis. Comput. **22**(14), 1185–1196 (2004)

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

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

54. Krämer, P., Benois-Pineau, J., Domenger, J.P.: Local object-based super-resolution mosaicing from low-resolution video. Signal Process. **91**(8), 1771–1780 (2011)

55. Krizhevsky, A., Sutskever, I., Hinton, G.E.: Imagenet classification with deep convolutional neural networks. In: Proceedings of Advances in neural information processing systems, pp. 1097–1105 (2012)

56. Lai, W., Huang, J., Ahuja, N., Yang, M.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 5835–5843 (2017)

57. Lai, W., Huang, J., Ahuja, N., Yang, M.: Fast and accurate image super-resolution with deep laplacian pyramid networks. CoRR arXiv:abs/1710.01992 (2017)

58. Lai, W.S., Huang, J.B., Ahuja, N., Yang, M.H.: Deep laplacian pyramid networks for fast and accurate super-resolution. In: Proceedings of Computer Vision and Pattern Recognition (2017)

59. LeCun, Y., Bottou, L., Bengio, Y., Haffner, P., et al.: Gradient-based learning applied to document recognition. Proc. IEEE **86**(11), 2278–2324 (1998)

60. Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., Aitken, A.P., Tejani, A., Totz, J., Wang, Z., Shi, W.: Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition, pp. 105–114 (2017)

61. Li, J., Feng, J., Kuo, C.C.J.: Deep convolutional neural network for latent fingerprint enhancement. Signal Process. Image Commun. **60**, 52–63 (2018)

62. Li, K., Zhu, Y., Yang, J., Jiang, J.: Video super-resolution using an adaptive superpixel-guided auto-regressive model. Pattern Recognit. **51**, 59–71 (2016)

63. Li, X., Orchard, M.T.: New edge-directed interpolation. IEEE Trans. Image Process. **10**(10), 1521–1527 (2001)

64. Li, Y., Agustsson, E., Gu, S., Timofte, R., Van Gool, L.: Carn: convolutional anchored regression network for fast and accurate single image super-resolution. In: Proceedings of European Conference on Computer Vision, pp. 166–181 (2018)

65. Lim, B., Son, S., Kim, H., Nah, S., Lee, K.M.: Enhanced deep residual networks for single image super-resolution. In: Proceedings of 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 1132–1140 (2017)

66. Lin, M., Chen, Q., Yan, S.: Network in network. CoRR arXiv:abs/1312.4400 (2013)

67. Mao, X., Shen, C., Yang, Y.: Image restoration using convolutional auto-encoders with symmetric skip connections. CoRR arXiv:abs/1606.08921 (2016)

68. Martin, D., Fowlkes, C., Tal, D., Malik, J.: A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. Proc. Int. Conf. Comput. Vis. **2**, 416–423 (2001)

69. Metz, L., Poole, B., Pfau, D., Sohl-Dickstein, J.: Unrolled generative adversarial networks. CoRR arXiv:abs/1611.02163 (2016)

70. Milanfar, P.: Super-resolution imaging. CRC Press, Boca Raton (2010)

71. Minaee, S., Abdolrashidi, A.: Iris-gan: Learning to generate realistic iris images using convolutional gan. CoRR arXiv:abs/1812.04822 (2018)

72. Minetto, R., Thome, N., Cord, M., Leite, N.J., Stolfi, J.: Snoopertext: a text detection system for automatic indexing of urban scenes. Comput. Vis. Image Underst. **122**, 92–104 (2014)

73. Nasrollahi, K., Escalera, S., Rasti, P., Anbarjafari, G., Baro, X., Escalante, H.J., Moeslund, T.B.: Deep learning based super-resolution for improved action recognition. In: Proceedings of Image Processing Theory, Tools and Applications, pp. 67–72 (2015)

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

75. Nguyen, K., Fookes, C., Sridharan, S., Tistarelli, M., Nixon, M.: Super-resolution for biometrics: a comprehensive survey. Pattern Recognit. **78**, 23–42 (2018)

