Optics and Lasers in Engineering 146 (2021) 106717

Contents lists available at ScienceDirect

Optics and Lasers in Engineering

journal homepage: www.elsevier.com/locate/optlaseng

Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections

Yan Zou[a,b,c,1], Linfei Zhang[a,b,1], Chengqian Liu[a,b,c], Bowen Wang[a,b,*], Yan Hu[a,b], Qian Chen[a,b,*]

[a] Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
[b] Smart Computational Imaging (SCI) Laboratory, Nanjing University of Science and Technology, Nanjing, Jiangsu Province 210094, China
[c] Military Representative Office of army equipment department in Nanjing, Nanjing, Jiangsu Province 210094, China

ARTICLE INFO

Keywords:
Super-resolution
Infrared image
Convolutional neural network
Skip connection

ABSTRACT

Image super-resolution technology successfully overcomes the limitation of excessively large pixel size in infrared detectors and meets the increasing demand for high-resolution infrared image information. In this paper, the super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections is reported. The introduction of global residual learning and local residual learning reduces computational complexity and accelerates network convergence. Multiple convolution layers and deconvolution layers respectively implement the extraction and restoration of the features in infrared images. Skip connections and channel fusion are introduced to the network to increase the number of feature maps and promote the deconvolution layers to restore image details. Compared with the other previously proposed methods for infrared information restoration, our proposed method shows obvious advantages in the ability of high-resolution details acquisition.

1. Introduction

Modern military operations increasingly require optical imaging systems to be able to obtain detailed information in the theater of operations and make corresponding measures based on this information to ensure the successful completion of combat missions. High-resolution (HR) imaging systems have become one of the indispensable key information acquisition equipment for modern military operations. According to the Nyquist sampling theory, the highest spatial frequency information that an imaging sensor can collect is twice the size of the detector pixel. Therefore, when the resolution of the infrared imaging system is ultimately constrained by the pixel size of the infrared detector, the simplest way to improve the infrared imaging resolution is to increase the resolution by reducing the size of the pixel. However, limited by the manufacturing process of the detector, the pixel dimension cannot be further machined to a smaller size. Therefore, scholars propose to use computational imaging methods [1,2] to reconstruct an HR image from one or more low-resolution (LR) images. At present, most scholars utilize various super-resolution methods [3–8] to enhance the visual effect of infrared images.

Deep learning [9–14] is an important branch of machine learning and has been widely applied in the indirect acquisition of information. The basic principle of deep learning is to construct a representation learning structure containing multiple processing layers of a complex structure, and use a series of nonlinear transformation operations to recombine and extract simple features from the original data, so as to represent the data more effectively. Compared with traditional methods, deep learning has the advantages of higher accuracy, better robustness, and faster-testing speed. For the past few years, deep learning is widely used in computer vision [15–18]. In particular, the convolutional neural network (CNN) has been one of the great technological breakthroughs in various fields of computer vision, such as target detection [19,20] and image recognition [21,22], which has affected the domain of image super-resolution reconstruction. The CNN is introduced into the domain of super-resolution reconstruction for the first time in Super-Resolution using Convolutional Neural Network (SRCNN) [23]. In this network structure, the LR image is firstly enlarged to the target size by the bicubic interpolation, then a three-layer network structure is used to fit the nonlinear mapping, and finally, the HR image is output. This method is currently a typical CNN-based super-resolution reconstruction method.

The concept of Generative Adversarial Networks (GAN) [24] was first proposed in 2014. The generative network and the discriminant network in GAN play off against each other. When the samples generated by the generative network are input to the discriminant network, the discriminant network output value is close to 0.5, and the Nash equilibrium

* Corresponding authors.
E-mail addresses: wangbowen@njust.edu.cn (B. Wang), chenqian@njust.edu.cn (Q. Chen).
[1] The authors contributed equally to this work.

https://doi.org/10.1016/j.optlaseng.2021.106717
Received 3 January 2021; Received in revised form 1 June 2021; Accepted 9 June 2021

[25] is reached at this time. In other words, the discriminant network cannot judge the authenticity of this image. The super-resolution generative adversarial network (SRGAN) [26] applied the GAN to the domain of image super-resolution reconstruction. The mean-opinion-score (MOS) was added to the evaluation criteria for the first time. The peak signal-to-noise ratio (PSNR) of the SRGAN reconstructed image may not be the highest, but its visual quality is obviously better than other methods. However, the SRGAN network is difficult to implement the Nash equilibrium on account of its high computational complexity.

Due to the low resolution and blurred edge of the infrared image, the super-resolution reconstruction of the infrared image becomes more difficult, and the result of using the above network directly is not ideal. In this paper, the super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections is reported. The role of convolution layers is to extract image details, while the function of deconvolutional layers is to restore image details. Convolutional layers and deconvolutional layers in the network structure are connected by skip connections. Therefore, skip connections transmit convolutional layers information to deconvolutional layers. Skip connections also have settled the problem of gradient disappearance. The proposed network structure introduces both global and local residual learning [27–29], which greatly reduces the complexity of the network and accelerates the network convergence. The number of feature maps is greatly increased through channel fusion, which is helpful for deconvolution layers to restore image details. When the scales are three and four, the average PSNR of our results is 1.07dB and 1.31dB higher than Very Deep Convolutional Networks (VDSR) [30], respectively. Experiment results show that our network structure has excellent infrared image super-resolution reconstruction capabilities.

2. Related works

Since high-quality HR images possess rich and valuable details, they have been broadly used in medical imaging [31–34], satellite imaging [35,36], and security imaging [37]. Benefiting from the development of efficient hardware capabilities and the progress of complex algorithms, deep learning has shown obvious advantages in many domains such as computer vision and natural language processing [38,39]. The single-image super-resolution network has drawn widespread attention in society because of its high efficiency.

SRCNN introduced CNN into the field of super-resolution reconstruction for the first time. This method is currently a typical CNN-based super-resolution reconstruction. Fast Super-Resolution Convolutional Neural Networks (FSRCNN) [40] is a betterment of the previous SRCNN. FSRCNN directly inputs the original LR image into the network without pre-sampling. In addition, it uses a smaller convolution kernel and more mapping layers. Therefore, FSRCNN has a higher speed than SRCNN.

The input LR images are similar to the output HR images to a great extent, so in fact, the net only needs to learn the residual between HR images and LR images. VDSR is the most direct and obvious learning residual structure. VDSR sends the interpolated LR image and the original HR image into the network for end-to-end training. The output HR image of the network is obtained by adding the residual image learned from the network and the input image. VDSR deepens the network structure by introducing skip connections so that the deeper network model has a larger receptive field. Meanwhile, skip connections also alleviate the gradient vanishing problem caused by deep networks.

