

Article

Low-Light-Level Image Super-Resolution Reconstruction Based on a Multi-Scale Features Extraction Network

Bowen Wang ^{1,2,†} , Yan Zou ^{1,2,3,†}, Linfei Zhang ^{1,2}, Yan Hu ^{1,2}, Hao Yan ⁴, Chao Zuo ^{1,2,*} and Qian Chen ^{1,2}

¹ Jiangsu Key Laboratory of Spectral Imaging and Intelligent Sense, Nanjing University of Science and Technology, Nanjing 210094, China; wangbowen@njjust.edu.cn (B.W.); zouyan@njjust.edu.cn (Y.Z.); zhanglinfei@njjust.edu.cn (L.Z.); huyan@njjust.edu.cn (Y.H.); chenqian@njjust.edu.cn (Q.C.)

² Smart Computational Imaging (SCI) Laboratory, Nanjing University of Science and Technology, Nanjing 210094, China

³ Military Representative Office of Army Equipment Department in Nanjing, Nanjing 210024, China

⁴ Military Representative Office of Army Equipment Department in Taian, Taian 271000, China; yanhao@njjust.edu.cn

* Correspondence: zuochao@njjust.edu.cn

† These authors contributed equally to this work.

Abstract: Wide field-of-view (FOV) and high-resolution (HR) imaging are essential to many applications where high-content image acquisition is necessary. However, due to the insufficient spatial sampling of the image detector and the trade-off between pixel size and photosensitivity, the ability of current imaging sensors to obtain high spatial resolution is limited, especially under low-light-level (LLL) imaging conditions. To solve these problems, we propose a multi-scale feature extraction (MSFE) network to realize pixel-super-resolved LLL imaging. In order to perform data fusion and information extraction for low resolution (LR) images, the network extracts high-frequency detail information from different dimensions by combining the channel attention mechanism module and skip connection module. In this way, the calculation of the high-frequency components can receive greater attention. Compared with other networks, the peak signal-to-noise ratio of the reconstructed image was increased by 1.67 dB. Extensions of the MSFE network are investigated for scene-based color mapping of the gray image. Most of the color information could be recovered, and the similarity with the real image reached 0.728. The qualitative and quantitative experimental results show that the proposed method achieved superior performance in image fidelity and detail enhancement over the state-of-the-art.

Keywords: super resolution; low-light-level; deep learning network; multi-scale feature extraction



Citation: Bowen, W.; Yan, Z.; Linfei, Z.; Hu, Y.; Yan, H.; Zuo, C.; Chen, Q. Low-Light-Level Image Super-Resolution Reconstruction Based on a Multi-Scale Features Extraction Network. *Photonics* **2021**, *8*, 321. <https://doi.org/10.3390/photonics8080321>

Received: 21 July 2021

Accepted: 8 August 2021

Published: 10 August 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Super-resolution (SR) algorithms [1,2] serve the purpose of reconstructing high-resolution (HR) images from either single or multiple low-resolution (LR) images. Due to the inherent characteristics of the photoelectric imaging system, normally, it is challenging to obtain HR images [3]. In this regard, the SR algorithm provides a feasible solution to restore HR images from LR images recorded by sensors. As one of the essential sources of information acquisition in a low illumination environment, a low light level (LLL) imaging detection system [4] employs high sensitivity optoelectronic devices to enhance and record weak target and valuable environment information.

Unfortunately, due to the insufficient spatial sampling of the image detector and the trade-off between pixel size and photosensitivity, the ability of current imaging sensors to obtain both high spatial resolution and a large field-of-view is limited [5]. However, in traditional optical imaging, obtaining a slight improvement in imaging performance usually means a dramatic increase in hardware cost and, thus, causes difficulty in engineering applications.

The emergence of computational imaging [6–8] ideas has reversed this circumstance, drawing an exceptional opportunity for the remote sensing field. Image resolution is no longer only dependent on physical devices and, in turn, on the joint design of front-end optics and back-end image processing, to achieve sub-pixel imaging. With the development of deep learning methods, single image SR [9] has made significant progress. Utilizing this method, the non-linear function can be generated more effectively to deal with complex degradation models.

To date, image SR methods can be generally classified into traditional multi-frame image SR methods [10,11] and deep-learning-based methods [12–14]. In the conventional passive multi-frame image SR imaging algorithm, multiple frames with relative sub-pixel shifts based on the target scene are formed by the random relative motion [15–17] between the target and the sensor. Similarly, scholars proposed using computational imaging methods to reconstruct an HR image from one or more low-resolution images. However, non-ideal global panning, alignment errors, and non-uniform sampling are still problems in multi-image reconstruction algorithms.

The appearance of the convolutional neural network [18,19] reversed this situation. Deep-learning-based methods focus on exploiting external training data to learn a mapping function in accordance with the degradation process [20]. Its extraordinary fitting ability and efficient processing algorithms have enabled it to be generally utilized. Similarly, the processing method of single-frame image SR based on deep learning is not omnipotent, which has the disadvantages [21] of a slow training speed, poor network convergence, and the demand for an abundant amount of data. Therefore, how to achieve high speed, accurate, and effective image enhancement is still an essential issue to be solved.

In order to solve the above problems and make full use of the advantages of deep learning networks in feature extraction and information mapping, an SR neural network based on feature extraction is proposed, as shown in Figure 1. Compared with other single image SR methods, the significant advantage of MSFE is that it can draw out otherwise available information from the different scale observations of the same scene. The innovations of this paper are mainly in the following four aspects:

1. High-frequency information and low-frequency components can be fused in different scales by applying the skip connection structure.
2. The channel attention module and the residual block are combined to make the network focus on the most promising high-frequency information, mostly overcoming the main locality limitations of convolutional operations.
3. The dimension of the extracted high-frequency information is improved by sub-pixel convolution, and the low-frequency components are fused by element addition.
4. The network structure is expanded to realize grayscale image colorization and procure HR color images with a low-cost, LR monochromatic LLL detector in real-time.

The rest of the paper is structured as follows. In Section 2, some classical super-resolution imaging methods are briefly reviewed. In Section 3, the basic structure of our network is described in detail. In Section 4, the method of dataset building, details of training, comparison of the SR results, and color imaging results are presented. Section 5 is a summary of the article.