76. Nguyen, T., Le, T., Vu, H., Phung, D.Q.: Dual discriminator generative adversarial nets. In: Proceedings of Advances in Neural Information Processing Systems, pp. 2670–2680 (2017)

77. Park, S.C., Park, M.K., Kang, M.G.: Super-resolution image reconstruction: a technical overview. IEEE Signal Process. Mag. **20**(3), 21–36 (2003)

78. Patti, A.J., Sezan, M.I., Tekalp, A.M.: Super-resolution video reconstruction with arbitrary sampling lattices and nonzero aperture time. IEEE Trans. Image Process. **6**(8), 1064–1076 (1997)

79. Pérez-Pellitero, E., Salvador, J., Ruiz-Hidalgo, J., 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)

80. Petrou, M., Jaward, M.H., Chen, S., Briers, M.: Super-resolution in practice: the complete pipeline from image capture to super-resolved subimage creation using a novel frame selection method. Mach. Vis. Appl. **23**(3), 441–459 (2012)

81. Qi, L., Wang, R., Hu, C., Li, S., He, Q., Xu, X.: Time-aware distributed service recommendation with privacy-preservation. Inf. Sci. **480**, 354–364 (2019)

82. Qi, L., Zhang, X., Dou, W., Hu, C., Yang, C., Chen, J.: A two-stage locality-sensitive hashing based approach for privacy-preserving mobile service recommendation in cross-platform edge environment. Future Gener. Comput. Syst. **88**, 636–643 (2018)

83. Qi, L., Zhang, X., Dou, W., Ni, Q.: A distributed locality-sensitive hashing-based approach for cloud service recommendation from multi-source data. IEEE J. Sel. Areas Commun. **35**(11), 2616–2624 (2017)

84. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. CoRR arXiv:abs/1511.06434 (2015)

85. Radford, A., Metz, L., Chintala, S.: Unsupervised representation learning with deep convolutional generative adversarial networks. In: Proceedings of 4th International Conference on Learning Representations (2016)

86. Robinson, M.D., Chiu, S.J., Toth, C.A., Izatt, J.A., Lo, J.Y., Farsiu, S.: New applications of super-resolution in medical imaging. In: Super-Resolution Imaging, pp. 401–430. CRC Press (2017)

87. Roweis, S.T., Saul, L.K.: Nonlinear dimensionality reduction by locally linear embedding. Science **290**(5500), 2323–2326 (2000)

88. Russakovsky, O., Deng, J., Su, H., Krause, J., Satheesh, S., Ma, S., Huang, Z., Karpathy, A., Khosla, A., Bernstein, M., et al.: Imagenet large scale visual recognition challenge. Int. J. Comput. Vis. **115**(3), 211–252 (2015)

89. Ryoo, M.S., Rothrock, B., Fleming, C., Yang, H.J.: Privacy-preserving human activity recognition from extreme low resolution. In: Proceedings of AAAI Conference on Artificial Intelligence, pp. 4255–4262 (2017)

90. Sajjadi, M.S., Schölkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. In: Proceedings of IEEE International Conference on Computer Vision, pp. 4501–4510. IEEE (2017)

91. Sajjadi, M.S.M., Schölkopf, B., Hirsch, M.: Enhancenet: Single image super-resolution through automated texture synthesis. In: Proceedings of IEEE International Conference on Computer Vision, pp. 4501–4510 (2017)

92. Salimans, T., Goodfellow, I.J., Zaremba, W., Cheung, V., Radford, A., Chen, X.: Improved techniques for training gans. In: Proceedings of Neural Information Processing Systems, pp. 2226–2234 (2016)

93. Salvador, J., Perez-Pellitero, E.: Naive bayes super-resolution forest. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 325–333 (2015)

94. Schulter, S., Leistner, C., 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)

95. Schultz, R.R., Stevenson, R.L.: A bayesian approach to image expansion for improved definition. IEEE Trans. Image Process. **3**(3), 233–242 (1994)