Residual Encoder-Decoder Networks (RED-Net) [41] is composed of a symmetrical convolutional-deconvolutional layer. As an encoding-decoding framework, it learns the nonlinear mapping from LR images to original HR images. The network structure uses convolutional layers to acquire the feature of images while uses deconvolutional layers to restore image details. Skip connections are introduced to settle the gradient disappearance problem. At the same time, skip connections transmit the details of convolutional layers to deconvolution layers, which can restore the image with higher clarity.

Based on the successful use of the above neural network, traditional methods such as dictionary learning and wavelet analysis are combined with deep learning super-resolution [42,43]. Their respective advantages are used to obtain more precise high-resolution image reconstruction methods. However, these two methods have high computational complexity in image preprocessing, and the restoration of images depends on the pre-defined relationship between input and output images.

Furthermore, for the low resolution of infrared imaging, large-scale super-resolution methods have some limitations. An intermediate scale is established in the low-resolution infrared image and the high-resolution infrared image, which concatenates two simple deep networks [44]. Two depth networks with different receptive fields are trained by the multi-scale loss function. This method reduces the complexity of the network and improves the accuracy of high-resolution reconstruction, which is a typical case of large-scale image super-resolution reconstruction and restoration. In addition, for the lack of details of infrared data, some scholars use the high-frequency information of visible images in low light to restore and reconstruct the high-resolution infrared image to improve the image reconstruction effect [45,46]. Nevertheless, both methods are based on the fact that visible images can provide corresponding high-frequency information. Unfortunately, in the case of night vision, visible images can not provide relevant information, limiting the application of those methods. In contrast, an unsupervised depth generator imaging model based on recurrent attention network [47] is proposed. The biggest difference of the network is the unsupervised imaging model, which does not need to provide the ground truth image for training, and finally recovers the reconstructed image in line with human visual characteristics. However, the GAN network model is more complex than the supervised learning network model, and it is more challenging to train. It is difficult to achieve Nash equilibrium, which is also a problem to be solved.

Inspired by the above approach and building on our previous research, this paper proposes super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections, which contains multiple convolutional layers and deconvolutional layers. The role of convolutional layers is to extract detailed features in images, while the function of deconvolution layers is to restore image details. Convolutional layers and deconvolutional layers in the network structure are connected by skip connections. Therefore, skip connections transmit convolutional layers information to deconvolutional layers. Skip connections also have settled the problem of gradient disappearance. Compared with the RED-Net structure, our net introduced residual blocks in the network structure to reduce the computational complexity of the network and accelerate the network convergence. We change the skip connections mode and increase the image information by adding the number of feature maps. This method significantly improves the ability of deconvolution to recover HR images, which makes the network more effective in the super-resolution of infrared images. The experimental results show that when the scale is three and four, the average PSNR of our results is 1.07dB and 1.31dB higher than that of VDSR, respectively.

3. Proposed methods

In this section, we introduce the proposed super-resolution network in detail. The network diagram is shown in Fig. 1. The network structure includes convolution layers and deconvolution layers. Rectified Linear Unit (ReLU) [48] layers are added after each convolution layer and deconvolution layer. To alleviate the problem of network overfitting, dropout layers [49] with a sampling threshold of 0.3 are added after each convolution layer and deconvolution layer. The number of layers in the network structure is shown in Table 1.

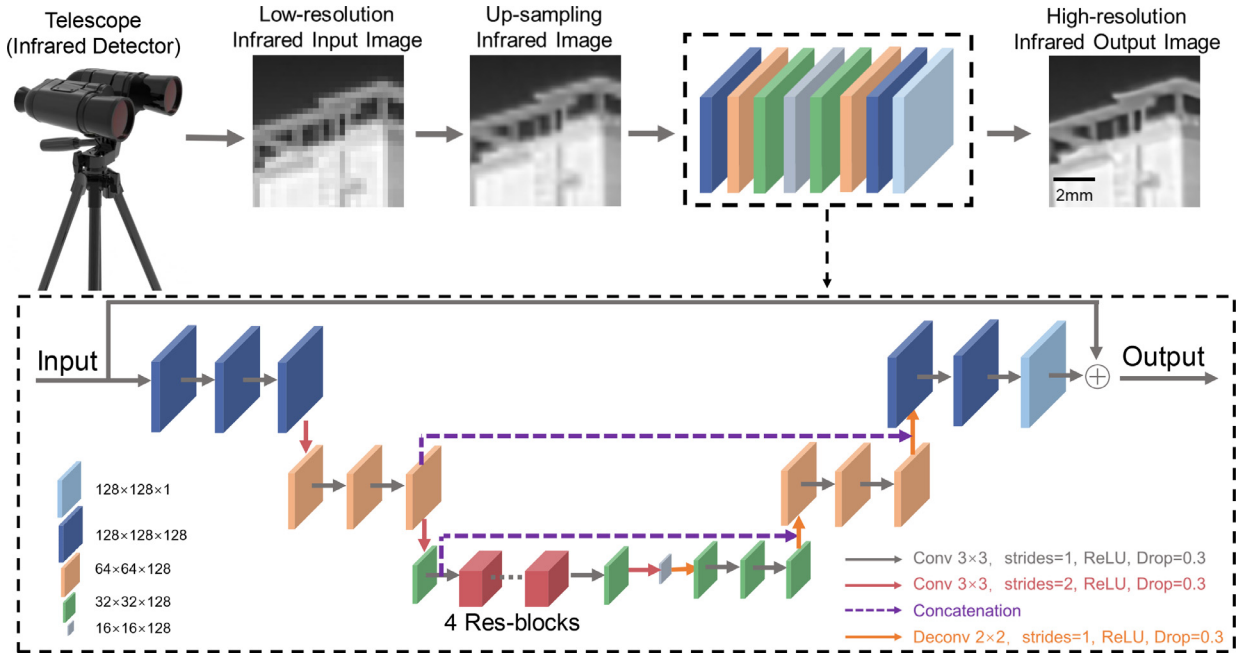

Fig. 1. The network structure of the proposed algorithm.

Table 1
The number of layers in the network structure.

