*sustainability*

MDPI

*Article*

# Deep USRNet Reconstruction Method Based on Combined Attention Mechanism

**Long Chen [1,2], Shuiping Zhang [1,2,*], Haihui Wang [1,2], Pengjia Ma [1,2], Zhiwei Ma [1,2] and Gonghao Duan [1,2]**

[1] Hubei Provincial Key Laboratory of Intelligent Robots, Wuhan Institute of Technology, Wuhan 430205, China
[2] School of Computer Science and Engineering, Wuhan Institute of Technology, Wuhan 430205, China
* Correspondence: zhangshuiping2007@126.com

**Abstract:** Single image super-resolution (SISR) based on deep learning is a key research problem in the field of computer vision. However, existing super-resolution reconstruction algorithms often improve the quality of image reconstruction through a single network depth, ignoring the problems of reconstructing image texture structure and easy overfitting of network training. Therefore, this paper proposes a deep unfolding super-resolution network (USRNet) reconstruction method under the integrating channel attention mechanism, which is expected to improve the image resolution and restore the high-frequency information of the image. Thus, the image appears sharper. First, by assigning different weights to features, focusing on more important features and suppressing unimportant features, the details such as image edges and textures are better recovered, and the generalization ability is improved to cope with more complex scenes. Then, the CA (Channel Attention) module is added to USRNet, and the network depth is increased to better express high-frequency features; multi-channel mapping is introduced to extract richer features and enhance the super-resolution reconstruction effect of the model. The experimental results show that the USRNet with integrating channel attention has a faster convergence rate, is not prone to overfitting, and can be converged after 10,000 iterations; the average peak signal-to-noise ratios on the Set5 and Set12 datasets after the side length enlarged by two times are, respectively, 32.23 dB and 29.72 dB, and are dramatically improved compared with SRCNN, SRMD, PAN, and RCAN. The algorithm can generate high-resolution images with clear outlines, and the super-resolution effect is better.

**Keywords:** super-resolution; USRNet network; attention mechanism; multi-channel mapping; loss function



**Citation:** Chen, L.; Zhang, S.; Wang, H.; Ma, P.; Ma, Z.; Duan, G. Deep USRNet Reconstruction Method Based on Combined Attention Mechanism. *Sustainability* **2022**, *14*, 14151. https://doi.org/10.3390/su142114151

Academic Editors: Haoran Wei, Zhendong Wang, Yuchao Chang and Zhenghua Huang

Received: 22 September 2022
Accepted: 26 October 2022
Published: 30 October 2022



## 1. Introduction

People obtain all kinds of information through vision in life, and images are an important source that allow people to obtain a large amount of information. With the vigorous development of computers and the urgent need to process image information quickly, people began to use computers to process images. However, for many reasons, such as hardware equipment or a bad environment, the image resolution and the amount of information is reduced. Thus, improving the resolution of the image is a very important topic in the field of image enhancement. At present, following the in-depth research on super-resolution reconstruction technology, this technology has important application value in many fields, such as facial recognition, video clarity reconstruction, medical image processing and remote sensing satellite image processing.

Traditional image super-resolution refers to an algorithm that restores high-resolution images, rather than an algorithm that uses deep learning. It is mainly divided into the following three types: super-resolution based on a reconstructed single image, learning-based single image super-resolution, and interpolation-based super-resolution algorithms [1]. The main purpose of the single image super-resolution reconstruction algorithm is to infer the corresponding high-resolution image based on the low-resolution image, which is a

typical inversion problem. The core idea of the super-resolution algorithm based on the reconstructed single image is: first, it is assumed that the low-resolution image is obtained by the high-resolution image through downsampling, adding noise and other degradation models, and then using this as a constraint to establish an image prior model to optimize and solve, and to reconstruct a high-resolution image. The difficulty of this algorithm lies in how to build a good mathematical model. Although the reconstruction-based super-resolution algorithm preserves more image details, the amount of computation is too large, and when the upsampling factor $r$ is greater than four, the gap between the prior hypothesis and the actual situation will be too large, resulting in unsatisfactory reconstruction results. The core idea of the learning-based super-resolution algorithm is: first possess the image dataset to be trained for a given scene, then learn the mapping relationship between high- and low-resolution images according to the training set, and then build a nonlinear model according to the mapping relationship, and finally use the nonlinear model to reconstruct the input low-resolution image into a high-resolution image. In 2013, Timofte et al. [2] proposed the Anchored Neighborhood Regression (ANR) algorithm by combining the two methods of sparse dictionary and neighborhood embedding, and adjusted ANR in 2014 to obtain the algorithm A+ [3]. In 2015, Huang et al. [4] proposed an improved self-similar method SelfExSR (Self-Exemplars Super Resolution), which expanded the search range of its own internal image blocks. The core idea of the interpolation-based super-resolution algorithm is: first, add a low-resolution image to the grid with the same size as the high-resolution image, and then use a mathematical model to calculate the pixel value of the point to be interpolated based on the surrounding pixels. The advantages of this type of method are less computation, low complexity, and easy implementation, so it is widely used. Common interpolation techniques in traditional super-resolution reconstruction algorithms include nearest neighbor interpolation [5], bilinear interpolation [6], and Bicubic Interpolation [7]. Although the super-resolution reconstruction based on interpolation is not ideal, considering the excellent performance of the interpolation method in real-time scenes, the bicubic interpolation method is used to downsample the high-resolution image in the data preprocessing stage, so that it becomes a low-resolution image and is input to the network model for training.