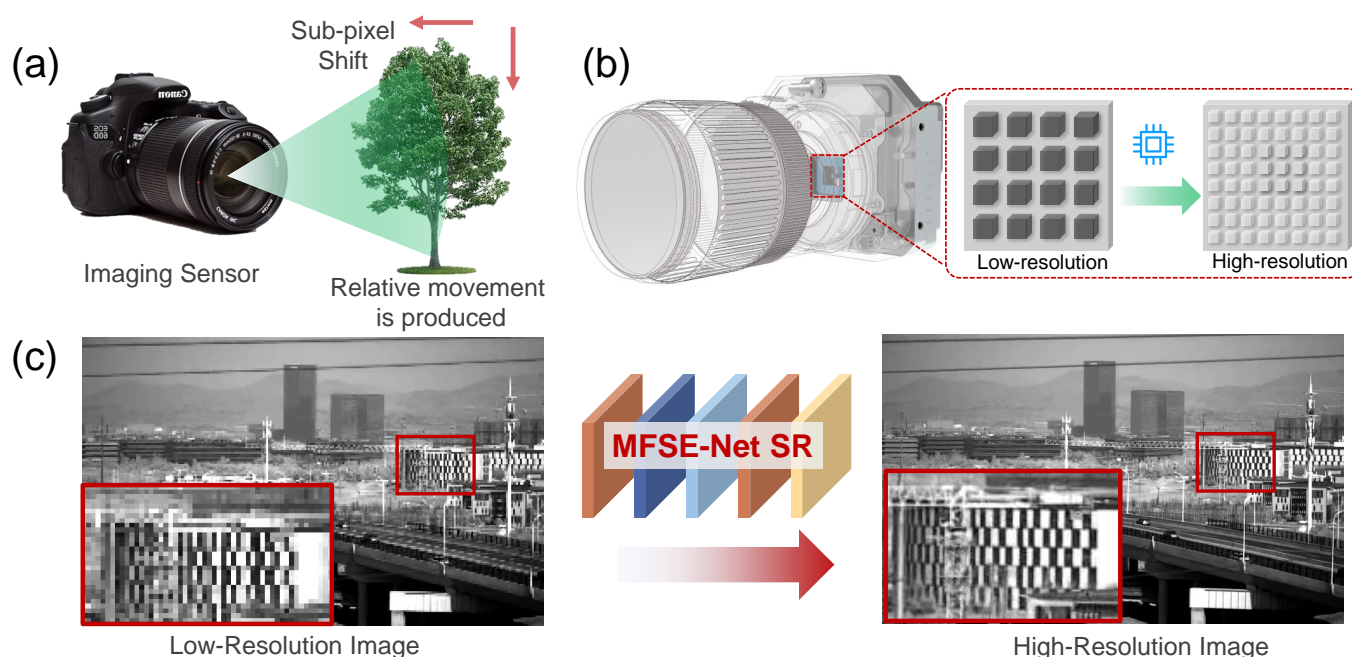


Figure 1. Schematic diagram of super-resolution reconstruction. (a) Multi-frame subpixel offset (b) Pixel super-resolution diagram (c) Super-resolution of a deep neural network.

2. Related Works

The most common SR method is based on interpolation, including bicubic linear interpolation and bilinear interpolation. Although the speed of these methods is perfect, each pixel is calculated according to the surrounding pixels, which only enlarges the image and cannot effectively restore the details of the image. The multi-reconstruction method [22] is to establish an observation model for the image acquisition process and then achieve SR reconstruction by solving the inverse problem of the observation model. However, the degradation model is often different from the actual situation, which cannot predict the image correctly.

The super-resolution method based on the sparse representation (SCSR) method [23,24] treats HR images and LR images as a dictionary multiplied by atoms. By training the dictionary, HR images can be obtained from multi-LR images. Insufficiently, it is evident that the dictionary training time is longer, and the reconstructed image has an apparent sawtooth phenomenon. Another key technology to realize SR reconstruction is micro-scanning SR imaging [25,26], which is realized by the vibration of the detector or the rotation of the scanning mirror or the flat plate to obtain the steerable displacement between the optical system and the sensor.

In addition, the method of controlled micro scanning can simplify the steps of image registration and improve the accuracy of the reconstruction results. However, sub-pixel scanning requires HR optoelectronic devices, such as motors and piezoelectric drivers, dramatically increasing the cost and complicating the whole imaging system. Therefore, the traditional image SR reconstruction method [27] still has the limitations of multi-frame image reconstruction, algorithm complexity, a complex imaging system, and a precise control system.

Learning-based methods build upon the relation between LR-HR images, and there have been many recent advancements in this approach, mostly due to deep convolutional neural networks. In the pursuit of higher image quality, the super resolution convolutional neural network (SRCNN) [28] applied a convolutional neural network (CNN) to image SR reconstruction for the first time, and the SR images obtained were superior to traditional SR algorithms.

Based on SRCNN, the fast super-resolution convolutional neural network (FSRCNN) [29] was proposed, which directly sent LR images to the network for training and used a deconvolution structure to obtain reconstructed images. The improved network not only improved the depth of the network but also significantly reduced the amount of calculations and improved the speed of calculations. In deep residual networks (ResNet) [30], the residuals were directly learned through skip connections, which effectively solved the problem of gradient disappearance or explosion in deep network training.

Next, the Laplacian pyramid networks (LapSRN) [31,32] realized parameter sharing in different levels of amplification modules, reduced the calculation amount, and effectively improved the accuracy through the branch reconstruction structure. The wide activation super-resolution network (WDSR) [33] ensured more information passing through by expanding the number of channels before the activation function and also provided nonlinearity fitting of the neural network.

After that, the channel attention (CA) mechanism was used in deep residual channel attention networks (RCAN) [34], which can focus on the most useful channels and improve the SR effect. However, to achieve the additional benefits of the CNN, a multitude of problems needs to be solved. For the most current super-resolution imaging methods, if the input image is a single channel gray image, the output image is also a single channel gray image. Human eye recognition objects can be identified by the brightness information and color information.

Colorizing images can produce more completed and accurate psychological representation, leading to better scene recognition and understanding, faster reaction times, and more accurate object recognition. Therefore, color fusion images [35] can make the naked eye recognize the target more accurately and faster. Directed against the above problems, this paper proposes a convolution neural network based on feature extraction, which is employed to SR imaging of LLL images under dark and weak conditions to achieve HR and high sensitivity all-weather color imaging. The network uses a multi-scale feature extraction module and low-frequency fusion module to combine multiple LR images.

3. Super-Resolution Principles

According to the sampling theory, an imaging detector can sample the highest spatial frequency information that is the half of the sampling frequency of its sensor. When the sensor's pixel size becomes the main factor restricting the resolution of an imaging system, the simplest way to improve the imaging resolution is to reduce the pixel size to enhance the imaging resolution. However, in practical applications, due to the limitations of the detector manufacturing process, the pixel size of some detectors, such as the LLL camera, cannot be further reduced.

This means that, in some cases, the LR image will lose some information compared with the HR image; inevitably, there will exist the pixel aliasing phenomenon. Therefore, the sampling frequency of the imaging sensor will be the key factor to limit the imaging quality. The SR method improves the spatial resolution of the image from LR to HR. From the perspective of imaging theory, the SR process can be regarded as the inverse solution of blur and down-sampling. SR is an inherently ill-posed problem in either case since multiple different solutions exist for any LR image. Hence, it is an underdetermined inverse problem with unique solutions. Based on a representation learning model, our proposed methodology aims to generate a super-resolved image function from the HR image to the LR image.

