doi:10.3772/j.issn.1006-6748.2021.03.008

A brief survey on deep learning based image super-resolution①

Zhu Xiaobin (祝晓斌)*, Li Shanshan**, Wang Lei②***
(* Department of Computer Science and Technology, University of Science and Technology Beijing, Beijing 100048, P. R. China)
(** School of Computer and Information Engineering, Beijing Technology and Business University, Beijing 100048, P. R. China)
(*** Academy of Broadcasting Science, National Radio and Television Administration, Beijing 100866, P. R. China)

Abstract

Image super-resolution (SR) is an important technique for improving the resolution and quality of images. With the great progress of deep learning, image super-resolution achieves remarkable improvements recently. In this work, a brief survey on recent advances of deep learning based single image super-resolution methods is systematically described. The existing studies of SR techniques are roughly grouped into ten major categories. Besides, some other important issues are also introduced, such as publicly available benchmark datasets and performance evaluation metrics. Finally, this survey is concluded by highlighting four future trends.

Key words: image super-resolution(SR), deep learning, convolutional neural network(CNN)

0 Introduction

Image super-resolution (SR) aims to transform a low-resolution (LR) image with coarse details into a counterpart high-resolution (HR) version with refined details and improved visual quality. With HR images various tasks in computer vision may enhance their performance, hence SR has been successfully introduced into many important tasks, such as object detection[1-2], face recognition[3-5], medical imaging[6], astronomical images[7]. However, as a typical ill-posed problem, image SR is challenging and there are many problems that remain to be solved.

According to the type of features, SR methods can be roughly divided into two categories: traditional methods and deep learning methods. Traditional methods adopt hand-craft features or independent feature learning processes, and their performances are always restricted by the quality of extracted feature. Recently, with the rapid development of deep convolutional neural networks (CNNs)[8-10], deep learning-based SR methods have achieved promising performances[11-15] and attracted ever-increasing attention from industrial and research communities. This work concentrates on deep learning methods[16].

The main contributions of this survey are threefold:

(1) A comprehensive review of deep learning-based single image SR techniques is described, including problem definitions, datasets, assessment metrics, representative algorithms, experimental comparisons.

(2) Providing systematic and extensive evaluations on publicly available image super-resolution datasets.

(3) The challenges, open issues, the new trends, and future directions are all discussed for providing insights on possible future directions.

1 Problem setting

1.1 Problem definitions

SR aims to convert an LR image into an HR image. In the process, the number of pixels in the input LR image is increased with required scaling factor, as shown in Fig. 1. The reconstruction strategies in existing image super-resolution methods concentrate on exploring prior information or deducting by rules to train an optimal model that can build amplified images from available LR images. Therefore, the reconstruction is reflected as the problem in an inverse route to calculate the original details around the geometrical symmetries of the SR image by merging one or more LR images.

1.2 Datasets

A series of classic datasets for image SR tasks have been proposed, among which some datasets pro-

① Supported by the National Key Research and Development Program of China (No. 2019YFB1405900).
② To whom correspondence should be addressed. E-mail: wanglei@abs.ac.cn
Received on July 27, 2020

vide LR-HR image pairs, and others only provide individual HR images. Representative benchmark datasets are Set5[17], Set14[18], BSD100[19], Urban100[20], DIV2K[13], Manga109[21], Flickr2K[14], OST300[22], PIRM[23]. Set5 and Set14 consist of 5 and 14 testing images, respectively. BSD100 contains 100 testing images from natural images and specific objects. Urban100 is a relatively new dataset taken from urban scenes. DIV2K is originally constructed for NITRE (new trends in image restoration and enhancement) challenges, and it contains 800 images for training and 100 images for testing and validation. PIRM consists of 200 images, covering diverse contents, including people, objects, etc.

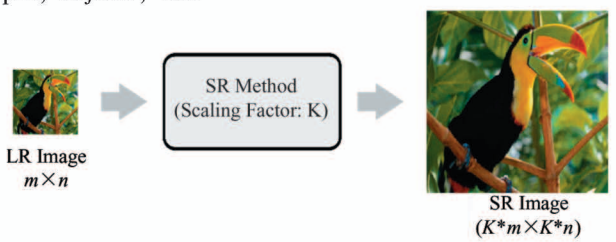

Fig. 1 Illustration of image super-resolution

1.3 Assessment

Image quality assessment (IQA) refers to determine the important visual attributes of images, focusing on the evaluation of human perception. IQA methods can be divided into subjective methods and objective methods. And the subjective and objective methods may be bit consistent, for the latter often cannot accurately capture human visual perception[24-25]. In this section, the five commonly used IQA metrics are briefly introduced.

1.3.1 Peak signal-to-noise ratio

In SR task, peak signal-to-noise ratio (PSNR) is calculated by the maximum possible pixel value (denoted as L) and mean squared error (MSE) between a ground-truth image I and a reconstructed image $\hat{I}$, which are formulated as

$$MSE = \frac{1}{N}\sum_{i=1}^{N}(I(i) - \hat{I}(i))^2 \tag{1}$$

$$PSNR = 10 \cdot \log_{10}\left(\frac{L^2}{MSE}\right) \tag{2}$$

where L is 255, N denotes the number of pixels. Since PSNR only relate to pixel-level MSE without visual perception, it may lead to poor reconstruction quality.