Layer	Kernel Size	Strides	Filters	Numbers
Convolution Layer	3×3	1	128	12
Convolution Layer	3×3	2	128	3
Deconvolution Layer	2×2	1	128	3
Residual Blocks(Res-blocks)	3×3	1	128	4

3.1. Network structure

There are four types of layers in the network: convolution, deconvolution, element-wise addition, and channel fusion. Except for the channel fusion, each layer is followed by a ReLU layer. Let X_i be the input image of the i-th layer, the convolutional and deconvolutional layers are expressed as:

$$F(\mathrm{X}_i) = \max\left(0, \mathrm{W}_k * \mathrm{X}_i + \mathrm{B}_k\right). \tag{1}$$

Where W_k and B_k represent filters and biases, and $*$ denotes either convolution or deconvolution operation for the convenience of the formulation. For the element-wise addition layer, the output is the element-wise addition of two inputs of the same size, followed by the ReLU activation:

$$F\left(\mathrm{X}_i, \mathrm{X}_j\right) = \max\left(0, \mathrm{X}_i + \mathrm{X}_j\right). \tag{2}$$

Where X_i and X_j represent the input images of the $(\mathrm{i}+1)$-th layer and the $(\mathrm{j}+1)$-th layer respectively. For the channel fusion layer, the output is the sum of two input channels of the same size:

$$F\left(X_i, X_j\right) = X_i \oplus X_j. \tag{3}$$

For the convenience of expression, we use F_c and F_d to represent convolution and deconvolution operations and temporarily omit the activation function ReLU behind each layer. The proposed network has 26 layers. According to the above equation, we can express the output of the network structure as follows:

$$OUT = F_c\left(X_{25}\right) + X_0. \tag{4}$$

where X_0 is the input infrared image of the network. Specifically, the result can be calculated recursively as follows:

$$\begin{aligned} OUT &= F_c\left(X_{25}\right) + X_0 \\ &= F_c^2\left(X_{24}\right) + X_0 \\ &\cdots \\ &= F_c^2\left(F_d\left(F_c^2\left(F_d\left(F_c^2\left(F_d\left(F_c\left(X_{15}\right)\right)\right)\right)\oplus\right.\right.\right. \\ &\quad \left.\left.\left. F_c^7\left(X_0\right)\right)\right) \oplus F_c^5\left(X_0\right)\right)\right) + X_0. \end{aligned} \tag{5}$$

where

$$\begin{aligned} X_{15} &= F_c^{15}\left(X_0\right) + X_{13} \\ &= \cdots \\ &= F_c^{15}\left(X_0\right) + F_c^{13}\left(X_0\right) \\ &\quad + F_c^{11}\left(X_0\right) + F_c^9\left(X_0\right) + F_c^7\left(X_0\right). \end{aligned} \tag{6}$$

The above formula only iterates to X_{15} to facilitate the display of the differences in the proposed network structure. If there are no residual blocks in the network structure, $X_{15} = F_c^{15}\left(X_0\right)$. Compared with the network without residual block, the bottom of our network contains more details, which can prevent the gradient from disappearing. If there are no channel fusions, the OUT can be expressed as:

$$\begin{aligned} OUT &= F_c^2\left(F_d\left(F_c^2\left(F_d\left(F_c^2\left(F_d\left(F_c\left(X_{15}\right)\right)\right)\right)\right)\right)\right) \\ &\quad + X_0. \end{aligned} \tag{7}$$

Compared with our network, many image details are missing, which will reduce the ability of deconvolution to recover HR images. By fusing image information of different layers, the ability of deconvolution layers to recover HR images can be improved.

Specifically, LR infrared images are exactly up-sampled to the target image size and then send to the convolutional neural network for end-to-end supervised learning. Convolution layers in the network are applied as feature extractors to extract features from infrared images. Convolution with a stride of two is devoted to diminishing the dimension of feature images. Deconvolution layers can up-sample feature maps and recover the detailed information. Skip connections in the structure connect the first half and the second half of the network so that the image information can be transmitted from the front end of the network to the back end. Skip connections also solve the problem of vanishing

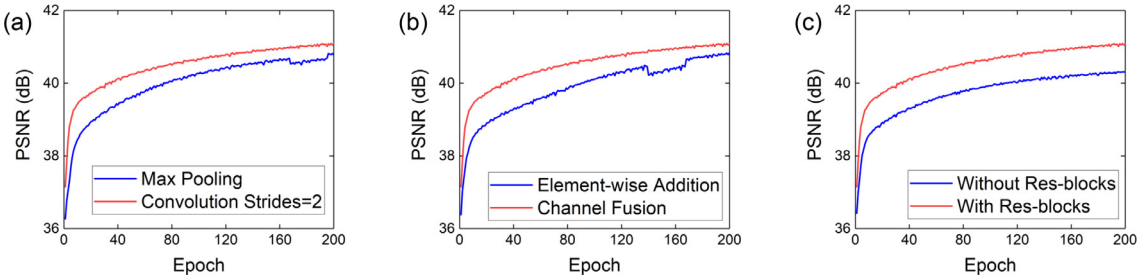

Fig. 2. The change curve of the PSNR on the training set in the verification experiment. (a) Test1: the influence of feature map dimensionality reduction method on super-resolution capability. (b) Test2: the influence of feature map fusion mode on super-resolution capability. (c) Test3: the influence of residual blocks on super-resolution capability.

gradient. Our net combines global residual learning with local residual learning, which reduces the computational complexity and accelerates the convergence of the network. Global residual learning means that the network only needs to learn the residual part between HR images and LR images, so it can greatly reduce the complexity of the network. The target HR image is obtained by adding the output image of the network and the input LR bicubic interpolation image.

3.2. Convolution and deconvolution

The proposed network structure contains a large number of convolutional layers and deconvolutional layers. The function of convolutional layers is to extract features from images. In previous works, AlexNet [50] uses some large convolution kernels, such as 11×11. The large convolution kernel has the advantage of a large receptive field and can extract the information of a large neighborhood of input images. But the large convolution kernel causes the network calculation to skyrocket. In consideration of deepening the network structure and enhancing the calculation performance, our net chooses 3×3 convolution kernels. We set the stride of convolutional layers to one and the padding to SAME in order to keep the image size unchanged in the feature extraction part. In addition, the network contains image dimensionality reduction processes. The reason why those processes do not use max-pooling operation for future maps dimensionality reduction is that max-pooling operation will eliminate image details. Max-pooling will reduce the performance of image restoration and reconstruction, which is contrary to the purpose of super-resolution. Therefore, using a convolution operation with a stride of two to reduce the dimensionality of images can retain the spatial information of feature maps with the most significant probability.

We take an experiment in Test 1 to verify the influence of the feature map dimensionality reduction method on the network super-resolution ability. In this verification experiment, the bicubic interpolation method is utilized to up-sample three times of the LR image as the input image and then keep parameters of the network structure consistent except the dimension reduction method. During the training process, the change in the PSNR is shown in Fig. 2 (a). The test results are shown in Fig. 3. In the network structure, the method of using convolution to reduce dimension is obviously better than max-pooling. Therefore, the convolution can be applied to reduce the dimension of feature maps, which is helpful to obtain higher resolution results.

Convolutional layers and Deconvolutional layers are located at the front and back ends of the network structure, respectively, and they are symmetrical. The convolution reduces the dimensionality of feature maps, and then the deconvolution can up-sample feature maps to the corresponding size. The change of image size can extract image features of multiple dimensions. Besides, the information of convolutional layers is transferred to deconvolutional layers through skip connections, which can greatly enhance the HR image reconstruction ability of deconvolutional layers.

3.3. Skip connections

Only by increasing the depth or width can the network model be suitable for more complex mapping functions. Unfortunately, the cost of increasing width is usually higher than that of depth. Therefore, deepening the network structure is a better way in most cases. The deeper network structure model means better nonlinear expression ability, which can learn more complex feature transformations and fit more complex mapping functions. Based on this, we design a deep residual network with skip connections structure to preferably learn the high-frequency details information between the interpolated LR image and the ground truth image.