96. Shamsolmoali, P., Zhang, J., Yang, J., et al.: Image super resolution by dilated dense progressive network. Image Vis. Comput. **88**, 9–18 (2019)

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

98. Sheikh, H.R., Wang, Z., Cormack, L., Bovik, A.C.: Live image quality assessment database release **2**, 2005 (2005)

99. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A.P., Bishop, R., Rueckert, D., Wang, Z.: Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1874–1883 (2016)

100. Simonyan, K., Zisserman, A.: Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014)

101. Singh, K., Gupta, A., Kapoor, R.: Fingerprint image super-resolution via ridge orientation-based clustered coupled sparse dictionaries. J. Electr. Imaging **24**(4), 043015 (2015)

102. Song, H., Zhang, L., Wang, P., Zhang, K., Li, X.: An adaptive l 1–l 2 hybrid error model to super-resolution. In: Proceedings of IEEE International Conference on Image Processing, pp. 2821–2824 (2010)

103. Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., Rabinovich, A.: Going deeper with convolutions. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1–9 (2015)

104. Tai, Y., Yang, J., Liu, X.: Image super-resolution via deep recursive residual network. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 2790–2798 (2017)

105. Takeda, H., Milanfar, P., Protter, M., Elad, M.: Super-resolution without explicit subpixel motion estimation. IEEE Trans. Image Process. **18**(9), 1958–1975 (2009)

106. Tian, J., Ma, K.K.: Stochastic super-resolution image reconstruction. J. Vis. Commun. Image Represent. **21**(3), 232–244 (2010)

107. Tian, J., Ma, K.K.: A survey on super-resolution imaging. Signal Image Video Process. **5**(3), 329–342 (2011)

108. Timofte, R., Agustsson, E., Van Gool, L., Yang, M.H., Zhang, L.: Ntire 2017 challenge on single image super-resolution: Methods and results. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 114–125 (2017)

109. Timofte, R., De Smet, V., Van Gool, L.: Anchored neighborhood regression for fast example-based super-resolution. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 1920–1927 (2013)

110. Timofte, R., De Smet, V., Van Gool, L.: A+: Adjusted anchored neighborhood regression for fast super-resolution. In: Proceedings of Asian Conference on Computer Vision, pp. 111–126 (2014)

111. Timofte, R., Rothe, R., Van Gool, L.: Seven ways to improve example-based single image super resolution. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, pp. 1865–1873 (2016)

112. Tong, T., Li, G., Liu, X., Gao, Q.: Image super-resolution using dense skip connections. In: Proceedings of IEEE International Conference on Computer Vision, pp. 4809–4817 (2017)

113. Trinh, D.H., Luong, M., Dibos, F., Rocchisani, J.M., Pham, C.D., Nguyen, T.Q.: Novel example-based method for super-resolution and denoising of medical images. IEEE Trans. Image Process. **23**(4), 1882–1895 (2014)

114. Wallach, D., Lamare, F., Kontaxakis, G., Visvikis, D.: Super-resolution in respiratory synchronized positron emission tomography. IEEE Trans. Med. Imaging **31**(2), 438–448 (2012)

115. Wang, X., Tang, X.: Hallucinating face by eigentransformation. IEEE Trans. Syst. Man Cybern. Part C **35**(3), 425–434 (2005)

116. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., Qiao, Y., Loy, C.C.: ESRGAN: enhanced super-resolution generative adversarial networks. In: Proceedings of European Conference on Computer Vision Workshops, pp. 63–79 (2018)

117. Wang, Y., Perazzi, F., McWilliams, B., Sorkine-Hornung, A., Sorkine-Hornung, O., Schroers, C.: A fully progressive approach to single-image super-resolution. In: Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 864–873 (2018)

118. Wang, Y.H., Qiao, J., Li, J.B., Fu, P., Chu, S.C., Roddick, J.F.: Sparse representation-based mri super-resolution reconstruction. Measurement **47**, 946–953 (2014)

119. Wang, Z., Liu, D., Yang, J., Han, W., Huang, T.: Deep networks for image super-resolution with sparse prior. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 370–378 (2015)