1.3.2 Structural similarity

The structural similarity index (SSIM)[25] is proposed for measuring the structural similarity between images in terms of luminance, contrast, and structures. Due to the characteristic of HVS, SSIM can well meet the requirements of perceptual assessment[26] and is widely used in SR tasks. SSIM can briefly be formulated as

$$SSIM(I, \hat{I}) = [C_l(I, \hat{I})]^{\alpha}[C_C(I, \hat{I})]^{\beta}[C_S(I, \hat{I})]^{\gamma} \tag{3}$$

where α, β, γ are control parameters for adjusting importance, $C_l(\cdot)$, $C_C(\cdot)$ and $C_S(\cdot)$ are comparisons on luminance, contrast, and structures, respectively.

1.3.3 Mean opinion score

Mean opinion score (MOS) is a commonly used subjective IQA metric by human raters to determine a perceptual image quality score (generally ranges from 1-poor to 5-excellent). The final MOS is calculated as a mean of all rating scores. It is often discouraged by biases and variance of rating criteria, differences between the subjective views of different raters. In reality, some SR models perform poorly in PSNR or SSIM, but achieves good perceptual quality in MOS[27-31].

1.3.4 Learned perceptual image patch similarity

As a recently introduced full-reference IQA metric, the learned perceptual image patch similarity (LPIPS)[32] adopts linear deep classification networks to measure perceptual similarity. These networks are often trained on the BAPPS[32], which contains human perceptual judgments. The LPIPS is consistent in many models that have been trained to improve PSNR, SSIM scores, or those trained to improve the quality of perception. Hence, LPIPS can well balance objective evaluation (PSNR, SSIM) and subjective evaluation (MOS).

1.3.5 Natural image quality evaluator

Natural image quality evaluator (NIQE)[33] computes 36 identical natural scene statistic (NSS) features from image patches with the same size to analysis quality, fitting them with the multivariate Gaussian (MVG) model. The sharpness criterion is not applied to these patches. The quality of distorted images is expressed as the distance between the quality-aware NSS feature model and the MVG features extracted from the distorted image as

$$D(v_1, v_2, \Sigma_1, \Sigma_2) = \sqrt{(v_1 - v_2)^{\mathrm{T}}\left(\frac{\Sigma_1 + \Sigma_2}{2}\right)^{-1}(v_1 - v_2)} \tag{8}$$

where v_1, v_2 and Σ_1, Σ_2 are the mean vectors and covariance matrices of the natural MVG model and the distorted image's MVG model.

2 Deep learning based image SR

Various image SR algorithms based on deep learnings have shown impressive performance. According to model designs, the existing SR algorithms are roughly

divided into 10 categories.

2.1 Direct network

This type of method mainly consists of a single path without complex skip connections. In these networks, several convolution layers are often simply stacked[34-39]. SRCNN[35] is a pioneer deep learning based SR work which is mainly composed of three convolutional layers. Different from the shallow network structure of SRCNN, VDSR[40] is commonly referred as a deep VGG network and uses fixed-size convolutions (3×3) in all layers. The architecture of DnCNN[36] is simple and is similar to SRCNN, it only stacks convolutional layers, batch normalization layers, and ReLU layers. IRCNN[37] combines discriminative CNN networks with model-based optimization methods. FSRCNN[38] consists of four convolution layers and one deconvolution layer.

2.2 Residual learning based methods

Residual learning adopts local or global skip connections, which can benefit training and feature learning[14-15,41-42]. CARN[41] employs ResNet blocks[15] to learn the relationship between LR input and HR output. FormResNet[43] presents a formatted residual learning framework for image restoration. BTSRN[44] proposes a novel balanced two-stage residual networks with lightweight yet efficient two-layer residual blocks. REDNet[9] proposes a deep encoding and decoding framework for image restoration, in which convolution and deconvolution are combined for extracting primary image content and recovering details. HCNN[45] hierarchically assembles shallow CNNs with deep CNNs for effective image SR.

2.3 Dense connection based methods

Inspired by the success of the DenseNet[46], SR algorithms adopt densely connected CNN layers to improve performance. SRDenseNet[47] is directly constructed on DenseNet. In RDN[48], residual dense block (RDB) serves as a basic build module. In each RDB, the dense connections between each layer allow full usage of local layers. BBDP[49] iteratively performs back projections to explore the feedback error for improving texture details. Its motivation is that only feedforward mechanism is not optimal for modelling the mapping from LR to HR images. ESRN[50] adopts an efficient residual dense block to construct a fast, lightweight, and accurate super-resolution network.

2.4 Recursive network based methods

Recursive networks[51-52] either employ recursively convolutional layers or recursively link units for progressively breaking down the complex SR problem into a set of simple ones. DRCN[53] efficiently reuses weight parameters while exploiting large image context. To ease the difficulty of training the model, DRCN uses recursive-supervision and skip connection. As DRCN, DRRN[51] utilizes recursive learning, which replicates a basic skip-connection block to achieve a multi-path network block. In MemNet[52], a memory block adopts a gating mechanism for tackling the long-term dependency problem. SRFBN[54] enhances low-level representations by exploring high-level semantic information, in which a feedback block effectively handles feedback information as well as the feature reuse.