Nevertheless, the deeper the network structure is, the better the super-resolution performance will not be. There may be two reasons. Since convolution processes make the image features more and more abstract, a lot of image details may be lost in the network with more convolutional layers. It is very difficult to restore and reconstruct the HR images with only a small amount of image details. In addition, the gradient vanishing problem is often encountered in the deep network structure. To address the above problems, skip connections are added to the network structure, which connects convolution layers of the first half of the network with deconvolution layers of the latter half of the network. The specific structure is shown in Fig. 1. Skip connections transfer a large number of image details of convolution layers to deconvolution layers, promoting the image super-resolution reconstruction ability of

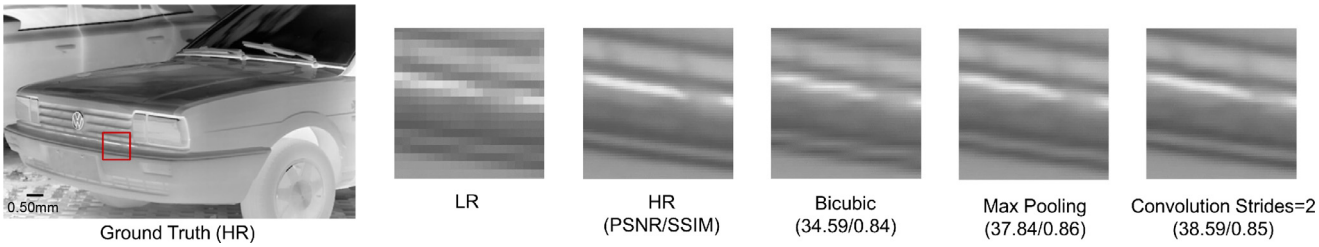

Fig. 3. Test1: the influence of feature map dimensionality reduction method on super-resolution capability.

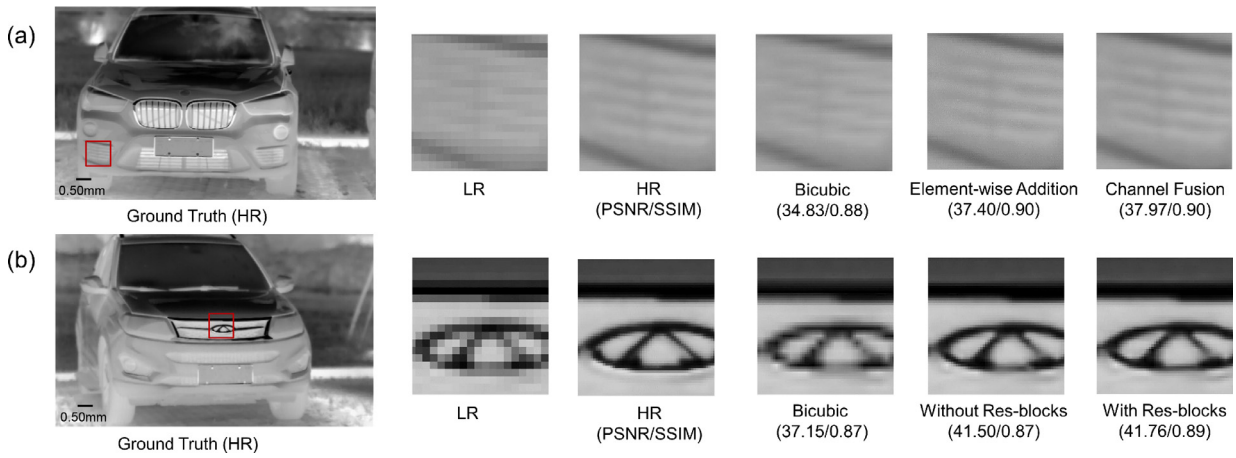

Fig. 4. (a) Test2: the influence of feature map fusion mode on super-resolution capability. (b) Test3: the influence of residual blocks on super-resolution capability.

deconvolution layers. Next, skip connections also have the advantage of backpropagation of gradient to the bottom layer, which solves the gradient disappearance problem. Consequently, training deeper networks becomes easier.

Compared with Image Super-Resolution Using Very Deep Residual Channel Attention Networks (RCAN) [51], the local residual and global residual modules are only built on the network structure at the same scale. The global residual is to find the smallest error in multiple local residual models and gradually approach the ground truth. The difference is that our network structure is analogous to the pyramid imaging model. A local residual module is added to the bottom layer of the network, so that an initial value of the high-resolution infrared image can be obtained at the starting point of the high-resolution image restoration. The high-resolution images of different scales are recovered by a step-by-step up-sampling and skip connection model. Moreover, the local residual model can reduce the network parameters at the bottom of the network, making the operation more efficient. Furthermore, we add the global residual module at the top of the network and add the extracted high-frequency information while retaining more low-frequency information of the image itself, which can restore the high-resolution image of the target to the greatest extent. The global residual learning structure and the local residual learning structure are shown in Fig. 6. On the contrary, only using the residual structure in the same dimension increases the computational complexity and lacks the extraction of high-frequency details in different dimensions, resulting in a decrease in the quality of reconstruction. Therefore, our paper combines global residual learning and local residual learning further to improve the overall performance of the network model super-resolution.

The results of Test2 and Test3 verify that the method of introducing residual block and changing feature fusion is effective in improving the super-resolution ability. The control variable method is introduced to ensure that the conditions other than the verification object remain unchanged. The curves of PSNR during training are shown in Fig. 2(b) and (c), and the experimental test results are shown in Fig. 4(a) and (b). The results show that the network structure proposed in this paper is more suitable for infrared image super-resolution.

4. Experiment and results

The infrared image data sets (DOI:10.6084/m9.figshare.14706183; https://figshare.com/s/2121562561211c0a8101) used in this experiment are all taken by the telescope in the thermal imaging mode. Typical training and test sets are shown in Figs. 11 and 12. The field of view (FOV) of the telescope is 16°. The image size of the infrared data set used in this experiment is 600×800. Down-sampling obtained images to acquire corresponding LR infrared images. Then, the bicubic interpolation method is used to up-sample LR infrared images to the target size. Infrared images are cut into 128×128 pieces and sent to the network for training.

Learning the end-to-end mapping from LR images to HR images needs to precisely estimate the weights θ represented by the convolutional and deconvolutional kernels. Specifically, there are N training sample pairs $\{X^z, Y^z\}$, where X^z is the LR image and Y^z is the HR version as the ground truth. We minimize the following Mean Squared Error (MSE):

$$L(\theta) = \frac{1}{N}\sum_{z=1}^{N} \|F(X^z) - Y^z\|^2. \tag{8}$$

In the network, the batch size is set to 16, and the epoch is set to 200. Empirically, we use Adam optimizer to optimize the network structure, and the initial learning rate is set to 10^{-4}. The hardware platform of the network for model training is Intel Core ™ i7-9700K CPU @ 3.60GHz×8, the graphics card is RTX2080Ti. The software platform used is TensorFlow 1.1.0 under Ubuntu 16.04 operating system.