In recent years, deep-learning-based super-resolution algorithms have made significant progress, but they still have some shortcomings. Many researchers have proposed various different and effective models, which can better improve the super-resolution effect. Dong et al. first proposed the SRCNN (Super-Resolution Convolutional Neural Network), which was the first network to apply convolutional layers to super-resolution tasks. Although the structure is simple, it has inspired researchers a lot [8]. It achieved image reconstruction by using three-layer convolutional neural networks to represent feature extraction, nonlinear mapping and final reconstruction functions, respectively. To this end, the FSRCNN (Fast Super-Resolution Convolutional Neural Networks) proposed by Dong et al. [8] in 2016 used a deconvolution layer to enlarge the size in the final reconstruction module to reduce the computational load of the network model, but deconvolution may result in a pixel-overlaid checkerboard phenomenon in the reconstructed image [9]. In addition, Kim et al. [10] proposed the DRCN (Deeply Recursive Convolutional Network) based on the VDSR network, and used the RNN (Recursive Neural Network) structure in the super-resolution network for the first time, but when the upsampling factor r is eight, the reconstructed image quality is not good. Furthermore, in 2018, Li et al. [11] proposed the MSRN (Multi-scale Residual Network) which combined local multi-scale features with global features to solve the problem of feature loss during transmission, but its reconstructed structure will lose the feature information of the original image. In 2020, the CFSRCNN (Coarse-to-Fine Super-Resolution Convolutional Neural Network) proposed by Tian et al. [12] used multiple refinement modules to increase the stability of the model, but the feature extraction capabilities of these two networks are insufficient, and the reconstruction effect still needs to be improved. In 2021, Qiao et al. [13] comprehensively evaluated the performance of existing super-resolution convolutional neural

network models on microscopic image super-resolution tasks; proposed the adversarial network model generated by Fourier domain attention convolutional neural network, which achieved optimal super-resolution prediction of microscopic images and super-resolution reconstruction of structured light under different imaging conditions; and achieved a more robust prediction effect of microscopic images than other existing convolutional neural network models. This model can replace the existing super-resolution imaging methods in actual biological imaging experiments, and its application scenarios have been greatly expanded. The super-resolution reconstruction method using deep learning can further help the network pay more attention to the geometry of the image, which is expected to improve the image resolution and restore the high-frequency information of the image.

Based on the USRNet (Unfolding Super-Resolution Network), the method in this paper makes changes to the network structure, considering the interdependence between feature channels to adapt and readjust features, and introducing the channel attention mechanism module, which can achieve deeper depth than previous CNN-based methods and obtains better super-resolution performance. Furthermore, many researchers have also introduced various attention mechanisms in super resolution to improve the effect of super-resolution. Wang et al. proposed the use of nonlocal neural networks to apply nonlocal operations in terms of spatial attention, which greatly improved the experimental results [14]. Hu et al. proposed SENet (Squeeze-and-Excitation Networks) to exploit the relationship between channels to improve performance in object classification [15]. Because SENet only utilizes first-order information, Dai et al. used SAN (Second-order Attention Network) with second-order feature statistics to obtain more discriminative feature representations [16]. Inspired by SE-Block (Squeeze-and-Excitation block), Zhang et al. proposed the Image Super-Resolution Using Very Deep Residual Channel Attention Networks [17], which introduced the channel attention mechanism into the SR task and achieved good qualitative and quantitative results. The channel attention mechanism used global average pooling to extract channel statistics. Zhao et al. proposed PAN (Pyramid Attention Network) and CBAM (Convolutional Block Attention Module) by introducing a pixel attention mechanism. Unlike CBAM, the module PA was used to generate a 3D attention map instead of a 1D or 2D attention vector diagram [18]. This attention strategy introduced fewer parameters, but produced better SR results. Since the loss function of most SR models is the mean square error loss, it is easy to make the image after super-resolution recovery appear smooth and lose details, so Ledig et al. proposed SRGAN (Super Resolution Generative Adversaria Network) to use a generative adversarial network on the super-resolution reconstruction problem, and on this basis, perceived loss function instead of the mean square error loss were used as the target loss function to improve the perceptual quality of the image [19]. In 2021, Gao et al. proposed a mutually supervised few-shot segmentation network; the graph attention network is adopted to avoid losing spatial information and increase the number of pixels in the support image to guide the query image segmentation. This method is first applied to semantic segmentation. It is time consuming, because training a model requires a large number of pixel-level annotated samples [20,21]. Gao et al. proposed a deep feature and attention mechanism-based method for health assessment, which aims to apply HDLGN (Hand-Deep Local–Global Net) for image recognition. The local attention mechanism is introduced to identify key areas of the image, and color features are extracted to learn deep features. This method has a low recognition rate when the number of samples is limited and the knowledge is limited [22].