3.1. Image Super-Resolution Forward Model

A flow chart for the observation model is illustrated in Figure 2. For an imaging system, the image degradation process is first affected by the optical system's lens, leading to the diffraction limit, aberration, and defocusing in an optical lens, which can be modeled as linear space invariant (LSI). Unfortunately, it is more problematic that pixel aliasing will

occur if the detector pixel size exceeds a specific limitation, making the high-frequency part of the imaging object unavailable in the imaging process.

It is inevitable to introduce the subsampling matrix, which generates aliased LR images from forward-generating the blurring HR image into the imaging model. Conventionally, SR image reconstruction technology utilizes multiple LR observation images to reconstruct the underlying HR scene with noisy and slight movement. Consequently, assuming a perfect registration between each HR and LR image, we can derive the expressions of the observation HR image, sampled scene x in matrix form as:

$$y = BDx + n \quad (1)$$

where D is a subsampling matrix, B typifies a blurring matrix, x symbolizes the desired HR image, y represents the observed LR image, and n is the zero-mean white Gaussian noise associated with the observation image. Different from the traditional multi-frame super-resolution imaging, the deep learning reconstruction method establishes the mapping relationship between the LR image and the HR image. Through the information extraction of different dimensions, the problem of image pixelation imaging is effectively solved, and super-pixel resolution imaging is realized.

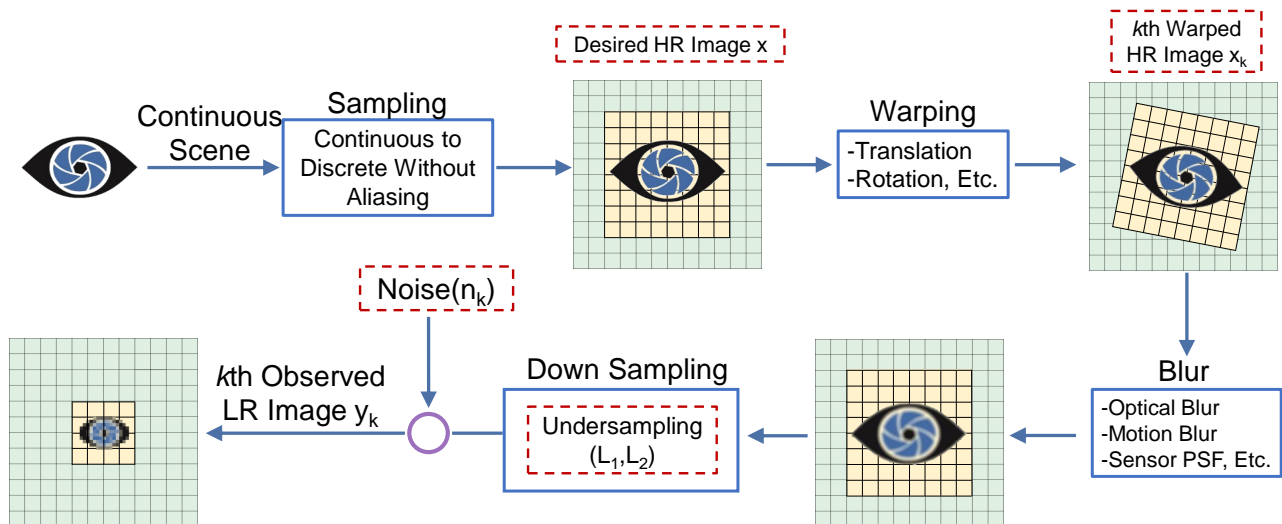


Figure 2. Image super-resolution forward model.

3.2. Network Structure

An overview of the RAMS network is depicted in Figure 3. The whole network is a pyramid model, which is cascaded by two layers of the model, and each layer of the model realizes twice the LR LLL image feature extraction. Vertically, the model is composed of two branches, the upper part is the feature extraction branch of the LLL image, and the lower part is the reconstruction branch of the LLL image.

The feature extraction branch obtained the high-frequency information of the corresponding input image. High-frequency features are more helpful for HR reconstruction, while LR images contain abundant low-frequency information, which directly forwards to the network tail-ends. The reconstruction branch obtains the up-sampled image corresponding to the size of the HR image. In order to express the SR of the network more clearly, the network model can be defined as:

$$I_{\text{out}}(x, y) = F_{\omega, \theta}[I_{LR}(x, y)] \quad (2)$$

where $F_{\omega, \theta}[\bullet]$ represents the nonlinear mapping function of the network, ω and θ , respectively, depict the trainable parameters of weight and deviation in the network, I_{LR} describes the LR LLL input image, and I_{out} typifies the HR image predicted by the network.

The specific convolution layer number and parameters of the super-resolution network structure are shown in Table 1.

Table 1. The number and parameters of convolution layer in the super-resolution network structure.

Layer	Numbers
Convolution layer ($3 \times 3 \times 32$)	10
Convolution layer ($3 \times 3 \times 192$)	8
Convolution layer ($3 \times 3 \times 4$)	4
Sub-pixel convolution layer	4
Deconvolution layer ($2 \times 2 \times 32$)	1
Global Average Pooling layer	8
Fully connected layer (FC) (24)	8
Fully connected layer (FC) (192)	8

The main task of low-level feature extraction, high-level feature extraction, and feature mapping is mainly to collect the promising compositions from the input I_{LR} image into the CNN network and to express all information into feature maps. It is noteworthy that the corner, edge, and line can be dug out from the feature maps. The attention mechanism module focuses on the most promising features and reduces the interference of irrelevant features.

We formulate the procedure as $F_{w,\theta}[\bullet]$, which consists of four parts:

1. Low-Level Feature Extraction: This step aims to extract the fundamental information from the input I_{LR} image and forward it as a series of feature maps.
2. High-Level Feature Extraction: In this operation, through a convolution operation of different dimensions and channel attention mechanism module, the calculation of the network is mostly focused on the acquisition of high-frequency information.
3. Features Mapping: In order to reduce the hyperparameters, the mapping process from high-dimensional vector to low ones is designed.
4. Reconstruction: This operation integrates all the information to reconstruct an HR image I_{out} .

In the following paragraphs, we present the overall architecture of the network with a detailed overview of the main blocks. Finally, we conclude the methodology section with precise details of the optimization process for training the network.