The total training time of our network is 10.44 hours, and the average test time for each image is 0.57 seconds. In order to show the superiority of our network model, the results obtained by training with the network structure of this paper are compared with SRCNN and VDSR. The PSNR curve during training is shown in Fig. 5.

Table 2 shows the test results of each network under three different scales (scale=2, 3, 4). When the scale is two, the average PSNR of our net result is 6.43dB higher than the bicubic interpolation result, 0.90dB higher than SRCNN, and 0.07dB higher than VDSR. When the scale is small (scale = 2), our average PSNR is significantly higher than that of bicubic interpolation. Compared with VDSR, our PSNR is not evidently improved, and there are even cases where the PSNR is lower than VDSR. The reason for those situations may be that the scale is small, so that most of the information in the LR image is still retained. Therefore, it is not difficult to restore and reconstruct the HR image. When the scale is three, the average PSNR of our result is 3.37dB higher than the bicubic interpolation result, 1.89dB higher than SRCNN, and 1.07dB higher than VDSR. When the scale is four, the average PSNR of our result is 2.91dB higher than the bicubic interpolation result, 1.69dB higher than SRCNN, and 1.31dB higher than VDSR. When the scale is large (scale=3 or 4), the average PSNR improvement value of our results is not as good as when the scale is small, because it becomes difficult to restore HR images at this time. Compared with SRCNN and VDSR, our average PSNR has improved significantly, showing the superiority of our network in super-resolution reconstruction. In terms of structural similarity (SSIM), our results are slightly improved compared with SRCNN and VDSR. Part of the results is shown in Figs. 7, 8, and 9. Finally, we compare the results

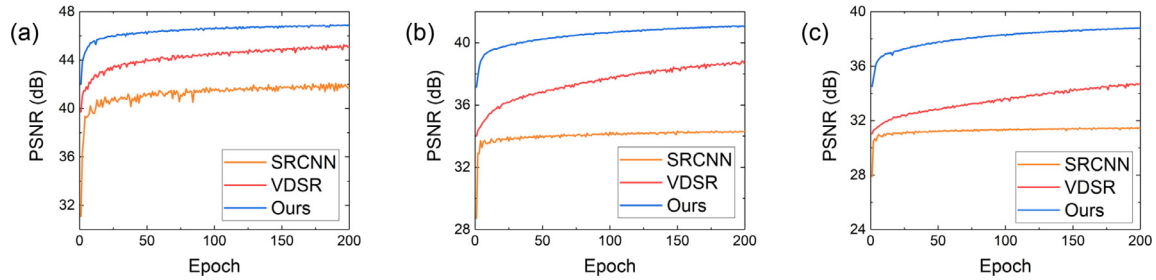

Fig. 5. Under different up-sample scales, the PSNR change curve of each network during training. (a) Scale=2. (b) Scale=3. (c) Scale=4.

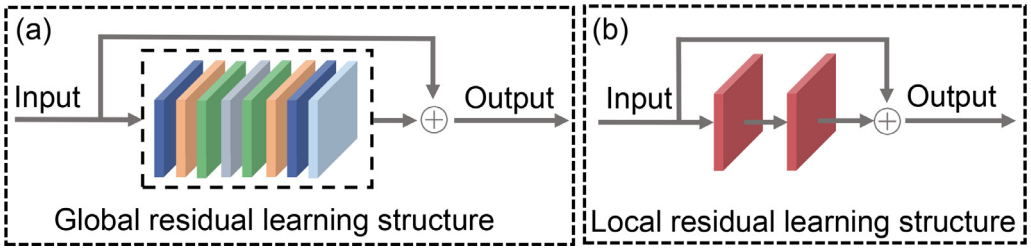

Fig. 6. Global residual structure and local residual structure.

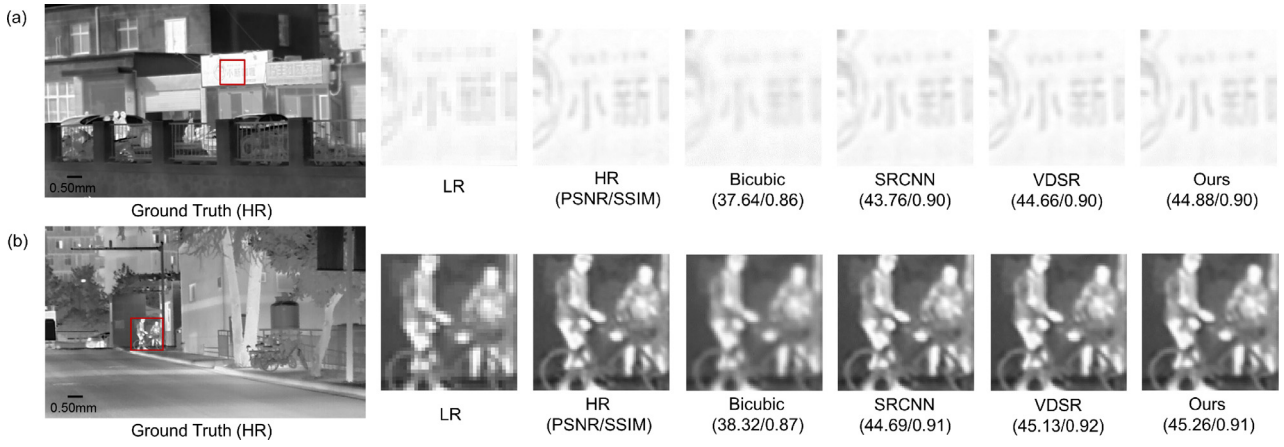

Fig. 7. Experimental results with an up-sampling scale of two.

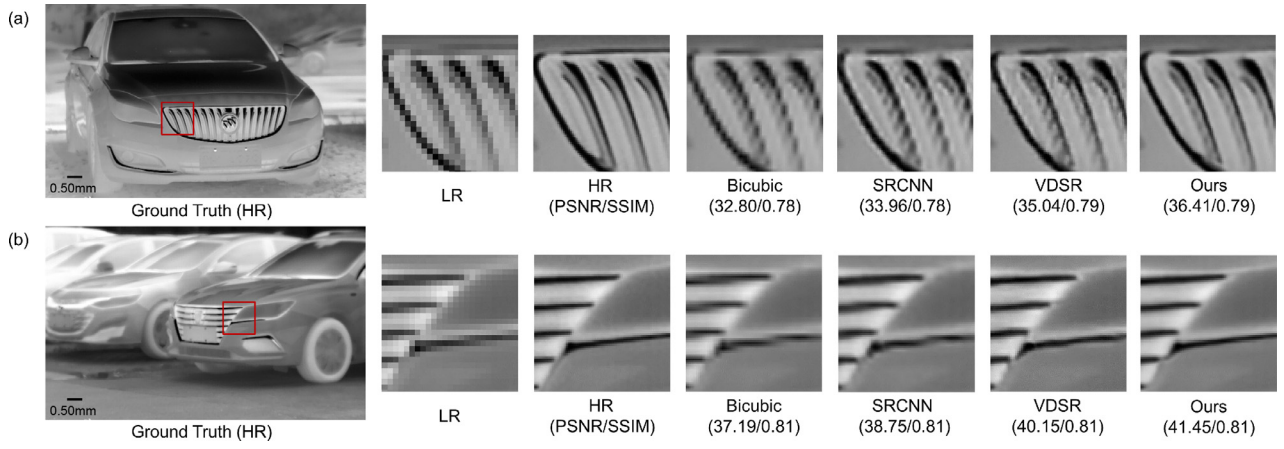

Fig. 8. Experimental results with an up-sampling scale of three.