In order to better solve this problem, a large number of deep learning-based methods have been proposed for learning the mapping of low-resolution images to high-resolution images. Through research, it can be found that a large number of experiments have proved that traditional methods and deep learning-based methods have their advantages and disadvantages, and this study combines the advantages of the two:

(1) Considering the interdependence between feature channels to adaptively readjust the features, the channel attention mechanism is introduced, so that high-frequency

information and low-frequency information are input into subsequent convolutional layers with different weights, which is helpful for the optimization of the algorithm.

(2) Combining with the end-to-end USRNet network model for training, so as to ensure the effectiveness and efficiency, so that the features extracted by the network are richer, the robustness to image size is enhanced, and the network performance is improved.

## 2. Relevant Work

In the process of super-resolution reconstruction, since low-resolution image sequences are often affected by optical blur, motion blur, noise level, and aliasing factors, super-resolution reconstruction techniques cover image restoration techniques. The difference between the two is that image restoration technology restores an image without changing the size of the image, so image restoration technology and image super-resolution reconstruction are quite closely related, and it can be considered that image super-resolution reconstruction is a theoretical second-generation image restoration problem. On the one hand, the study of image super-resolution reconstruction technology has important theoretical significance to promote the further development of image restoration technology; on the other hand, it has important practical significance to overcome the limitations of optical imaging system hardware. In some cases, the original low-resolution imaging system can still be used. In the case of training on smaller datasets, images that meet specific resolution requirements can still be obtained.

### *2.1. Degradation Model*

In recent years, deep convolutional neural network (CNN) methods have made very great progress in the field of single image super-resolution (SISR). However, the existing CNN-based SISR methods mainly assume that the low-resolution (LR) image is obtained from the high-resolution (HR) image by bicubic downsampling, so when the degradation process of the real image does not follow this assumption, the super-resolution result will be unsatisfactory.

The degradation model is critical to the success of SISR, as it defines how LR images are degraded from HR images [22,23]. In addition to the classical degradation model and the bicubic degradation model, in the work of earlier researchers, the degradation model assumed that the LR image was directly downsampled from the HR image without blurring, which corresponds to the image interpolation problem [24]. By assuming that the bicubic downsampled HR images are also sharp, the degradation model is considered as a synthesis of deblurring on LR images and SISR with bicubic degradation. Although many degradation models have been proposed, the classical degradation model SISR based on CNN has received little attention and deserves further study. The degradation of the image can be described by the linear system as shown in Figure 1, and it is the model process diagram of the image degradation-restoration process, and its mathematical expression is:

$$g(x,y) = f(x,y) * h(x,y) + n(x,y) \tag{1}$$

where $g(x,y)$ is the degraded image, $f(x,y)$ is the original image, $h(x,y)$ is the point spread function (PSF) of the image degradation model, $n(x,y)$ is the noise that is not correlated with the original image, and * represents the convolution operation. In this model, the process of image degradation is modeled as $f(x,y)$ with $h(x,y)$, and acts in conjunction with $n(x,y)$ to produce $g(x,y)$. The task of image recovery is to make some estimation of $f(x,y)$ from the degraded image $g(x,y)$ based on some prior knowledge of $f(x,y)$, $h(x,y)$ and $n(x,y)$. The Fourier transform of the degraded model is as follows:

$$G(u,v) = F(u,v) * H(u,v) + N(u,v) \tag{2}$$

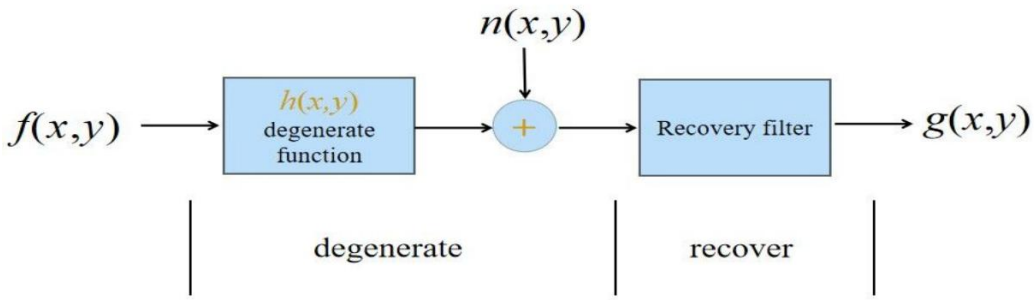



**Figure 1.** Model of image degradation/restoration process.

*2.2. SISR Method*

CNN-based super-resolution reconstruction methods have achieved good results in handling bicubic degradation models, but applying them to other more practical problems is not very effective. Considering the feature of practicability, a flexible super-resolution reconstruction algorithm is designed, which mainly considers the following three key factors: scale factor, blur kernel, and noise level. Researchers have proposed several methods to solve the bicubic degradation problem with different scale factors through a single model, such as LapSR (Laplacian Pyramid Super Resolution) with progressive upsampling [25], MDSR (Multi-scale Super Resolution) with specific scale factors [26], and Meta-SR (Meta-Super-Resolution) with amplification module [27]. These methods are limited to Gaussian blur kernel. The CNN-based super-resolution reconstruction method can deal with various blur kernels, scale factors and noise levels, which is the deep plug-and-play method. The main idea of this approach is to insert the learned CNN prior knowledge into iterative solutions under the MAP (Maximum a Posteriori) framework, which are basically model-based methods that are computationally expensive and involve manually selected hyperparameters.