2.5 Progressive design based methods

Facing large scaling factors, SR algorithms[55] often progressively predict output image by multiple steps. SCN[56] combines a sparse coding model particularly designed for super-resolution with a neural network and trained in a cascaded structure from end to end. LapSRN[55] reconstructs the sub-band residuals of high-resolution images through the progressive reconstruction. CMSC[57] models the super-resolution reconstruction by a sequence of cascaded sub-networks to gradually refine high resolution features with cascaded-supervision in a coarse-to-fine manner.

2.6 Multi-branch network based methods

Multi-branch networks aim to obtain a diverse set of features at multiple context scales. CNF[58] aims to fuse multiple CNNs for image SR. IDN[59] consists of feature extraction block, stacked information distillation blocks, and reconstruction block. EBRN[60] advocates that the limitation of existing methods is caused by under-fitting of the models on complex textures and overfitting on simple structures. Hence, a block residual module is adopted to restore parts of the image information while passing the remained information to deeper layers. DRN[61] proposes a novel dual regression scheme for paired and unpaired data. SeaNet[62] adopts an image reconstruction network, a soft-edge reconstruction network, and a refinement network.

2.7 Attention mechanism based methods

Combined with deep networks, attention-based SR methods have shown promising performance. SelNet[63] proposes a novel selection unit, which is a multiplication of an identity mapping and a sigmoid-based selection module. In RCAN[64], a residual in residual (RIR) structure combined with channel attention (CA) mechanism is proposed to adaptively rescale

channel-wise features by considering interdependencies among channels. SRRAM[65] extensively evaluates a range of attention mechanisms with common SR architectures. In Ref. [66], authors explore the cross-scale patch recurrence property of a natural image through a cross-scale internal graph neural network.

2.8 Multiple degradation based methods

The majority of existing super-resolution methods only consider a single degradation. However, in reality, there often exist multiple types of degradations. ZSSR[67] only relies on a small image-specific CNN, which is trained at test time on internal examples extracted solely from a LR test image. SRMD[68] has high scalability of handling multiple degradations by taking both a LR image and its degradation maps as input. Recently, UDVD[69] proposes a unified network to accommodate the variations from inter-image (cross-image variations) and intra-image (spatial variations).

2.9 GAN based methods

GAN[70-72] based methods employ a game-theoretic, the generator tries to generate SR image that the discriminator cannot distinguish as a real HR image or an artificially super-resolved one. SRGAN[24] is a pioneer GAN-based network for optimizing a new perceptual loss. EnhanceNet[73] achieves promising results by a combination of adversarial training, perceptual losses, and a newly proposed texture transfer loss. SRFeat[11] produces perceptually pleasing images by employing an image discriminator together with a feature discriminator. The feature discriminator encourages the generator to generate high-frequency details instead of noisy artifacts. ESRGAN[12] achieved Top-1 rank in the PIRM-SR Challenge. In ESRGAN, the discriminator relativistic GAN learns to judge whether one image is more realistic than another, guiding the generator to recover more detailed textures.

2.10 Domain-specific applications

Intuitively, images with good quality and high resolution can facilitate visual related tasks. From this key observation, some work has explored the learning of SR for domain-specific applications[74-78]. Specifically, in Ref. [75], a clear high-resolution face is directly generated from a blurry small one by adopting a GAN for tiny face detection. In Ref. [77], authors propose a novel feature-level super-resolution approach that can be extended into any proposal based detectors. Zhang et al. [79] proposed two SR methods for the spatial and temporal streams, tailored for two-stream action recognition networks. Zhu et al. [80] addressed image classification by developing an end-to-end architecture that internally elevates representations of an LR image to "super-resolved" ones.

3 Experiments

Table 1 lists the 2 × SR results, and Table 2 lists 4 × SR results. Apparently, Bicubic has poor performance compared with the deep learning ones. As a pioneer deep learning work, SRCNN outperforms Bicubic by 2.98 dB PSNR and 0.0238 SSIM on Set5 in 2 × SR task. Although the experimental results of other direct networks, e.g., ESPCN and FSRCNN, are greatly improved in PSNR and SSIM, the details are still not promising. By deepening networks, these methods (EDSR[14] and FormResNet[43]) adopt residual learning, which is conducive for training and feature learning, to improve the details of reconstructed images. RDN[48] makes full use of hierarchical feature information and provides a lot of reference information for reconstruction. Consequently, its PSNR and SSIM are respectively 32.47 dB and 0.8990 on Set5 for 4 × SR. DRCN[53], DRRN[51] and MemNet[52] employ recursive learning to progressively break down the harder SR problem into a set of simpler ones for further boosting reconstruction performance. And they all achieve advanced PSNR and SSIM results in SR tasks. LapSRN[55] progressively predicts output image by multiple steps, which alleviates issues with undesired artifacts and greatly reduces the computational complexity. To exploit the features of multiple context scales, SeaNet[62] employs a multi-branch network to obtain complementary information and achieves 32.44 dB PSNR and 0.8981 SSIM results on Set5 for 4 × SR. RCAN[64] focuses on necessary and effective image information and achieves 38.27 dB PSNR and 0.9614 SSIM on Set5 for 2 × SR. The PSNR and SSIM of ZSSR[67], SRMD[68], and UDVD[69] on Set5 are not the best, but they can reconstruct visually satisfactory SR images for real-world images. Although SRGAN has relatively low PSNR and SSIM, its reconstruction results are visually promising. IRN[81] achieves the best PSNR and SSIM in 2 × SR, which are respectively 43.99 dB and 0.9871 on Set5. And in 4 × SR, its PSNR and SSIM are respectively 36.19 dB and 0.9451 on Set5. For 4 × SR, DRN[61] achieves the best PSNR and SSIM, showing the effectiveness of multi-scale feature learning and dual regression scheme.