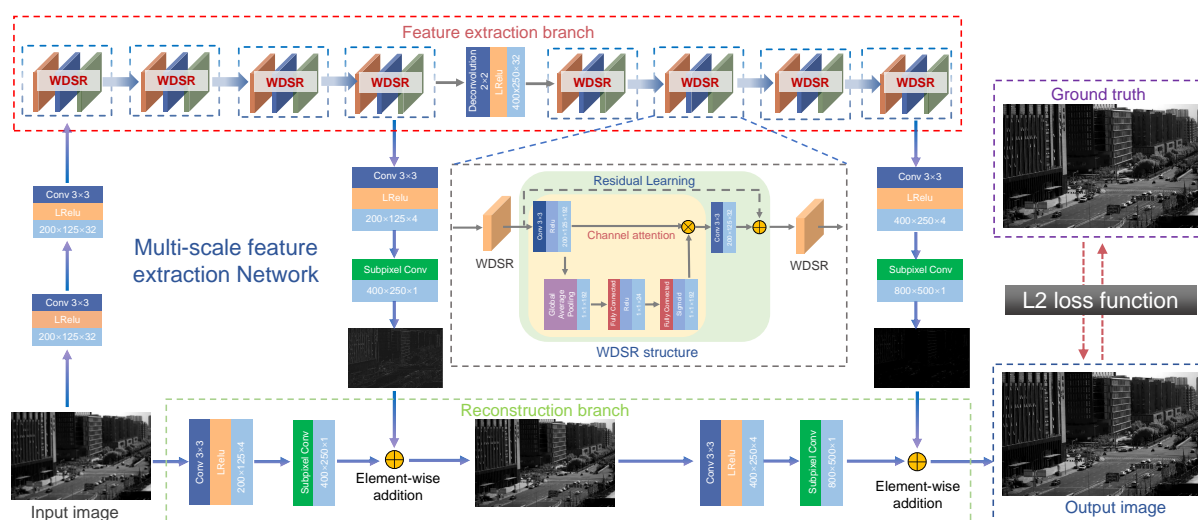


Figure 3. Structure diagram of the super-resolution deep learning network based on multi-scale feature extraction.

3.2.1. Feature Extraction Branch

The feature extraction branch of each pyramid model mainly includes a convolution layer, wide activation residual module, and sub-pixel convolution layer. The purpose of the feature extraction branch is to realize the feature extraction of the LLL image. The wide activation residual module mainly includes a channel attention mechanism and skip connection. The channel attention mechanism is similar to the human selective visual attention mechanism. The core goal is to select the more critical information to the current task from several details.

In the deep learning network, the channel attention mechanism can adjust each channel's weight and retain valuable information beneficial to obtain HR LLL image to achieve SR reconstruction of the LLL image. In addition, the current mainstream network structure models are developing in a deeper direction. A deeper network structure model means a more dependable nonlinear expression ability, acquiring more complex transformation, and fitting more input complex features. To this end, we employed a long skip connection for the shallow features and several short skip connections inside each feature attention block to let the network focus on more valuable high-frequency components.

In addition, the skip connection in the residual structure efficiently enhanced the gradient propagation and alleviated the problem of gradient disappearance caused by the deepening of the network. Therefore, the skip connection was introduced in the wide activation residual module, to extract the image detail information and improve the super-resolution performance of the network structure. As shown in Figure 4a, the experiments of wide activation residual module without skip connection and with skip connection were carried out, respectively. The network structure with a skip connection produced a more robust SR performance and better expressed the details of the image.

Similarly, as shown in Figure 4b, the number of wide activation residual modules in each pyramid model was verified. In the verification experiment, only the number of wide activation residual modules was changed. By this, it can be seen from the impact that the network structure with multiple wide activated residuals had a higher fidelity SR performance than those with single wide activated residuals as shown in Figure 4b.

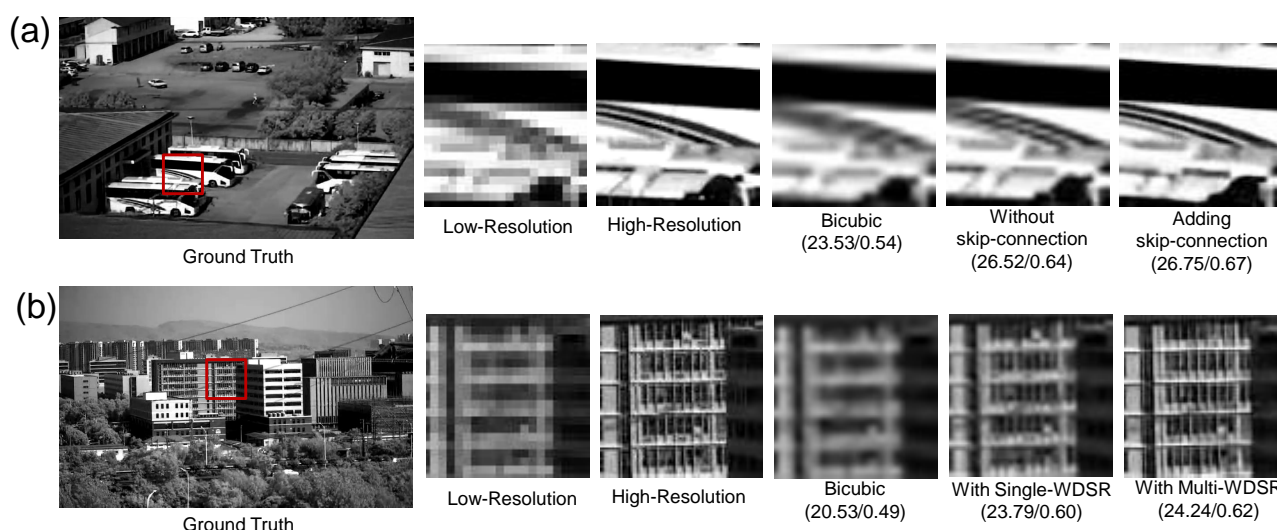


Figure 4. The comparison experiment of various network structure parameters. (a) Comparison of the skip connection residual structure. (b) Comparison of multiple wide activation residual modules.

3.2.2. Reconstruction Branch

The reconstruction branch mainly includes a convolution layer and sub-pixel convolution layer to enlarge the feature image. In general deconvolution, there will be several values of zero, which may reduce the performance of SR reconstruction. In order to max-

imize the effectiveness of the image information to enhance the imaging resolution, we employed the sub-pixel convolution with reconstruction from the LR image to the HR image by pixel shuffling. Subpixel convolution combines a single pixel on a multichannel feature map into a unit on a feature map. That is to say, the pixels on each feature map are equivalent to the subpixels on the new feature map.

In the reconstruction process, we set the number of output layers to a specified size to ensure that the total number of pixels is consistent with the number of pixels of the HR image. By doing so, the pixels can be rearranged through the sub-pixel convolution layer, and we finally obtain the enlarged LLL image. This method utilizes the ability of the sub-pixel deconvolution process to learn complex mapping functions and effectively reduces the error caused by spatial aliasing. The model predicts and regresses the HR image gradually in the process of reconstruction through gradual reconstruction.

This feature makes the SR method more applicable. For example, depending on the available computing resources, the same network can be applied to enhance the different video resolutions. The existing techniques based on the convolutional neural network cannot provide such flexibility for the scene with limited computing resources. In contrast, our model with four-times magnification can still perform two times better than the SR and only requires bypassing the more refined residual calculation.