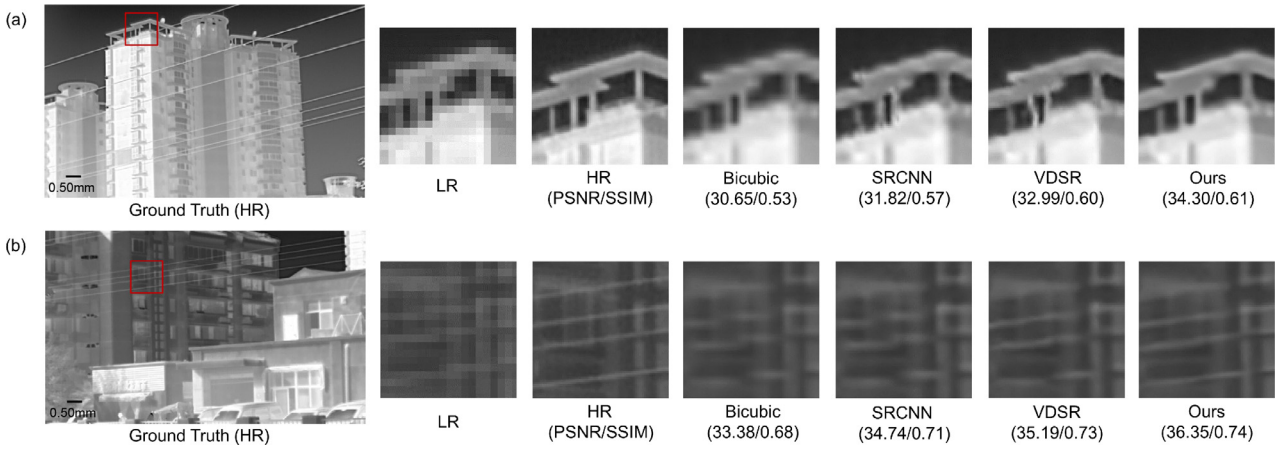

Fig. 9. Experimental results with an up-sampling scale of four.

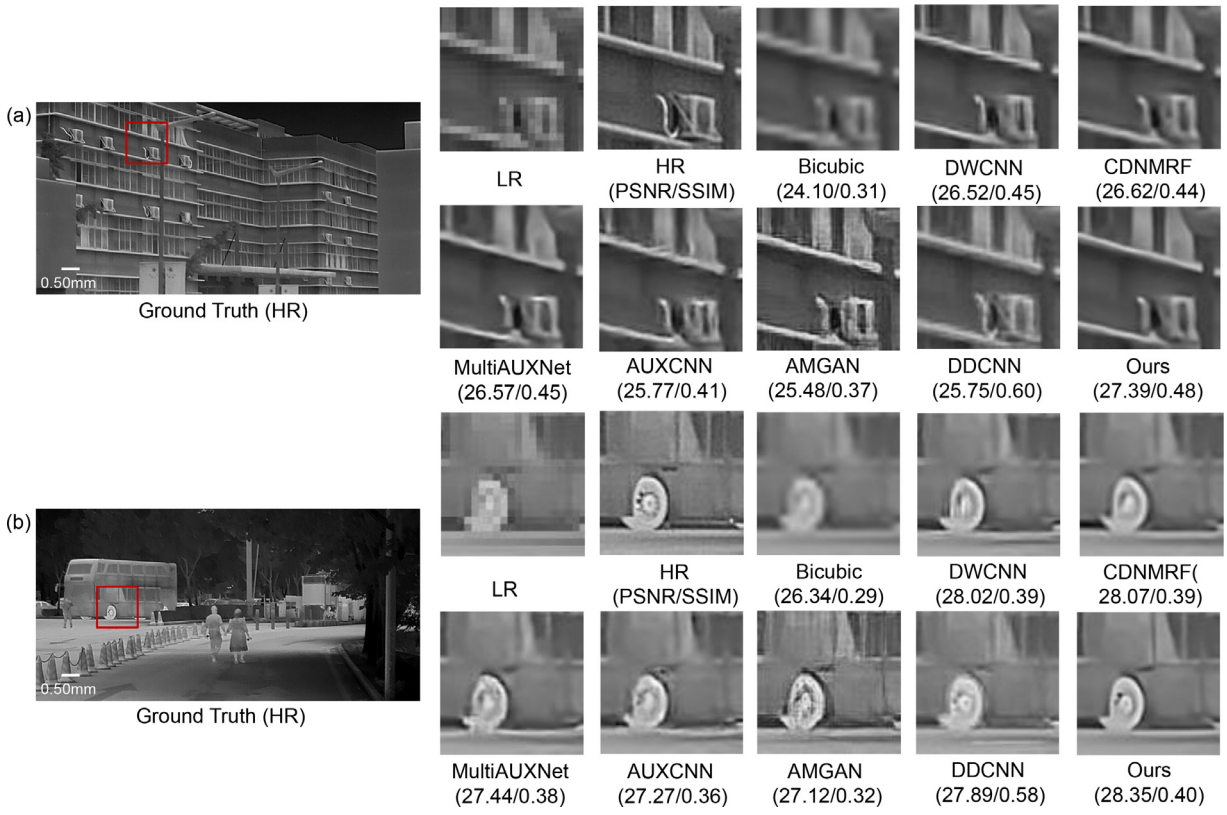

Fig. 10. The comparison of super-resolution imaging results with different networks (Scale=4).

Fig. 11. Ten representative groups of data in the training set.

Fig. 12. Ten representative groups of data in the testing set.

Table 2
Test set results under three different scale factors (scale factors=2, 3, 4). The bold text indicates the best result.