In image super-resolution, it is difficult to compare methods, because there are many involved factors, such as network complexity, training data, patch size for training, number of features maps. In Table 3,

Table 1 PSNR and SSIM on the benchmark datasets for a scaling factor 2 ×

		Set5		Set14		BSD100		Urban100		DIV2K		Manga109	
Method	Category	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	-	33.68	0.9304	30.24	0.8691	29.56	0.8435	26.88	0.8405	32.45	0.9040	31.05	0.9350
SRCNN[35]	Direct Network	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.51	0.8946	34.59	0.9320	35.72	0.9680
VDSR[40]	Direct Network	37.53	0.9587	33.05	0.9127	31.90	0.8960	30.77	0.9141	35.43	0.9410	37.16	0.9740
DnCNN[36]	Direct Network	37.58	0.9590	33.03	0.9128	31.90	0.8961	30.74	0.9139	-	-	-	-
FSRCNN[38]	Direct Network	36.98	0.9556	32.62	0.9087	31.50	0.8904	29.85	0.9009	34.74	0.9340	36.62	0.9710
CARN[41]	Residual Learning	37.76	0.9590	33.52	0.9166	32.09	0.8978	31.92	0.9256	36.04	0.9451	38.36	0.9764
EDSR[14]	Residual Learning	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	35.03	0.9695	39.10	0.9773
MDSR[14]	Residual Learning	38.11	0.9602	33.85	0.9198	32.29	0.9007	32.84	0.9347	34.96	0.9692	38.96	0.9780
ERN[45]	Residual Learning	38.18	0.9610	33.88	0.9195	32.30	0.9011	32.66	0.9332	-	-	-	-
RDN[48]	Dense Connection	38.24	0.9614	34.01	0.9212	32.34	0.9017	32.89	0.9353	-	-	39.18	0.9780
ESRN[50]	Dense Connection	38.04	0.9607	33.71	0.9185	32.23	0.9005	32.37	0.9310	-	-	-	-
DRRN[51]	Recursive Network	37.74	0.9591	33.23	0.9136	32.05	0.8973	31.23	0.9188	35.63	0.9410	37.92	0.9760
MemNet[52]	Recursive Network	37.78	0.9597	33.28	0.9142	32.08	0.8978	31.31	0.9195	-	-	37.72	0.9740
SRFBN[54]	Recursive Network	38.11	0.9609	33.82	0.9196	32.29	0.9010	32.62	0.9328	-	-	39.08	0.9779
DRCN[53]	Recursive Network	37.63	0.9588	33.06	0.9121	31.85	0.8942	30.76	0.9133	35.45	0.9400	37.57	0.9730
SCN[56]	Progressive Design	36.52	0.9530	32.42	0.9040	31.24	0.8840	29.50	0.8960	34.98	0.9370	35.51	0.9670
D-DBPN[49]	Progressive Design	38.09	0.9600	33.85	0.9190	32.27	0.9000	32.55	0.9324	-	-	38.89	0.9775
LapSRN[55]	Progressive Design	37.52	0.9591	32.99	0.9124	31.80	0.8949	30.41	0.9101	35.31	0.9400	37.53	0.9740
IDN[59]	Multibranch Network	37.83	0.9600	33.30	0.9148	32.08	0.8985	31.27	0.9196	-	-	38.02	0.9749
IRN[81]	Multibranch Network	43.99	0.9871	40.79	0.9778	41.32	0.9876	39.92	0.9865	44.32	0.9908	-	-
SeaNet[62]	Multibranch Network	38.15	0.9611	33.86	0.9198	32.31	0.9013	32.68	0.9332	-	-	38.97	0.9779
RCAN[64]	Attention Mechanism	38.27	0.9614	34.12	0.9216	32.41	0.9027	33.34	0.9384	36.63	0.9491	39.44	0.9786
ZSSR[67]	Multiple Degradation	37.37	0.9570	33.00	0.9108	31.65	0.8920	-	-	-	-		-
SRMDNF[68]	Multiple Degradation	37.79	0.9601	33.32	0.9159	32.05	0.8985	31.33	0.9204	35.54	0.9414	38.07	0.9761