3.2.3. Loss Function

Let I_{LR} denote the LR LLL input image, and I_{out} represents the HR LLL image predicted by the network. ω and θ depict the trainable parameters weight and deviation in the network. Our goal is to learn a nonlinear mapping function $F_{\omega,\theta}[\cdot]$ to generate a HR LLL image $I_{out}(x,y) = F_{\omega,\theta}[I_{LR}(x,y)]$, which is as close as possible to the real image I_{HR} . The loss function used in the training is the mean square error, which can be expressed by the following formula:

$$L(\omega, \theta) = \frac{1}{N} \sum_{i=1}^N \|F_{\omega,\theta}[I_{LR}^i(x,y)] - I_{HR}^i(x,y)\|_{HR}^2, \quad (3)$$

where I_N is the number of training samples. The curve of the loss function during training is shown in Figure 5.

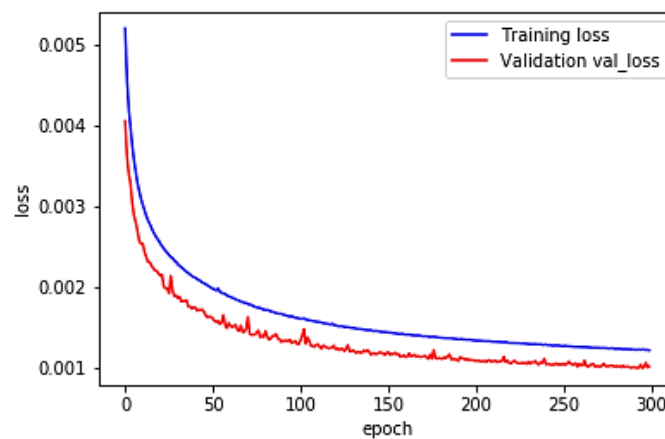


Figure 5. The training loss function of the super-resolution network.

4. Analysis and Discussion

In this section, first, the details of the data set and the experimental setup are introduced, and then the LR LLL images are input into four different networks to quantitatively evaluate the SR reconstruction results. After that, the network is extended to RGB color image reconstruction, and its colorization ability is verified.

4.1. Data Set Establishment

We utilized the telescope to obtain the LLL image resolution of 800×600 . After clipping, LLL images with the size of 800×500 were obtained. Then, multiple images with the size of 128×128 were cropped to the size of 800×500 . In this paper, 500 images were input as the training set, 50 images were input as the verification set, and some of the representative training sets are shown in Figure 6. Finally, the original LLL images size of 128×128 were taken as the ground truth, and then the LLL images were down-sampled four times to obtain the LR LLL images resolution of 32×32 as input to form the training set.



Figure 6. The representative training sets for super resolution.

4.2. Experimental Setup

In the network, the batch size was set to 4, and the epoch was set to 300. Empirically, we employed an Adam optimizer to optimize the network structure, and the initial learning rate was set to 10^{-4} . The activation function was Leaky Rectified Linear Unit (LReLU), and the parameter was 0.2. The hardware platform of the network for model training was an Intel Core™ i7-9700K CPU @ 3.60 GHz \times 8, and the graphics card was RTX2080Ti. The software platform was TensorFlow 1.1.0 under the Ubuntu 16.04 operating system.

The LR LLL image dimension of 32×32 and the corresponding HR LLL image dimension of 128×128 were sent into the program as the original image to train the neural network. The network training took 3.2 h. In the test, the input LLL image size was 200×125 , and the HR LLL image size was 800×500 . The test time of each image was 0.016 s. Therefore, our proposed network not only realized SR imaging but also realized all-weather real-time imaging. Part of the real-time images are shown in Figure 7.



Figure 7. The output part of the real-time image in the video stream.

4.3. Comparison of Super-Resolution Results with Different Networks

The imaging ability of three traditional SR neural networks (CDNMRF [36], VDSR [37], and MultiAUXNet [38]) was compared with our network. We utilized the peak signal-to-noise ratio (PSNR) and structural similarity (SSIM) as specific numerical evaluation indexes, and the particular results are shown in Table 2. In the case of four up-sampling scales, we compared the experimental results with CDNMRF, VDSR, and MultiAUXNet, and the results are shown in Figure 8. Subjectively, our method reconstructed the most similar details of the wall, iron frame, window, car, and so on, and the edges of the images were the clearest. In the objective evaluation, PSNR and SSIM were calculated and compared.