120. Wu, H., Zheng, S., Zhang, J., Huang, K.: GP-GAN: towards realistic high-resolution image blending. CoRR arXiv:abs/1703.07195 (2017)

121. Wu, Y., Shivakumara, P., Lu, T., Tan, C.L., Blumenstein, M., Kumar, G.H.: Contour restoration of text components for recognition in video/scene images. IEEE Trans. Image Process. **25**(12), 5622–5634 (2016)

122. Wu, Y., Shivakumara, P., Wei, W., Lu, T., Pal, U.: A new ring radius transform-based thinning method for multi-oriented video characters. Int. J. Doc. Anal. Recognit. **18**(2), 137–151 (2015)

123. Xu, X., Fu, S., Qi, L., Zhang, X., Liu, Q., He, Q., Li, S.: An iot-oriented data placement method with privacy preservation in cloud environment. J. Netw. Comput. Appl. **124**, 148–157 (2018)

124. Xu, X., Ma, Y., Sun, W.: Towards real scene super-resolution with raw images. CoRR arXiv:abs/1905.12156 (2019)

125. Xu, X., Zhang, X., Khan, M., Dou, W., Xue, S., Yu, S.: A balanced virtual machine scheduling method for energy-performance trade-offs in cyber-physical cloud systems. Future Generation Computer Systems (2017)

126. Xu, X., Zhao, X., Ruan, F., Zhang, J., Tian, W., Dou, W., Liu, A.X.: Data placement for privacy-aware applications over big data in hybrid clouds. Secur. Commun. Netw. **2017**, 2376484:1–2376484:15 (2017)

127. Xu, Y., Qi, L., Dou, W., Yu, J.: Privacy-preserving and scalable service recommendation based on simhash in a distributed cloud environment. Complexity **2017**, 3437854:1–3437854:9 (2017)

128. Yamanaka, J., Kuwashima, S., Kurita, T.: Fast and accurate image super resolution by deep cnn with skip connection and network in network. In: Proceedings of Neural Information Processing, pp. 217–225 (2017)

129. Yang, C.Y., Yang, M.H.: Fast direct super-resolution by simple functions. In: Proceedings of the IEEE International Conference on Computer Vision, pp. 561–568 (2013)

130. Yang, J., Wright, J., Huang, T.S., Ma, Y.: Image super-resolution via sparse representation. IEEE Trans. Image Process. **19**(11), 2861–2873 (2010)

131. Ye, Q., Doermann, D.S.: Text detection and recognition in imagery: a survey. IEEE Trans. Pattern Anal. Mach. Intell. **37**(7), 1480–1500 (2015)

132. Yu, X., Fernando, B., Ghanem, B., Porikli, F., Hartley, R.: Face super-resolution guided by facial component heatmaps. In: Proceedings of the European Conference on Computer Vision (ECCV), pp. 217–233 (2018)

133. Yuan, Y., Liu, S., Zhang, J., Zhang, Y., Dong, C., Lin, L.: Unsupervised image super-resolution using cycle-in-cycle generative adversarial networks. In: Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition Workshops, pp. 701–710 (2018)

134. Yuan, Z., Wu, J., Kamata, S.i., Ahrary, A., Yan, P.: Fingerprint image enhancement by super resolution with early stopping. In: Proceedings of IEEE International Conference on Intelligent Computing and Intelligent Systems, vol. 4, pp. 527–531 (2009)

135. Yue, L., Shen, H., Jie, L., Yuan, Q., Zhang, H., Zhang, L.: Image super-resolution: the techniques, applications, and future. Signal Process. **128**, 389–408 (2016)

136. Zareapoor, M., Zhou, H., Yang, J.: Perceptual image quality using dual generative adversarial network. Neural Computing and Applications pp. 1–11 (2019)

137. Zeyde, R., Elad, M., Protter, M.: On single image scale-up using sparse-representations. In: Proceedings of International Conference on Curves and Surfaces, pp. 711–730 (2010)