Currently, learning-based blind image restoration has received considerable attention [28–32], but we note that the goal of blind image restoration is to find an inverse transformation such that the blurred image $g(x, y)$ can be transformed to obtain the original clear image $f(x, y)$, as shown in Figure 1. Because both the point spread function and the clear image are unknown, there are infinitely many solutions, which are ill posed, so deconvolution is an ill-conditioned problem. The presence of noise further reinforces this ill-posedness. Existing image deblurring optimization algorithms rely too much on the prior knowledge of images and convolution kernels. Both the blind restoration method and the nonblind restoration method need to make certain assumptions about the prior knowledge of the image or blur kernel, and then a repeated iterative solution is performed, and repeated iterations will result in a large amount of computation. If the estimated blur kernel is not accurate enough, the restoration result will be poor.

Blind image super resolution aims to super-enhance low-resolution images of unknown degradation types, and after input observation images, they are divided into three categories through blur-type identification. They are motion blur, defocus blur and other types, among which the former two can be restored by the parameter method, the latter can be restored by iterative method, and the restored images need to be processed and evaluated through the ringing effect, as shown in Figure 2.

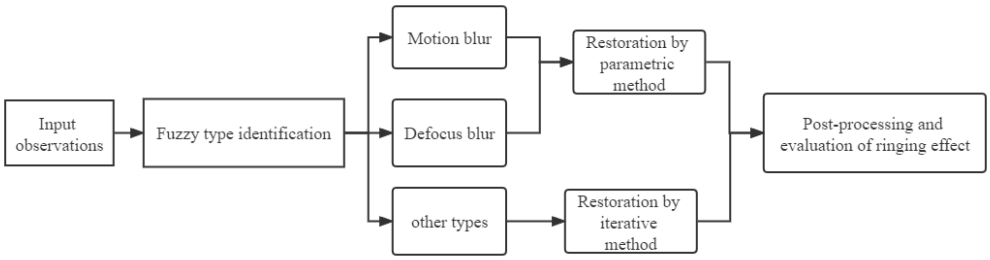



**Figure 2.** Processing framework of image blind restoration problem.

### *2.3. Attention Mechanism Method*

The attention mechanism is summed up by the laws of humans' habit of observing the environment. When human beings observe the environment, it is transmitted to the brain through vision. The brain only focuses on some particularly important local parts, identifies the characteristic information of the object that needs to be obtained, and constructs a certain description about the object in the environment. The initial attention mechanism was first introduced into natural language processing and achieved good results. Later, researchers applied attention to deep learning computer vision. The network model can learn the importance of different parts of the image through the attention mechanism, and then combine them to improve performance.

The attention mechanism pays attention to the high weight of the network model; that is, the high degree of importance. This mechanism will allow the network to extract more attention and more needed feature information, so that the model can achieve better results when predicting. Another advantage of this is that while improving the effect, the resource consumption and memory pressure of the lock will not be too great. It is based on these two advantages that the attention mechanism is widely used. The attention model has been widely used in various fields of deep learning in recent years. In image processing, speech recognition, or various types of natural language processing tasks, it is easy to encounter the existence of the attention model. Furthermore, understanding the working principle of the attention mechanism is very necessary for technicians who are concerned about the development of deep-learning technology.

The attention mechanism can make the neural network focus on the features with high importance. There are two main types of attention mechanisms: channel attention and spatial attention. Channel attention mainly explores the dependencies between channels. Spatial attention can capture the relationship between pixels. Of course, there are also mixed dimensions that add attention mechanisms in both the spatial dimension and the channel dimension at the same time. Attention mechanism has been deeply studied by many researchers.

## 3. Deep USRNet Algorithm Based on Attention Algorithm

### *3.1. Network Model*

For image super-resolution, the degradation process of LR image is generally described by the following Formula (3):

$$\mathbf{y} = (\mathbf{x} \otimes \mathbf{k}) \downarrow_{\mathbf{s}} + \mathbf{n} \tag{3}$$

where **x** denotes the degraded blur kernel, **s** denotes downsampling, and **n** denotes additive white Gaussian noise. It can be seen that the above degradation process includes blurring, downsampling and noise, while the traditional image super-resolution algorithm only considers the downsampling of the blur kernel. Among them, the most studied is bicubic interpolation degradation. In fact, bicubic interpolation degradation can select a suitable fuzzy approximation through the above formula. At the same time, this paper can solve the above kernel estimation problem in a data-driven way, and the optimization objectives are as follows:

$$\mathbf{k}_{\text{bicubic}}^{\times \mathbf{s}} = \arg\min_{\mathbf{k}} \|(\mathbf{x} \otimes \mathbf{k}) \downarrow_{\mathbf{s}} - \mathbf{y}\| \tag{4}$$

USRNet is based on the flexibility of model-based methods and the advantages of learning-based methods. For the first time, a single end-to-end model is used to deal with classical degradation models with different scale factors, blur kernels, and noise levels. Due to this, bicubic degradation has been well researched; it is very important to study its relationship with the classical degradation model. Actually, the bicubic degradation can be approximated by setting a suitable blur kernel in Formula (4).