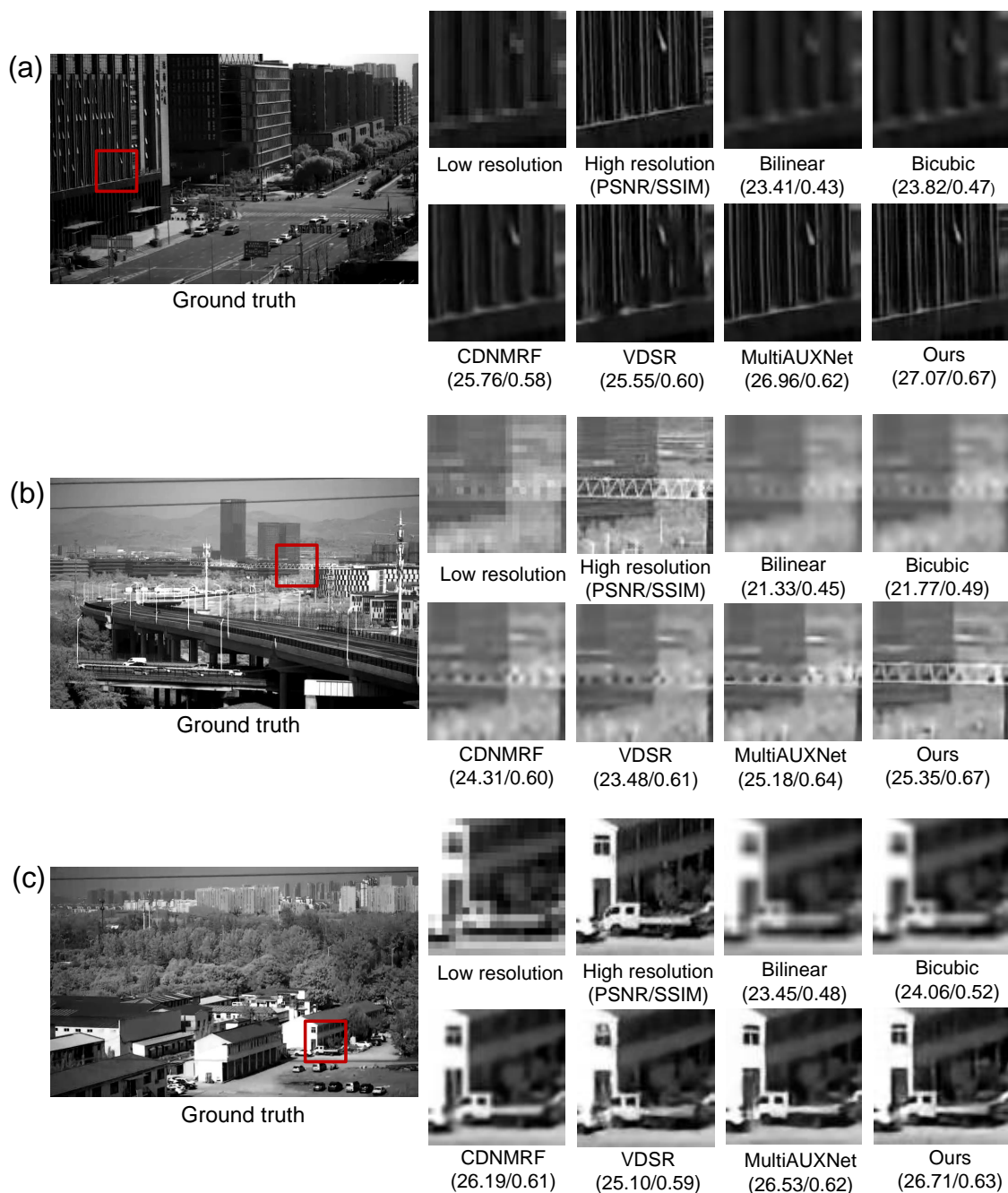


Figure 8. Comparison of the super-resolution results with different networks. (a–c) Different test scenes.

Table 2. The PSNR and SSIM results of different super-resolution networks.

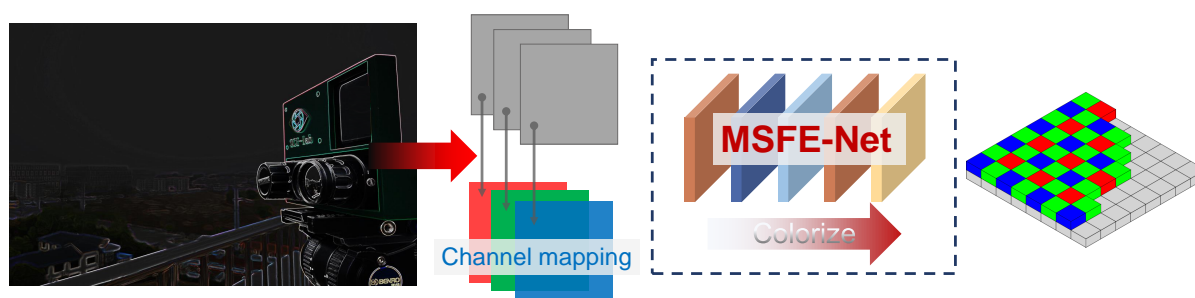
	Image 1	Image 2	Image 3
Methods	PSNR/SSIM	PSNR/SSIM	PSNR/SSIM
Bilinear	23.41/0.43	21.33/0.45	23.45/0.48
Bicubic	23.82/0.47	21.77/0.49	24.06/0.52
CDNMRF	25.76/0.58	24.31/0.60	26.19/0.61
VDSR	25.55/0.60	23.48/0.61	25.10/0.59
MultiAUXNet	26.96/0.62	25.18/0.64	26.53/0.62
Ours	27.07/0.67	25.35/0.67	26.71/0.63

In terms of PSNR, our results were 0.96 db higher than CDNMRF, 1.67 db higher than VDSR, and 0.15 dB higher than MultiAUXNet. In terms of SSIM, our results were 0.06 higher than CDNMRF, 0.06 higher than VDSR, and 0.03 higher than MultiAUXNet. In general, our network structure showed better super-resolution performance in the wide FOV LLL images.

4.4. Application for RGB Images

Remote imaging detection requires the imaging system to provide detailed high-frequency information and visible spectral information in the imaging process. The color information of most color images is quite different from that of the natural scene, which is not realistic. Nevertheless, the observer can further segment the image by distinguishing the color contrast of the fused image to recognize different objects in the image.

In addition to the SR task of the gray image, we also extended the network performance in our work. As shown in Figure 9, by expanding the number of channels in the original network, the color image's RGB channel corresponded to one gray level output, and gray images of different scenes were colorized under the condition of LLL imaging.

**Figure 9.** The colorization network framework for the RGB images.

The proposed LLL image colorization was combined with the existing scene image library for supervised learning. First, the input LLL gray image was classified, and the category label was obtained. Then, the natural color fusion image was recovered by color transfer. Compared with the color look-up table method, the proposed method can adaptively match the most suitable reference image for color fusion without acquiring the natural image of the scene in advance. As shown in Figure 10, we realized the color image reconstruction based on the jungle and urban environments, respectively.

Similarly, we also evaluated the color image output by the network, as shown in Figure 11. Figure 11c describes the difference between the network output image and the actual color image captured by the visible light detector. We can see that only the local color information was wrong. Furthermore, we quantitatively evaluated the histogram distribution similarity of the two images. The color distribution was basically the same, and the similarity of the final histogram distribution was 0.728, as shown in Figure 11. In general, the imaging results met the requirements of human visual characteristics and intuitively handled the scene information in HR.