		Image 1	Image 2	Image 3	Image 4	Image 5	Image 6	Image 7	Image 8	Image 9
Methods	Scale	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bicubic	2	37.21/0.80	37.64/0.86	40.16/0.88	41.39/0.87	36.43/0.87	38.69/0.88	40.92/0.88	38.32/0.87	42.95/0.83
SRCNN	2	43.41/0.86	43.76/0.90	45.64/0.92	46.26/0.91	41.77/0.88	44.48/0.91	46.24/0.90	44.69/0.91	47.17/0.86
VDSR	2	45.15/0.87	44.66/0.90	46.23/**0.93**	46.14/0.91	44.36/0.89	44.86/0.92	**46.80/0.91**	45.13/**0.92**	47.27/0.87
Ours	2	**45.33/0.87**	**44.88/0.90**	**46.42**/0.92	**46.34/0.91**	**44.47/0.89**	**45.01/0.92**	46.77/0.90	**45.26**/0.91	**47.41/0.87**
Bicubic	3	33.64/0.67	33.85/0.76	36.41/0.78	37.74/0.79	32.80/0.78	35.02/0.79	37.19/0.81	34.76/0.77	39.50/0.73
SRCNN	3	35.23/0.69	35.53/0.78	37.66/0.80	39.55/0.79	33.96/0.78	36.72/0.80	38.75/0.81	36.75/0.78	40.10/0.73
VDSR	3	36.97/0.70	35.95/0.77	38.75/0.81	39.93/0.79	35.04/0.79	36.59/0.80	40.15/0.81	36.71/0.79	41.54/0.73
Ours	3	**38.55/0.71**	**36.64/0.78**	**39.36/0.81**	**40.76/0.80**	**36.41/0.79**	**38.12/0.81**	**41.45/0.81**	**37.62/0.80**	**42.37/0.73**
Bicubic	4	30.65/0.53	30.76/0.66	33.38/0.68	34.55/0.69	30.17/0.69	31.53/0.68	34.06/0.73	31.48/0.66	36.47/0.64
SRCNN	4	31.82/0.57	31.95/0.68	34.74/0.71	36.47/0.72	30.83/0.71	32.50/0.71	35.27/0.75	32.55/0.69	37.93/0.66
VDSR	4	32.99/0.60	31.85/0.68	35.19/0.73	36.67/0.72	31.48/0.72	32.77/0.71	35.43/0.75	32.89/0.69	38.22/0.66
Ours	4	**34.30/0.61**	**33.45/0.70**	**36.35/0.74**	**37.34/0.73**	**32.80/0.73**	**34.27/0.73**	**37.38/0.76**	**34.26/0.70**	**39.10/0.66**

with Deep Wavelet Residuals CNN (DWCNN) [43], Cascaded Deep Networks with Multiple Receptive Fields (CDNMRF) [44], Multi Auxiliary Network (MultiAUXNet) [45], Auxiliary Convolution Neural Network (AUXCNN) [46], Attention Mechanism Generative Adversarial Network (AMGAN) [47], and Discriminative Dictionary and Deep Residual Network (DDRes) [42], as shown in Fig. 10, respectively. Compared with other methods, our network has excellent super-resolution reconstruction capabilities at multiple scales.

5. Conclusion

This paper proposes the super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections. Image features are extracted by convolution layers, and deconvolution layers recover image details. Meanwhile, skip connections, and residual blocks are introduced to solve the problem of gradient vanishing and improve super-resolution performance. Moreover, the number of feature maps is significantly increased by channel fusion, which improves the super-resolution reconstruction capability of deconvolution layers. The up-sampling factor is set in three and four, and the average PSNR of our net is 1.07dB and 1.31dB higher than VDSR, respectively. Experiment results show that our network structure has excellent infrared image super-resolution reconstruction capabilities.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

CRediT authorship contribution statement

Yan Zou: Conceptualization, Writing - original draft, Formal analysis, Supervision. **Linfei Zhang:** Writing - original draft, Formal analysis, Investigation, Supervision. **Chengqian Liu:** Investigation, Supervision, Writing - original draft. **Bowen Wang:** Conceptualization, Formal analysis, Supervision, Writing - original draft. **Yan Hu:** Formal analysis, Supervision, Writing - original draft. **Qian Chen:** Supervision, Writing - original draft.

Acknowledgements

This work was supported by the National Natural Science Foundation of China (61722506, 11574152), National Defense Science and Technology Foundation of China (0106173), Outstanding Youth Foundation of Jiangsu Province (BK 20170034), The Key Research and Development Program of Jiangsu Province (BE2017162), National Defense Science and technology innovation project (2016300TS00908801), Equipment Advanced Research Fund of China (61404130314), and Open Research Fund of Jiangsu Key Laboratory of Spectral Imaging & Intelligent Sense (3091801410411).

References

[1] Zuo C, Li J, Sun J, Fan Y, Zhang J, Lu L, et al. Transport of intensity equation: a tutorial. Opt Lasers Eng 2020:106187.
[2] Pan A, Zhang Y, Wen K, Zhou M, Min J, Lei M, et al. Subwavelength resolution fourier ptychography with hemispherical digital condensers. OptExpress 2018;26(18):23119–31.
[3] Park SC, Park MK, Kang MG. Super-resolution image reconstruction: a technical overview. IEEE Signal ProcessMag 2003;20(3):21–36.
[4] Tian J, Ma K-K. A survey on super-resolution imaging. Signal Image Video Process 2011;5(3):329–42.
[5] Farsiu S, Robinson D, Elad M, Milanfar P. Advances and challenges in super-resolution. Int J Imaging SystTechnol 2004;14(2):47–57.
[6] Yue L, Shen H, Li J, Yuan Q, Zhang H, Zhang L. Image super-resolution: The techniques, applications, and future. Signal Process 2016;128:389–408.
[7] Zhang K, Li J, Xiong Z, Liu X, Gao X. Optimized multiple linear mappings for single image super-resolution. Opt Commun 2017;404:169–76.
[8] Siu W-C, Hung K-W. Review of image interpolation and super-resolution. In: Proceedings of the 2012 Asia Pacific signal and information processing association annual summit and conference. IEEE; 2012. p. 1–10.
[9] LeCun Y, Bengio Y, Hinton G. Deep learning. Nature 2015;521(7553):436–44.
[10] Ahn N, Kang B, Sohn K-A. Fast, accurate, and lightweight super-resolution with cascading residual network. In: Proceedings of the European conference on computer vision (ECCV); 2018. p. 252–68.