Table 2 PSNR and SSIM on the benchmark datasets for a scaling factor 4 ×

		Set5		Set14		BSD100		Urban100		DIV2K		Manga109	
Method	Category	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	-	28.43	0.8109	26.00	0.7023	25.96	0.6678	23.14	0.6574	28.11	0.7750	25.15	0.7890
SRCNN[35]	Direct Network	30.48	0.8628	27.50	0.7513	26.90	0.7103	24.52	0.7226	29.33	0.8090	27.66	0.8580
VDSR[40]	Direct Network	31.35	0.8838	28.02	0.7678	27.29	0.7252	25.18	0.7525	29.82	0.8240	28.82	0.8860
DnCNN[36]	Direct Network	31.40	0.8845	28.04	0.7672	27.29	0.7253	25.20	0.7521	-	-	-	-
FSRCNN[38]	Direct Network	30.70	0.8657	27.59	0.7535	26.96	0.7128	24.60	0.7258	29.36	0.8110	27.89	0.8590
CARN[41]	Residual Learning	32.13	0.8937	28.60	0.7806	27.58	0.7349	26.07	0.7837	30.43	0.8374	30.40	0.9082
EDSR[14]	Residual Learning	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	29.25	0.9017	31.02	0.9148
MDSR[14]	Residual Learning	32.50	0.8973	28.72	0.7857	27.72	0.7418	26.67	0.8041	29.26	0.9016	31.11	0.9150
ERN[45]	Residual Learning	32.39	0.8975	28.75	0.7853	27.70	0.7398	26.43	0.7966	-	-	-	-
RDN[48]	Dense Connection	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028	-	-	31.00	0.9151
ESRN[50]	Dense Connection	32.26	0.8957	28.63	0.7818	27.62	0.7378	26.24	0.7912	-	-	-	-
DRRN[51]	Recursive Network	31.68	0.8888	28.21	0.7720	27.38	0.7284	25.44	0.7638	29.98	0.8270	29.46	0.8960
MemNet[52]	Recursive Network	31.74	0.8893	28.26	0.7723	27.40	0.7281	25.50	0.7630	-	-	29.42	0.8942
SRFBN[54]	Recursive Network	32.47	0.8983	28.81	0.7868	27.72	0.7409	26.60	0.8015	-	-	31.15	0.9160
DRCN[53]	Recursive Network	31.53	0.8854	28.03	0.7673	27.24	0.7233	25.14	0.7511	29.83	0.8230	28.97	0.8860
SCN[56]	Progressive Design	30.39	0.8620	27.48	0.7510	26.87	0.7100	24.52	0.7250	29.47	0.8130	27.39	0.8570
D-DBPN[49]	Progressive Design	32.47	0.8980	28.82	0.7860	27.72	0.7400	26.38	0.7946	-	-	30.91	0.9137
LapSRN[55]	Progressive Design	31.54	0.8866	28.09	0.7694	27.32	0.7264	25.21	0.7553	29.88	0.8250	29.09	0.8900
IDN[59]	Multibranch Network	31.82	0.8903	28.25	0.7730	27.41	0.7297	25.41	0.7632	-	-	29.40	0.8936
DRN[61]	Multibranch Network	32.74	0.9020	28.98	0.7920	27.83	0.7450	27.03	0.8130	-	-	31.73	0.9220
IRN[81]	Multibranch Network	36.19	0.9451	32.67	0.9015	31.64	0.8826	31.41	0.9157	35.07	0.9318	-	-
SeaNet[62]	Multibranch Network	32.44	0.8981	28.81	0.7872	27.70	0.7399	26.50	0.7976	-	-	31.05	0.9154
RCAN[64]	Attention Mechanism	32.63	0.9002	28.87	0.7889	27.77	0.7436	26.82	0.8087	30.77	0.8459	31.22	0.9173
SRGAN[24]	GAN	32.05	0.8910	28.53	0.7804	27.57	0.7354	26.07	0.7839	28.92	0.8960	-	-
ZSSR[67]	Multiple Degradation	31.13	0.8796	28.01	0.7651	27.12	0.7211	-	-	-	-	-	-
SRMDNF[68]	Multiple Degradation	31.96	0.8925	28.35	0.7787	27.49	0.7337	25.68	0.7731	30.01	0.8278	30.09	0.9024

the classical and advanced SR methods are compared in many factors. Methods with direct reconstruction perform one-step up-sampling from the LR to HR space, while progressive reconstruction predicts HR images in multiple up-sampling steps. Global residual learning indicates that the network learns the difference between the ground truth HR image and the up-sampled LR images. Local residual learning stands for the local skip connections between intermediate convolutional layers. Multi-scale training indicates that the network design adopts multi-scale architecture. Table 3 shows the comparison of parameters for different SR algorithms. More parameters mean higher computation costs. Due to the improvement of the network structure and the convolutional layer, the parameters of FSRCNN and SCN are less than SRCNN, and the parameters are 12 K and 24 K, respectively. Although EDSR has an excellent performance, it has more parameters than other methods, reaching 43 000 K.