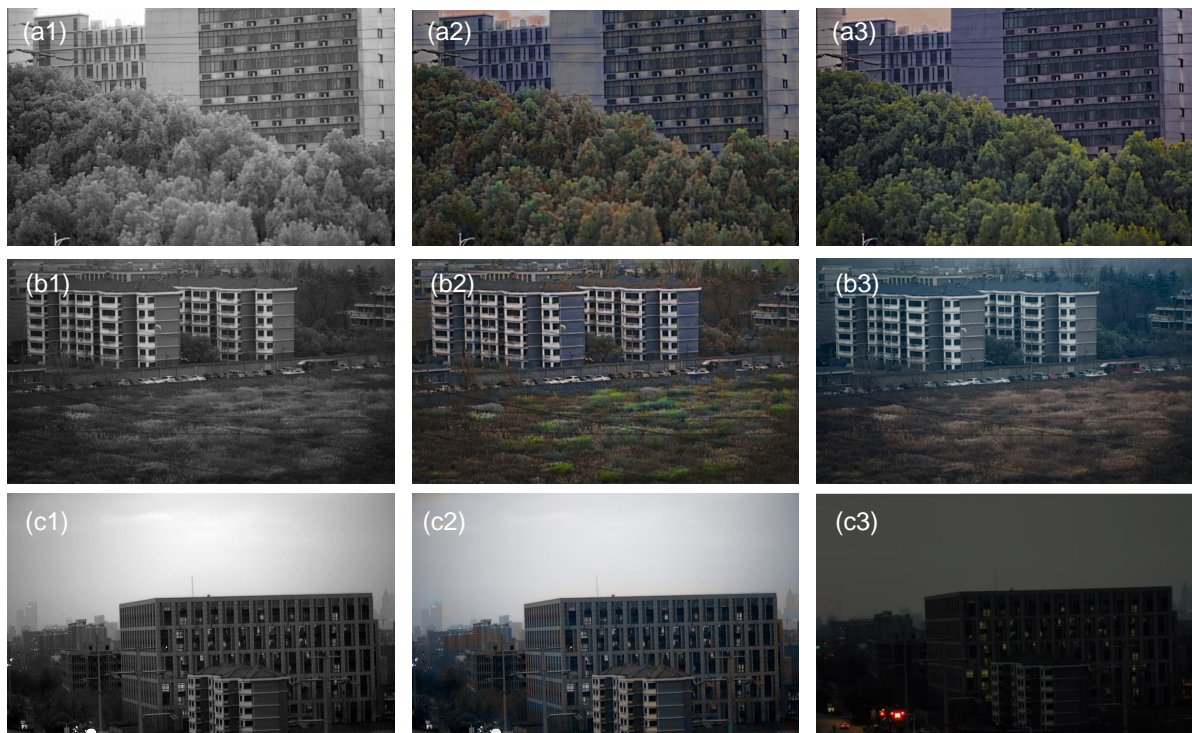


Figure 10. The image color reconstruction results based on the scene. (a1–c1) Input grayscale images; (a2–c2) output color images; and (a3–c3) color images captured by the visible sensor.

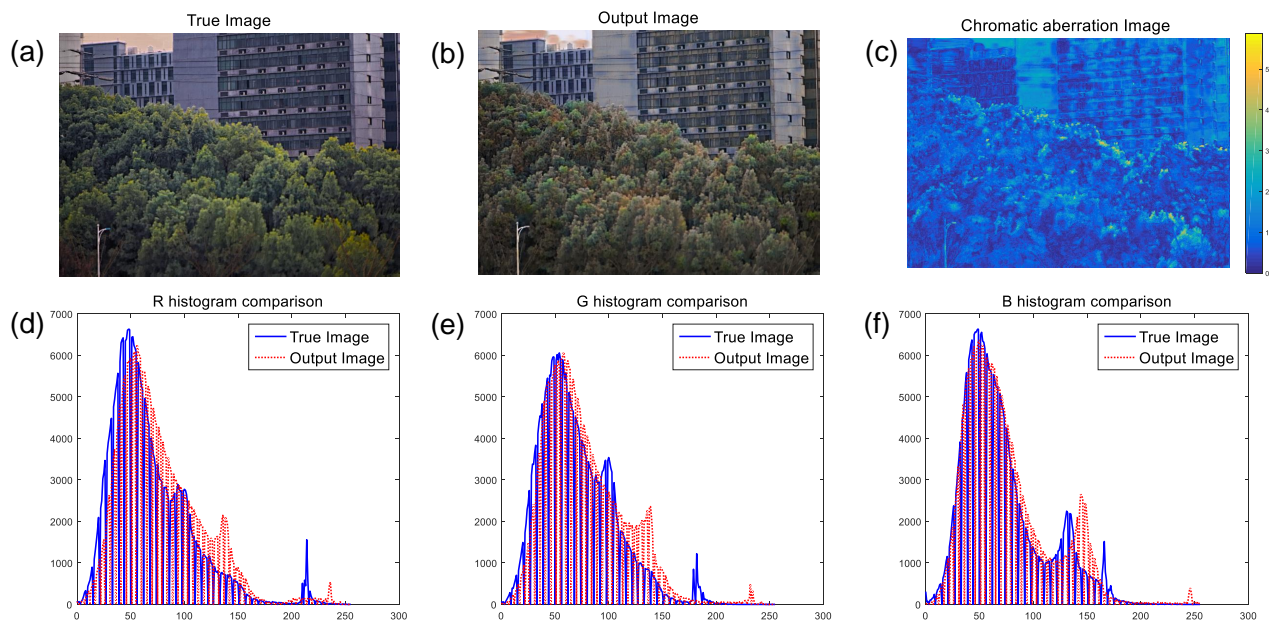


Figure 11. Quantitative evaluation of color images. (a) The true image; (b) the output image; (c) the chromatic aberration diagram; and (d–f) histogram comparison of the R, G, and B channels.

5. Conclusions

In summary, we demonstrated an SR network structure based on multi-scale feature extraction. The proposed network learned an end-to-end mapping function to reconstruct an HR image from its LR version, which could robustly reproduce the visual richness of natural scenes under different conditions and output photos with high quality. The network,

which is based on high-frequency component calculation, effectively improved the peak signal-to-noise ratio of the reconstructed image by 1.67 dB.

The effective network structure realized the image output of 0.016 s per frame and, thus, guarantees real-time imaging. In order to realize the output of the color image, we expanded the number of channels of the network to achieve the mapping of a single channel image to a three-channel image. The similarity between the histogram distribution of the final output color image and the authentic image captured by the visible detector reached 0.728. The experimental results indicate that the proposed method offers superior image fidelity and detail enhancement, which suggests promising applications in remote sensing, detection, and intelligent security monitoring.

Author Contributions: C.Z. and B.W. proposed the idea. Y.Z. and B.W. jointly wrote the manuscript and analyzed the experimental data. L.Z. and B.W. performed the experiments. Y.H. and H.Y. analyzed the data. Q.C. and C.Z. supervised the research. All authors have read and agreed to the published version of the manuscript.

Funding: 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), Postgraduate Research & Practice Innovation Program of Jiangsu Province (KYCX21_0274).

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

MSFE	Multi-Scale feature extraction
LLL	Low-Light-level
HR	High-Resolution
LR	Low-Resolution
CDNMRF	Cascaded deep networks with multiple receptive fields
VDSR	Very deep super resolution
MultiAUXNet	Multi auxiliary network
WDSR	Wide Activation Super-Resolution

References

1. Milanfar, P. *Super-Resolution Imaging*, 1st ed.; CRC Press: Boca Raton, FL, USA, 2011.
2. Park, S.C.; Park, M.K.; Kang, M.G. Super-resolution image reconstruction: A technical overview. *IEEE Signal Process. Mag.* **2003**, *20*, 21–36. [\[CrossRef\]](#)
3. Katsaggelos, A.K.; Molina, R.; Mateos, J. Super resolution of images and video. *Synth. Lect. Image Video Multimed. Process.* **2007**, *1*, 1–134. [\[CrossRef\]](#)
4. Hyneczek, J.; Nishiwaki, T. Excess noise and other important characteristics of low light level imaging using charge multiplying CCDs. *IEEE Trans. Electron. Devices* **2003**, *50*, 239–245. [\[CrossRef\]](#)
5. Zheng, G.; Horstmeyer, R.; Yang, C. Wide-field, high-resolution Fourier ptychographic microscopy. *Nat. Photonics* **2013**, *7*, 739–745. [\[CrossRef\]](#)
6. Nguyen, N.; Milanfar, P.; Golub, G. A computationally efficient superresolution image reconstruction algorithm. *IEEE Trans. Image Process.* **2001**, *10*, 573–583. [\[CrossRef\]](#) [\[PubMed\]](#)
7. Zuo, C.; Li, J.; Sun, J.; Fan, Y.; Zhang, J.; Lu, L.; Zhang, R.; Wang, B.; Huang, L.; Chen, Q. Transport of intensity equation: A tutorial. *Opt. Laser Eng.* **2020**, *135*, 106187. [\[CrossRef\]](#)
8. Holloway, J.; Wu, Y.; Sharma, M.K.; Cossairt, O.; Veeraraghavan, A. SAVI: Synthetic apertures for long-range, subdiffraction-limited visible imaging using Fourier ptychography. *Sci. Adv.* **2017**, *3*, e1602564. [\[CrossRef\]](#) [\[PubMed\]](#)