According to the MAP framework, the HR picture can be estimated by the energy equation below. First, the mapping reasoning is expanded through the semi-quadratic splitting algorithm, and a fixed number of iterations consisting of alternately solving the data sub-problem and the prior sub-problem can be obtained by Formulas (5) and (6). Then, these two sub-problems can be solved by the neural module of Formula (7). Thus, the proposed network inherits the flexibility of model-based methods and can super resolve blurred, noisy images of different scale factors using a single model, while keeping the advantages of learning-based approaches.

$$E_{\mu}(\mathbf{x},\mathbf{z}) = \frac{1}{2\sigma^2}\|\mathbf{y} - (\mathbf{z} \otimes \mathbf{k}) \downarrow_{\mathrm{s}}\|^2 + \lambda\varphi(\mathbf{x}) + \frac{\mu}{2}\|z - \mathbf{x}\|^2 \tag{5}$$

$$E_{\mu}(\mathbf{x},\mathbf{z}) = \frac{1}{2\sigma^2}\|\mathbf{y} - (\mathbf{z} \otimes \mathbf{k}) \downarrow_{\mathrm{s}}\|^2 + \lambda\varphi(\mathbf{x}) + \frac{\mu}{2}\|z - \mathbf{x}\|^2 \tag{6}$$

$$\begin{cases} \mathbf{z}_k = \arg\min_{\mathbf{z}} \|\mathbf{y} - (\mathbf{z} \otimes \mathbf{k}) \downarrow_{\mathbf{s}}\|^2 + \mu\sigma^2\|\mathbf{z} - \mathbf{x}_{k-1}\|^2 \\ \mathbf{x}_k = \arg\min_{\mathbf{x}} \frac{\mu}{2}\|\mathbf{z}_k - \mathbf{x}\|^2 + \lambda\varphi(\mathbf{x}) \end{cases} \tag{7}$$

Figure 3 is the USRNet network model structure, and the meanings of the input on the left side of the network structure are: $\mathbf{y}$ is the input low-resolution image, $\mathbf{k}$ is the fuzzy kernel,$\sigma$ is the noise level, and $\mathbf{s}$ is the image scaling factor. The whole model has three modules: the data module (Formula (5)), the prior module (Formula (6)), and the hyperparameter module, which are model-based super resolution, learning-based denoising, and hyperparameter prediction, respectively. The structure of the prior module is a ResUNet, and the channel attention mechanism module is introduced after sampling on the ResUNet network, assigning different weights to each channel, so that the network pays attention to important features and suppresses unimportant features; the specific details are Section 3.3.

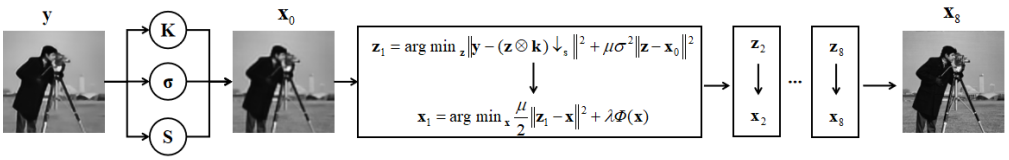



**Figure 3.** USRNet schematic diagram of the model Structure.

The whole process is as follows: first, the preset noise level $\sigma$ and scaling factor $\mathbf{s}$ are used as the input of the hyperparameter module to predict the hyperparameters; then, the image $\mathbf{y}$ is upsampled to the same size as $\mathbf{x_k}$ at the final output, as the initial input $X_0$ of the iteration. Finally, $\mathbf{x_0}$,$\mathbf{s}$,$\mathbf{k}$,$\mathbf{y}$,$\alpha$ are used as the input of the data module (Formula (5)) to obtain $\mathbf{z_1}$. Next, the solution obtained by the data module and the predicted hyperparameters are taken as input, and sent to the prior module to obtain an iteration of $\mathbf{x_1}$. Finally, the obtained $\mathbf{x}$ is sent to the next round of iterative calculation, and the HR image is generated.