Table 3 Comparisons of parameter numbers

Method	Input	Output	Parameters	Global residual	Local residual	Multi-scale
SRCNN[35]	Bicubic	Direct	57 K			
FSRCNN[38]	LR	Direct	12 K			
SCN[56]	Bicubic	Progressive	42 K			
VDSR[40]	Bicubic	Direct	665 K	√	√	
DRCN[53]	Bicubic	Direct	1775 K	√		
LapSRN[55]	LR	Progressive	812 K	√		
DRRN[51]	Bicubic	Direct	297 K	√	√	√
SRGAN[24]	LR	Direct	1500 K			
DnCNN[36]	Bicubic	Direct	566 K			√
EDSR[14]	LR	Direct	43 000 K	√	√	
MDSR[14]	LR	Direct	8000 K	√	√	√
ZSSR[67]	LR	Direct	225 K	√		
MemNet[52]	Bicubic	Direct	677 K	√	√	√
IDN[59]	LR	Direct	796 K	√	√	
SRMDNF[68]	LR	Direct	1482 K			
SRFeat[11]	LR	Direct	6189 K	√	√	
D-DBPN[49]	LR	Direct	10 000 K	√	√	
RDN[48]	LR	Direct	21 900 K	√	√	
SRFBN[54]	LR	Direct	3 500 K	√	√	√
RCAN[64]	LR	Direct	16 000 K	√	√	√
DRN[61]	Bicubic	Direct	9800 K	√	√	√
IRN[81]	LR	Direct	4350 K			√
ESRN[50]	LR	Direct	1014 K	√	√	
ERN[45]	LR	Direct	9530 K	√	√	
SeaNet[62]	LR	Direct	7397 K	√	√	√

4 Conclusions

This survey paper reviews most work published on the topic of deep learning based super-resolution. The possible four development trends are concluded as follows.

(1) Unsupervised SR. As a self-supervised task, SR often explores low-resolution images from high ones through simple degradation algorithms. The exploration of similarities in self-images should be a valuable topic.

(2) Light-weight network. The existing work often seeks to build a deeper and more integrated network for extracting informative features. However, they are too time-consuming, greatly restricting usages in real cases.

(3) Realistic SR. Some work, especially the GAN based ones, reconstruct super-resolved images with promising perceptual quality. However, the super-resolved images often have seriously semantic inconsistency with their LR counterparts.

(4) Domain-specific applications. Super-resolution can greatly benefit other vision tasks. Therefore, it is also a promising direction to apply SR to more spe-

cific applications, such as object detection, face recognition.

References

[1] Girshick R, Donahue J, Darrell T, et al. Region-based convolutional networks for accurate object detection and segmentation[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(1): 142-158

[2] Bai Y, Zhang Y, Ding M, et al. Sod-mtgan: small object detection via multi-task generative adversarial network [C]// Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 206-221

[3] Mudunuri S P, Biswas S. Low resolution face recognition across variations in pose and illumination [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(5): 1034-1040

[4] Peng X, Yu X, Sohn K, et al. Reconstruction-based disentanglement for pose-invariant face recognition [C]// Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 1623-1632

[5] Peng X, Feris R S, Wang X, et al. A recurrent encoder-decoder network for sequential face alignment[C]// European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 38-56

[6] Greenspan H. Super-resolution in medical imaging[J]. *The Computer Journal*, 2009, 52(1): 43-63

[7] Lobanov A P. Resolution limits in astronomical images [J]. *arXiv*: astro-ph/0503225, 2005

[8] Zhang X Y, Li C, Shi H, et al. AdapNet: adaptability decomposing encoder-decoder network for weakly supervised action recognition and localization[J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2020 (99):1-12

[9] Zhang X Y, Wang S, Yun X. Bidirectional active learning: a two-way exploration into unlabeled and labeled data set [J]. *IEEE Transactions on Neural Networks and Learning Systems*, 2015, 26(12): 3034-3044

[10] Zhang X Y, Shi H, Li C, et al. Multi-instance multi-label action recognition and localization based on spatio-temporal pre-trimming for untrimmed videos [C]// The 32nd Innovative Applications of Artificial Intelligence Conference, New York, USA, 2020: 12886-12893

[11] Park S J, Son H, Cho S, et al. Srfeat: single image super-resolution with feature discrimination [C]// Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 439-455

[12] Wang X, Yu K, Wu S, et al. Esrgan: enhanced super-resolution generative adversarial networks [C]// Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 63-79

[13] Agustsson E, Timofte R. Ntire 2017 challenge on single image super-resolution: dataset and study[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 126-135

[14] Lim B, Son S, Kim H, et al. Enhanced deep residual networks for single image super-resolution[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 136-144

[15] He K, Zhang X, Ren S, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 770-778

[16] Wang Z, Chen J, Hoi S C H. Deep learning for image super-resolution: a survey [J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2020, 43: 3365-3387

[17] Bevilacqua M, Roumy A, Guillemot C, et al. Low-complexity single-image super-resolution based on nonnegative neighbor embedding [C]// British Machine Vision Conference, Surrey, UK, 2012: 1-10