[11] Feng S, Zuo C, Yin W, Gu G, Chen Q. Micro deep learning profilometry for high-speed 3d surface imaging. Opt Lasers Eng 2019;121(OCT.):416–27.
[12] Yin W, Chen Q, Feng S, Tao T, Huang L, Trusiak M, et al. Temporal phase unwrapping using deep learning. SciRep 2019;9(1):1–12.
[13] Yin W, Zhong J, Feng S, Tao T, Han J, Huang L, et al. Composite deep learning framework for absolute 3d shape measurement based on single fringe phase retrieval and speckle correlation. J Phys 2020;2(4):045009.
[14] Yan K, Yu Y, Huang C, Sui L, Qian K, Asundi A. Fringe pattern denoising based on deep learning. Opt Commun 2019;437:148–52.
[15] Forsyth DA, Ponce J. Computer vision: a modern approach. Prentice Hall Professional Technical Reference; 2002.
[16] Voulodimos A, Doulamis N, Doulamis A, Protopapadakis E. Deep learning for computer vision: a brief review. ComputIntellNeurosci 2018;2018.
[17] Brunetti A, Buongiorno D, Trotta GF, Bevilacqua V. Computer vision and deep learning techniques for pedestrian detection and tracking: a survey. Neurocomputing 2018;300:17–33.
[18] O'Mahony N, Campbell S, Carvalho A, Harapanahalli S, Hernandez GV, Krpalkova L, et al. Deep learning vs. traditional computer vision. In: Science andinformation conference. Springer; 2019. p. 128–44.
[19] Mundhenk TN, Konjevod G, Sakla WA, Boakye K. A large contextual dataset for classification, detection and counting of cars with deep learning. In: European conference on computer vision. Springer; 2016. p. 785–800.
[20] Nam H, Han B. Learning multi-domain convolutional neural networks for visual tracking. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2016. p. 4293–302.
[21] Baccouche M, Mamalet F, Wolf C, Garcia C, Baskurt A. Sequential deep learning for human action recognition. In: International workshop on human behavior understanding. Springer; 2011. p. 29–39.
[22] Wen Y, Zhang K, Li Z, Qiao Y. A discriminative feature learning approach for deep face recognition. In: European conference on computer vision. Springer; 2016. p. 499–515.
[23] Dong C, Loy CC, He K, Tang X. Image super-resolution using deep convolutional networks. IEEE Trans Pattern Anal Mach Intell 2016;38(2):295–307.
[24] Goodfellow I, Pouget-Abadie J, Mirza M, Xu B, Warde-Farley D, Ozair S, Courville A, Bengio Y. Generative adversarial nets. In: Advances in neural information processing systems; 2014. p. 2672–80.
[25] Aumann R, Brandenburger A. Epistemic conditions for nash equilibrium. Econometrica 1995;63(5):1161–80.
[26] Ledig C, Theis L, Huszár F, Caballero J, Cunningham A, Acosta A, Aitken A, Tejani A, Totz J, Wang Z, et al. Photo-realistic single image super-resolution using a generative adversarial network. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2017. p. 4681–90.
[27] 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; 2016. p. 770–8.
[28] Zhang K, Zuo W, Chen Y, Meng D, Zhang L. Beyond a gaussian denoiser: Residual learning of deep CNN for image denoising. IEEE Transactions on Image Processing 2017;26(7):3142–55.
[29] Zhang Y, Tian Y, Kong Y, Zhong B, Fu Y. Residual dense network for image super-resolution. In: Proceedings of the IEEE conference on computer vision and pattern recognition; 2018. p. 2472–81.
[30] 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; 2016. p. 1646–54.
[31] Shen D, Wu G, Suk H-I. Deep learning in medical image analysis. AnnRevBiomedEng 2017;19:221–48.
[32] Litjens G, Kooi T, Bejnordi BE, Setio AAA, Ciompi F, Ghafoorian M, et al. A survey on deep learning in medical image analysis. MedImage Anal 2017;42:60–88.
[33] Ker J, Wang L, Rao J, Lim T. Deep learning applications in medical image analysis. IEEE Access 2017;6:9375–89.
[34] Razzak MI, Naz S, Zaib A. Deep learning for medical image processing: Overview, challenges and the future. In: Classification in BioApps. Springer; 2018. p. 323–50.
[35] Jean N, Burke M, Xie M, Davis WM, Lobell DB, Ermon S. Combining satellite imagery and machine learning to predict poverty. Science 2016;353(6301):790–4.
[36] Basu S, Ganguly S, Mukhopadhyay S, DiBiano R, Karki M, Nemani R. Deepsat: a learning framework for satellite imagery. In: Proceedings of the 23rd SIGSPATIAL international conference on advances in geographic information systems; 2015. p. 1–10.
[37] Brandt D.D., Hall K., Anderson M.B., Anderson C.D., Collins G.B.. System and methodology providing automation security analysis, validation, and learning in an industrial controller environment. 2014. US Patent 8,909,926.
[38] Collobert R, Weston J. A unified architecture for natural language processing: Deep neural networks with multitask learning. In: Proceedings of the 25th international conference on machine learning; 2008. p. 160–7.
[39] Socher R, Lin CC-Y, Ng AY, Manning CD. Parsing natural scenes and natural language with recursive neural networks. ICML; 2011.
[40] Dong C, Loy CC, Tang X. Accelerating the super-resolution convolutional neural network. In: European conference on computer vision. Springer; 2016. p. 391–407.
[41] Mao X-J, Shen C, Yang Y-B. Image restoration using convolutional auto-encoders with symmetric skip connections, arXiv preprint arXiv:1606.08921.
[42] Yao T, Luo Y, Hu J, Xie H, Hu Q. Infrared image super-resolution via discriminative dictionary and deep residual network. Infrared Phys Technol 2020;107:103314.
[43] Suryanarayana G, Tu E, Yang J. Infrared super-resolution imaging using multi-scale saliency and deep wavelet residuals. Infrared Phys Technol 2019;97:177–86.
[44] He Z, Tang S, Yang J, Cao Y, Yang MY, Cao Y. Cascaded deep networks with multiple receptive fields for infrared image super-resolution. IEEE TransCircuitsSystVideo Technol 2018;29(8):2310–22.
[45] Han TY, Kim DH, Lee SH, Song BC. Infrared image super-resolution using auxiliary convolutional neural network and visible image under low-light conditions. J Vis Commun Image Represent 2018;51:191–200.
[46] Zou Y, Zhang L, Chen Q, Wang B, Hu Y, Zhang Y. An infrared image super-resolution imaging algorithm based on auxiliary convolution neural network. In: Optics frontier online 2020: optics imaging and display, vol. 11571. International Society for Optics and Photonics; 2020. p. 115711B.
[47] Liu Q-M, Jia R-S, Liu Y-B, Sun H-B, Yu J-Z, Sun H-M. Infrared image super-resolution reconstruction by using generative adversarial network with an attention mechanism. Appl Intell 2021;51(4):2018–30.
[48] Glorot X, Bordes A, Bengio Y. Deep sparse rectifier neural networks. In: Proceedings of the fourteenth international conference on artificial intelligence and statistics; 2011. p. 315–23.
[49] Srivastava N, Hinton G, Krizhevsky A, Sutskever I, Salakhutdinov R. Dropout: a simple way to prevent neural networks from overfitting. J MachLearnRes 2014;15(1):1929–58.
[50] Krizhevsky A, Sutskever I, Hinton GE. Imagenet classification with deep convolutional neural networks. Commun ACM 2017;60(6):84–90.
[51] 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 (ECCV); 2018. p. 286–301.