### *3.2. Network Model with Attention Mechanism*

Attention models have been widely used in various fields of deep learning in recent years. Whether conducting image recognition, speech recognition or various types of natural language processing tasks, it is easy to encounter attention models. The attention mechanism module focusing on high frequency is called CA (Channel Attention), which is concluded after referring to the two models of RCAB (Residual channel attention block) [26]

and CBAM (Convolutional Block Attention Module). It consists of an average pooling layer, two convolutional layers and a ReLu activation function to form a module, while the other module consists of a max pooling layer, two convolutional layers and a ReLu activation function. The outputs of the two modules are added to form a residual block. The output is obtained by multiplying the input **x** and the output **y** processed by the convolutional layer. This method can make important channels, i.e., high-frequency features, have greater weights. The channel weight of the part with a small improvement in the quality of the image effect is reduced. The essence of the channel attention mechanism is to model the importance of each feature. For different tasks, feature assignment can be performed according to the input, which is simple and effective.

SENet (Sequeeze and Excitation Net) is essentially a channel-based attention model. It models the importance of each feature channel, and then enhances or suppresses different channels for different tasks. The network structure diagram is shown below (Figure 4).

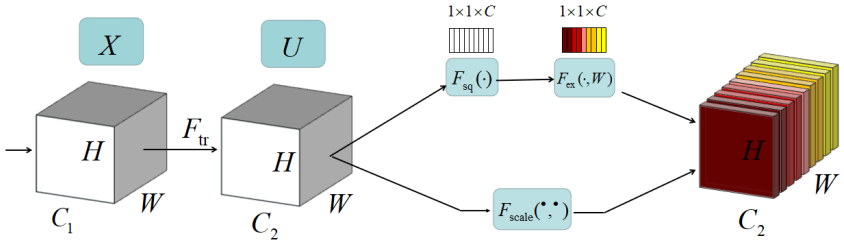



**Figure 4.** SE network structure diagram.

Given an input $X$, the number of its feature channel is $C_1$, after a series of general transformations such as convolution, a feature with the number of feature channel $C_2$ is obtained. Unlike the traditional CNN, the features obtained earlier are recalibrated through three operations.

The first is the Squeeze operation (i.e., $F_{sq}(\cdot)$ in the figure), which performs feature compression along the spatial dimension, and turns each two-dimensional feature channel into a real number. This real number has a global receptive field to some extent, and the dimensions of the output match the number of feature channels of the input.

The second is the Excitation operation (i.e., $F_{ex}(\cdot)$ in the figure). It generates weights for each feature channel through the parameter W, which is used to explicitly model the correlation between feature channels. In the experiment, a fully connected layer with a two-layer bottleneck structure (dimension down first, then up) and a Sigmoid function are used to achieve this.

The last is the operation of Reweight. The weight of the output of Excitation is regarded as the importance of each feature channel after feature selection, and then, channel by channel weighting is performed on the previous features by multiplication to complete the re-calibration of the original features in the channel dimension.

### *3.3. AttentionResNet Network Structure*

At present, the research methods of super-resolution mainly include traditional-based and deep learning-based image super-resolution algorithms. Although the deep-learning-based method has achieved great progress in the reconstruction quality and reconstruction efficiency of the single-frame super-resolution field compared with the traditional algorithm, the learning ability of the deep learning method is mainly determined by the quality of the training data, while the training data of existing super-resolution network models are all artificially synthesized. For example, during downsampling, a predefined blur kernel (such as bicubic downsampling) is used to make a low-resolution dataset, while the practical application the blur kernels involved in real scenes are very complex and unknown. The difference in the data distribution between the dataset used in the actual scene and the dataset used for training can lead to a severe drop in super-resolution performance.

Furthermore, how to only use the existing low-resolution images acquired in real scenes to achieve superior super-resolution in real scenes is still a challenging problem [33,34].

Based on the above problems, the channel attention mechanism is integrated to optimize the image super-resolution algorithm. The end-to-end expandable network model USRNet is used, which uses a single end-to-end training model to process different scale factors, blur kernels and noise levels of classical degradation models. The network has the advantages of both model-based methods and learning-based methods.

(1) First, a channel attention mechanism module is introduced into the deep expansion network, which assigns different weights to each channel, allowing the network to focus on important features and suppress unimportant features. It can be seen from Figure 5: the feature map is globally pooled, the next layer is input through the convolution of 1 × 1 to reduce the channel, a ReLU function is performed, and then a convolution of 1 × 1 is input for channel expansion, which is mapped into a real number (0,1) by the Sigmoid function, multiplied by the corresponding feature map, and processed by the channel attention mechanism.

(2) Then, the attention mechanism module is used to pay attention to the detailed texture information that is difficult to recover in the reconstruction process, and suppress the interference features.

(3) Finally, in order to alleviate the vanishing gradient problem and speed up the training, the residual structure and attention mechanism are introduced to reconstruct the image. The residual network structure diagram is shown below.