[18] Zeyde R, Elad M, Protter M. On single image scale-up using sparse-representations [C]// International Conference on Curves and Surfaces, Avignon, France, 2010: 711-730

[19] Martin D, Fowlkes C, Tal D, et al. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics[C]// Proceedings of 8th IEEE International Conference on Computer Vision, Vancouver, Canada, 2001: 416-423

[20] Huang J B, Singh A, Ahuja N. Single image super-resolution from transformed self-exemplars [C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, USA, 2015: 5197-5206

[21] Fujimoto A, Ogawa T, Yamamoto K, et al. Manga109 dataset and creation of metadata[C]// Proceedings of the 1st International Workshop on Comics Analysis, Processing and Understanding, Cancun, Mexico, 2016: 1-5

[22] Wang X, Yu K, Dong C, et al. Recovering realistic texture in image super-resolution by deep spatial feature transform[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 606-615

[23] Blau Y, Mechrez R, Timofte R, et al. The 2018 PIRM challenge on perceptual image super-resolution[C]// Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 334-355

[24] Ledig C, Theis L, Huszár F, et al. Photo-realistic single image super-resolution using a generative adversarial network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4681-4690

[25] Wang Z, Bovik A C, Sheikh H R, et al. Image quality assessment: from error visibility to structural similarity [J]. *IEEE Transactions on Image Processing*, 2004, 13 (4): 600-612

[26] Sheikh H R, Sabir M F, Bovik A C. A statistical evaluation of recent full reference image quality assessment algorithms [J]. *IEEE Transactions on Image Processing*, 2006, 15(11): 3440-3451

[27] Xu X, Sun D, Pan J, et al. Learning to super-resolve blurry face and text images[C]// Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 251-260

[28] Dahl R, Norouzi M, Shlens J. Pixel recursive super resolution[C]//Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 5439-5448

[29] Lai W S, Huang J B, Ahuja N, et al. Fast and accurate image super-resolution with deep Laplacian pyramid networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 41(11): 2599-2613

[30] Zhu X, Li Z, Zhang X, et al. Generative adversarial image super-resolution through deep dense skip connections[J]. *Computer Graphics Forum*, 2018, 37(7): 289-300

[31] Zhu X, Li Z, Zhang X Y, et al. Residual invertible spatio-temporal network for video super-resolution[C]//The 31st Innovative Applications of Artificial Intelligence Conference, Honolulu, USA, 2019, 33: 5981-5988

[32] Zhang R, Isola P, Efros A A, et al. The unreasonable effectiveness of deep features as a perceptual metric[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 586-595

[33] Mittal A, Soundararajan R, Bovik A C. Making a "completely blind" image quality analyzer[J]. *IEEE Signal Processing Letters*, 2012, 20(3): 209-212

[34] Dong C, Loy C C, He K, et al. Learning a deep convolutional network for image super-resolution[C]//European Conference on Computer Vision, Santiago, Chile, 2014: 184-199

[35] Dong C, Loy C C, He K, et al. Image super-resolution using deep convolutional networks[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2015, 38(2): 295-307

[36] Zhang K, Zuo W, Chen Y, et al. Beyond a Gaussian denoiser: residual learning of deep CNN for image denoising[J]. *IEEE Transactions on Image Processing*, 2017, 26(7): 3142-3155

[37] Zhang K, Zuo W, Gu S, et al. Learning deep CNN denoiser prior for image restoration[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 3929-3938

[38] Dong C, Loy C C, Tang X. Accelerating the super-resolution convolutional neural network[C]//European Conference on Computer Vision, Amsterdam, The Netherlands, 2016: 391-407

[39] Shi W, Caballero J, Huszár F, et al. Real-time single image and video super-resolution using an efficient subpixel convolutional neural network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1874-1883

[40] Kim J, Kwon Lee J, Mu Lee K. Accurate image super-resolution using very deep convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1646-1654

[41] Ahn N, Kang B, Sohn K A. Fast, accurate, and lightweight super-resolution with cascading residual network[C]//Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 252-268

[42] Li C, Wang X, Dong W, et al. Joint active learning with feature selection via cur matrix decomposition[J]. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2018, 41(6): 1382-1396

[43] Jiao J, Tu W C, He S, et al. Formresnet: formatted residual learning for image restoration[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 38-46

[44] Fan Y, Shi H, Yu J, et al. Balanced two-stage residual networks for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 161-168

[45] Lan R, Sun L, Liu Z, et al. Cascading and enhanced residual networks for accurate single-image super-resolution[J]. *IEEE Transactions on Cybernetics*, 2020, 51(1): 115-125

[46] Huang G, Liu Z, Van Der Maaten L, et al. Densely connected convolutional networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 4700-4708

[47] Tong T, Li G, Liu X, et al. Image super-resolution using dense skip connections[C]//Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 4799-4807

[48] Zhang Y, Tian Y, Kong Y, et al. Residual dense network for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 2472-2481

[49] Haris M, Shakhnarovich G, Ukita N. Deep back-projection networks for super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 1664-1673

[50] Song D, Xu C, Jia X, et al. Efficient residual dense block search for image super-resolution[C]//The 32nd Innovative Applications of Artificial Intelligence Conference, New York, USA, 2020: 12007-12014