138. Zhang, H., Yang, Z., Zhang, L., Shen, H.: Super-resolution reconstruction for multi-angle remote sensing images considering resolution differences. Rem. Sens. **6**(1), 637–657 (2014)

139. Zhang, H., Zhang, L., Shen, H.: A super-resolution reconstruction algorithm for hyperspectral images. Signal Process. **92**(9), 2082–2096 (2012)

140. Zhang, J., Pu, J., Chen, C., Fleischer, R.: Low-resolution gait recognition. IEEE Trans. Syst. Man Cybern. Part B **40**(4), 986–996 (2010)

141. Zhang, K., Zuo, W., Zhang, L.: Learning a single convolutional super-resolution network for multiple degradations. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3262–3271 (2018)

142. Zhang, Y., Li, K., Li, K., Wang, L., Zhong, B., Fu, Y.: Image super-resolution using very deep residual channel attention networks. In: Proceedings of the European Conference on Computer Vision, pp. 286–301 (2018)

143. Zhang, Y., Liu, S., Dong, C., Zhang, X., Yuan, Y.: Multiple cycle-in-cycle generative adversarial networks for unsupervised image super-resolution. IEEE transactions on Image Processing: (2019)

144. Zhang, Y., Tian, Y., Kong, Y., Zhong, B., Fu, Y.: Residual dense network for image super-resolution. In: Proceedings of 2018 IEEE Conference on Computer Vision and Pattern Recognition, pp. 2472–2481 (2018)

145. Zhao, S., Han, H., Peng, S.: Wavelet-domain hmt-based image super-resolution. In: Proceedings os International Conference on Image Processing, vol. **2**, p. 953 (2003)

146. Zhao, T., Ren, W., Zhang, C., Ren, D., Hu, Q.: Unsupervised degradation learning for single image super-resolution. CoRR arXiv:abs/1812.04240 (2018)

147. Zheng, H., Wang, X., Gao, X.: Fast and accurate single image super-resolution via information distillation network. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 723–731 (2018)

148. Zheng, Z., Zheng, L., Yang, Y.: Unlabeled samples generated by gan improve the person re-identification baseline in vitro. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition, pp. 3754–3762 (2017)

149. Zhu, J., Park, T., Isola, P., Efros, A.A.: Unpaired image-to-image translation using cycle-consistent adversarial networks. In: Proceedings of IEEE International Conference on Computer Vision, pp. 2242–2251 (2017)

150. Zhu, J.Y., Krähenbühl, P., Shechtman, E., Efros, A.A.: Generative visual manipulation on the natural image manifold. In: Proceedings of European Conference on Computer Vision, pp. 597–613. Springer (2016)

151. Zhuang, Y., Zhang, J., Wu, F.: Hallucinating faces: Lph super-resolution and neighbor reconstruction for residue compensation. Pattern Recognit. **40**(11), 3178–3194 (2007)

Xiaofang Li received the B.S. degree in computer science and technology the M.Sc. degree in power system automation and the Ph.D. degree in hydro informatics from Hohai University, in 1995, 2004, and 2011, respectively. She is currently an Associate Professor with the College of Computer and Information Engineering, Hohai University, Nanjing, China. She is also an Associate Professor with the Changzhou Institute of Technology, Chanzhou, China. Her research areas include information acquisition and wireless sensor networks. She is a Senior Member of the China Computer Federation.

Yirui Wu received the B.S., M.Sc. and the Ph.D. degree in computer science and technology from Nanjing University, in 2011, 2013, and 2016, respectively. He is currently an Assistant Professor with the College of Computer and Information Engineering, Hohai University, Nanjing, China. His research areas include pattern recognition, multimedia system and smart hydrology modeling.

Wen Zhang is currently a master student in Hohai University, Nanjing, China. Her research interests include multimedia computing and data-driven hydrology modeling.

Ruichao Wang is now working in School of Computer Science, University College Dublin, Ireland. His research interests include machine learning technologies and natural language processing.

Feng Hou is now working in School of Natural and Computational Science, Massey University, New Zealand. His research interests include wireless sensor network and machine learning modeling.