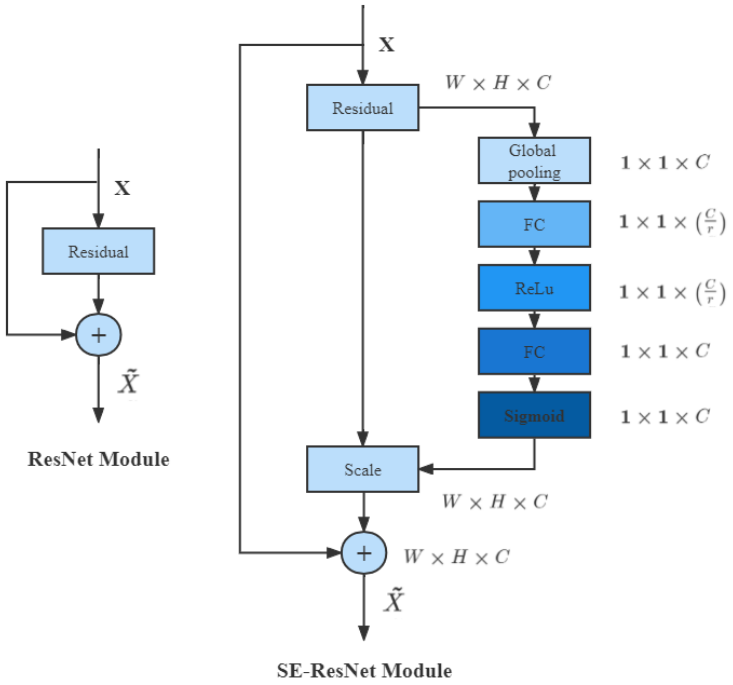



**Figure 5.** Residual module diagram.

Through the comparison test of ResNet and SE-ResNet, as shown in Figure 6, it is obvious from the figure that the ResNet after adding the SE module greatly reduces the error rate.

Therefore, in this paper, the super-resolution network based on USRNet is prone to artifacts in high-frequency details and is integrated into the channel attention mechanism. The attention mechanism pays attention to the part of the network model with high weight; that is, a high degree of importance. At the same time, because the deep learning convolutional neural network is too deep, there are basically problems of too long training

time and low super-resolution efficiency. So the unnecessary batch normalization operation convolution layer is deleted from the generative model. On the discriminant model, the original discriminator of the USRNet network is used to guide the training of the super-resolution model. The L1 loss function (Mean Absolute Error, MAE) is used on the loss function to further improve the visual effect of image super-resolution, and the image super-resolution algorithm incorporating the channel attention mechanism is optimized. Most of the current super-resolution algorithms generally use L1 (Mean Absolute Error, MAE), and L2 (Mean Squared Error, MSE)loss functions. The L2 loss function calculates the average of the squared differences between the actual and predicted values, which makes the reconstruction results tend to be smooth. It is necessary to make the network pay more attention to the reconstruction of high-frequency regions.

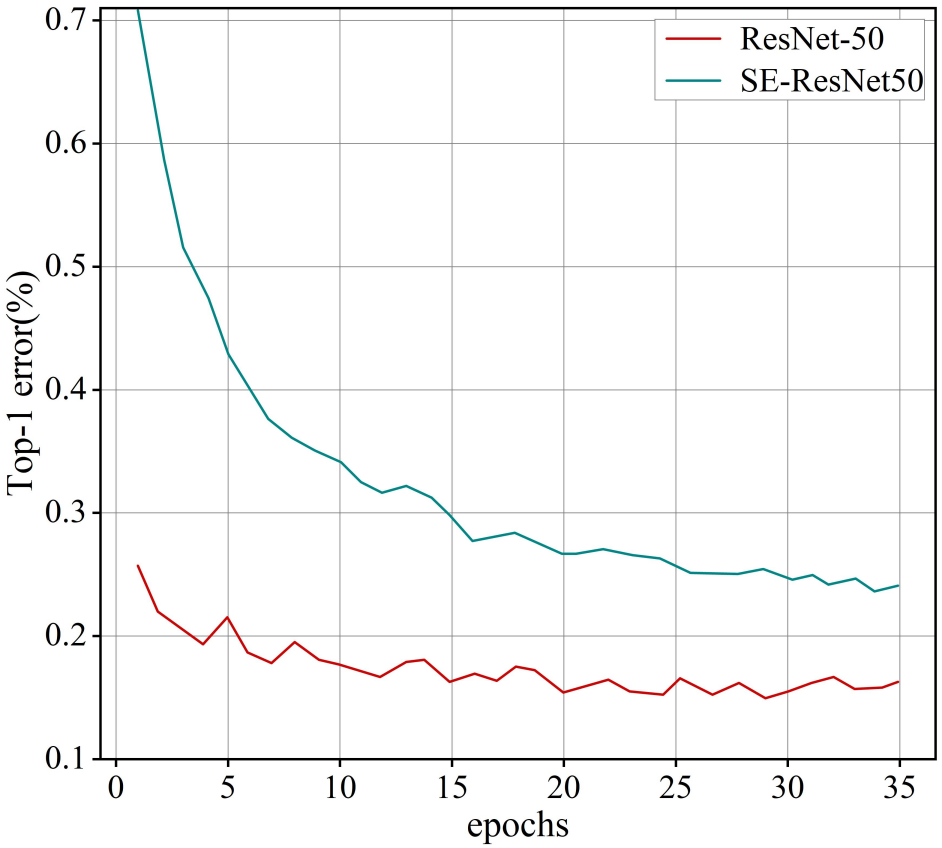



**Figure 6.** Adding SE module training diagram.

Fritsche et al. used high-pass filtering to extract the high-frequency details of the image and boosted the high-frequency details of the image by confronting the high-frequency details extracted from the generated image and the real image. Inspired by this, a high-frequency attention loss is proposed. Unlike that of Fritsche et al., this algorithm does not use adversarial loss, but pre-extracts high-frequency detail locations of training images and gives these locations additional weights, so that the network pays more attention to these areas during the training process.