[51] Tai Y, Yang J, Liu X. Image super-resolution via deep recursive residual network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 3147-3155

[52] Tai Y, Yang J, Liu X, et al. Memnet: a persistent memory network for image restoration[C]//Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 4539-4547

[53] Kim J, Kwon Lee J, Mu Lee K. Deeply-recursive convolutional network for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, USA, 2016: 1637-1645

[54] Li Z, Yang J, Liu Z, et al. Feedback network for image super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 3867-3876

[55] Lai W S, Huang J B, Ahuja N, et al. Deep Laplacian pyramid networks for fast and accurate super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, USA, 2017: 624-632

[56] Wang Z, Liu D, Yang J, et al. Deep networks for image super-resolution with sparse prior[C]//Proceedings of

the IEEE International Conference on Computer Vision, Santiago, Chile, 2015: 370-378
[57] Hu Y, Gao X, Li J, et al. Single image super-resolution via cascaded multi-scale cross network[J]. *arXiv*:1802.08808, 2018
[58] Ren H, El-Khamy M, Lee J. Image super resolution based on fusing multiple convolution neural networks[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 54-61
[59] Hui Z, Wang X, Gao X. Fast and accurate single image super-resolution via information distillation network[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 723-731
[60] Qiu Y, Wang R, Tao D, et al. Embedded block residual network: a recursive restoration model for single-image super-resolution[C]//Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 2019: 4180-4189
[61] Guo Y, Chen J, Wang J, et al. Closed-loop matters: dual regression networks for single image super-resolution [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 5407-5416
[62] Fang F, Li J, Zeng T. Soft-edge assisted network for single image super-resolution[J]. *IEEE Transactions on Image Processing*, 2020, 29: 4656-4668
[63] Choi J S, Kim M. A deep convolutional neural network with selection units for super-resolution[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, USA, 2017: 154-160
[64] Zhang Y, Li K, Li K, et al. Image super-resolution using very deep residual channel attention networks[C]//Proceedings of the European Conference on Computer Vision, Munich, Germany, 2018: 286-301
[65] Kim J H, Choi J H, Cheon M, et al. Ram: residual attention module for single image super-resolution[J]. *arXiv*:1811.12043, 2018
[66] Zhou S, Zhang J, Zuo W, et al. Cross-scale internal graph neural network for image super-resolution[J]. *arXiv*:2006.16673, 2020
[67] Shocher A, Cohen N, Irani M. "zero-shot" super-resolution using deep internal learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 3118-3126
[68] Zhang K, Zuo W, Zhang L. Learning a single convolutional super-resolution network for multiple degradations [C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 3262-3271
[69] Xu Y S, Tseng S Y R, Tseng Y, et al. Unified dynamic convolutional network for super-resolution with variational degradations[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, USA, 2020: 12496-12505
[70] Goodfellow I, Pouget-Abadie J, Mirza M, et al. Generative adversarial nets[C]//Advances in Neural Information Processing Systems, Montreal, Canada, 2014: 2672-2680
[71] Peng X, Tang Z, Yang F, et al. Jointly optimize data augmentation and network training: adversarial data augmentation in human pose estimation[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 2226-2234
[72] Radford A, Metz L, Chintala S. Unsupervised representation learning with deep convolutional generative adversarial networks[J]. *arXiv*:1511.06434, 2015
[73] Sajjadi M S M, Scholkopf B, Hirsch M. Enhancenet: single image super-resolution through automated texture synthesis[C]//Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 2017: 4491-4500
[74] Bulat A, Tzimiropoulos G. Super-fan: integrated facial landmark localization and super-resolution of real-world low resolution faces in arbitrary poses with gans[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 109-117
[75] Bai Y, Zhang Y, Ding M, et al. Finding tiny faces in the wild with generative adversarial network[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, USA, 2018: 21-30
[76] Bhunia A K, Das A, Bhunia A K, et al. Handwriting recognition in low-resource scripts using adversarial learning[C]//Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, USA, 2019: 4767-4776
[77] Noh J, Bae W, Lee W, et al. Better to follow, follow to be better: towards precise supervision of feature super-resolution for small object detection[C]//Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 2019: 9725-9734
[78] Wang W, Xie E, Liu X, et al. Scene text image super-resolution in the wild[J]. *arXiv*:2005.03341, 2020
[79] Zhang H, Liu D, Xiong Z. Two-stream action recognition-oriented video super-resolution[C]//Proceedings of the IEEE International Conference on Computer Vision, Seoul, Korea, 2019: 8799-8808
[80] Zhu X, Li Z, Li X, et al. Attention-aware perceptual enhancement nets for low-resolution image classification [J]. *Information Sciences*, 2020, 515: 233-247
[81] Xiao M, Zheng S, Liu C, et al. Invertible image rescaling[J]. *arXiv*:2005.05650, 2020

Zhu Xiaobin, born in 1981. He received his Ph. D degree in 2013 from Institute of Automation, Chinese Academy of Sciences. His research interests include machine learning, pattern recognition, video and image analysis, etc.