9. Glasner, D.; Bagon, S.; Irani, M. Super-resolution from a single image. In Proceedings of the 2009 IEEE 12th International Conference on Computer Vision, Kyoto, Japan, 29 September–2 October 2009; pp. 349–356.
10. Li, X.; Hu, Y.; Gao, X.; Tao, D.; Ning, B. A multi-frame image super-resolution method. *Signal Process.* **2010**, *90*, 405–414. [[CrossRef](#)]
11. Kato, T.; Hino, H.; Murata, N. Multi-frame image super resolution based on sparse coding. *Neural Netw.* **2015**, *66*, 64–78. [[CrossRef](#)] [[PubMed](#)]
12. Wang, Z.; Chen, J.; Hoi, S.C. Deep learning for image super-resolution: A survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**. [[CrossRef](#)] [[PubMed](#)]
13. Zou, Y.; Zhang, L.; Liu, C.; Wang, B.; Hu, Y.; Chen, Q. Super-resolution reconstruction of infrared images based on a convolutional neural network with skip connections. *Opt. Laser Eng.* **2021**, *146*, 106717. [[CrossRef](#)]
14. 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, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
15. Hardie, R.C.; Barnard, K.J.; Bognar, J.G.; Armstrong, E.E.; Watson, E.A. High-resolution image reconstruction from a sequence of rotated and translated frames and its application to an infrared imaging system. *Opt. Eng.* **1998**, *37*, 247–260. [[CrossRef](#)]
16. Elad, M.; Feuer, A. Restoration of a single superresolution image from several blurred, noisy, and undersampled measured images. *IEEE Trans. Image Process.* **1997**, *6*, 1646–1658. [[CrossRef](#)]
17. Zhang, X.; Huang, W.; Xu, M.; Jia, S.; Xu, X.; Li, F.; Zheng, Y. Super-resolution imaging for infrared micro-scanning optical system. *Opt. Express* **2019**, *27*, 7719–7737. [[CrossRef](#)] [[PubMed](#)]
18. Li, Y.; Hao, Z.; Lei, H. Survey of convolutional neural network. *J. Comput. Appl.* **2016**, *36*, 2508–2515.
19. Feng, S.; Chen, Q.; Gu, G.; Tao, T.; Zhang, L.; Hu, Y.; Yin, W.; Zuo, C. Fringe pattern analysis using deep learning. *Adv. Photonics* **2019**, *1*, 025001. [[CrossRef](#)]
20. Qiu, Y.; Wang, R.; Tao, D.; Cheng, J. Embedded block residual network: A recursive restoration model for single-image super-resolution. In Proceedings of the 2019 IEEE/CVF International Conference on Computer Vision (ICCV), Seoul, Korea, 27 October–2 November 2019; pp. 4180–4189.
21. Paola, J.D.; Schowengerdt, R.A. A review and analysis of backpropagation neural networks for classification of remotely-sensed multi-spectral imagery. *Int. J. Remote Sens.* **1995**, *16*, 3033–3058. [[CrossRef](#)]
22. Lin, Z.; Shum, H.Y. Fundamental limits of reconstruction-based superresolution algorithms under local translation. *IEEE Trans. Pattern Anal. Mach. Intell.* **2004**, *26*, 83–97. [[CrossRef](#)]
23. Yang, J.; Wright, J.; Huang, T.S.; Ma, Y. Image super-resolution via sparse representation. *IEEE Trans. Image Process.* **2010**, *19*, 2861–2873. [[CrossRef](#)]
24. Zhang, Y.; Liu, J.; Yang, W.; Guo, Z. Image super-resolution based on structure-modulated sparse representation. *IEEE Trans. Image Process.* **2015**, *24*, 2797–2810. [[CrossRef](#)]
25. Dai, S.S.; Liu, J.S.; Xiang, H.Y.; Du, Z.H.; Liu, Q. Super-resolution reconstruction of images based on uncontrollable microscanning and genetic algorithm. *Optoelectron. Lett.* **2014**, *10*, 313–316. [[CrossRef](#)]
26. Yang, C.Y.; Ma, C.; Yang, M.H. Single-image super-resolution: A benchmark. In Proceedings of the European Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; pp. 372–386.
27. Borman, S.; Stevenson, R.L. Super-resolution from image sequences—a review. In Proceedings of the 1998 Midwest symposium on circuits and systems (Cat. No. 98CB36268), Notre Dame, IN, USA, 9–12 August 1998; pp. 374–378.
28. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2015**, *38*, 295–307. [[CrossRef](#)] [[PubMed](#)]
29. Dong, C.; Loy, C.C.; Tang, X. Accelerating the super-resolution convolutional neural network. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 391–407.
30. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
31. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Deep laplacian pyramid networks for fast and accurate super-resolution. In Proceedings of the 2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
32. Lai, W.S.; Huang, J.B.; Ahuja, N.; Yang, M.H. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 2599–2613. [[CrossRef](#)] [[PubMed](#)]
33. Yu, J.; Fan, Y.; Yang, J.; Xu, N.; Wang, Z.; Wang, X.; Huang, T. Wide activation for efficient and accurate image super-resolution. *arXiv* **2018**, arXiv:1808.08718.
34. 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), Munich, Germany, 8–14 September 2018; pp. 286–301.
35. Du, J.; Zhou, H.; Qian, K.; Tan, W.; Zhang, Z.; Gu, L.; Yu, Y. RGB-IR cross input and sub-pixel upsampling network for infrared image super-resolution. *Sensors* **2020**, *20*, 281. [[CrossRef](#)] [[PubMed](#)]
36. He, Z.; Tang, S.; Yang, J.; Cao, Y.; Yang, M.Y.; Cao, Y. Cascaded deep networks with multiple receptive fields for infrared image super-resolution. *IEEE Trans. Circuits Syst. Video Technol.* **2018**, *29*, 2310–2322. [[CrossRef](#)]

-
37. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
 38. Han, T.Y.; Kim, D.H.; Lee, S.H.; Song, B.C. 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. [[CrossRef](#)]