Channel attention is mainly used to explore the dependencies between channels, and pay attention to the part of the network model with high weight; that is, the high degree of importance. This mechanism will allow the network to extract more attention and more needed feature information, so that the model can achieve better predictions. While it improves the effect, the resource and memory pressure consumed by the lock will not be too great. Based on these two advantages, an end-to-end deep expansion network AttentionResNet with attention mechanism is proposed. The network structure is shown in Figure 7:

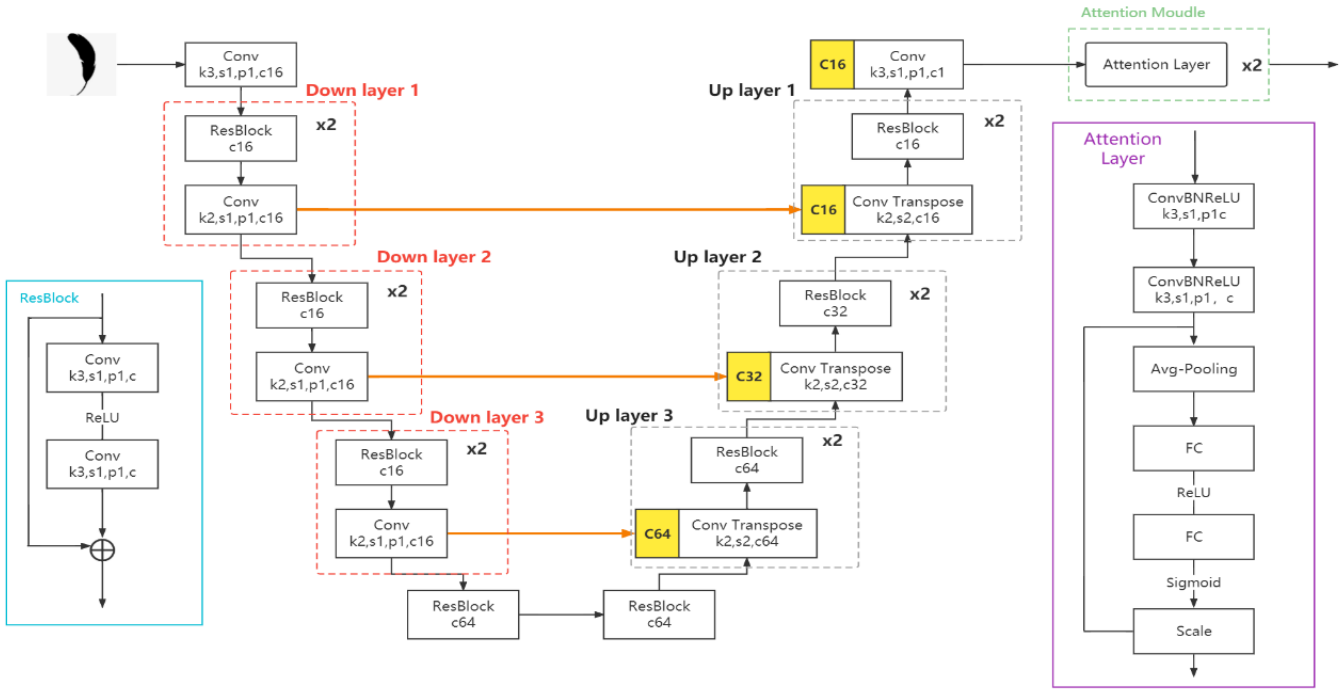



**Figure 7.** AttentionResNet network structure diagram.

## 4. Experiment

### *4.1. Experimental Environment and Its Dataset*

The GPU used in this experiment is NVIDIA GeForce GTX 2080Ti, the programming language is Python, the IDE is Pycharm 2021, the deep learning framework uses PyTorch1.5 or later, the graphics processing uses Python-based OpenCV, and the visualization uses Matplotlib. The above toolkits are all based on Python and the operating system is Windows 10.

The Train400 and trainH datasets are used for training, the magnifications are $\times 1$, $\times 2$, $\times 3$, and $\times 4$, and the three datasets of Set5 [35], Set14 [36], and real_faces are used for testing, as shown in Figure 8. The PAN (Pyramid Attention Network), RCAN (Residual Channel Attention Networks), SRMD (Super-Resolution Network for Multiple Degradations) and SRCNN (Super-Resolution Convolutional Neural Network) are compared horizontally. This experiment is a model modified on the basis of USRNet, so the focus is on comparing the PSNR value, SSIM value and the visual effect of the picture with the USRNet algorithm.

### *4.2. Training Details*

Using Train400 and trainH as the HR training datasets, the LR dataset is obtained by Formula (1). The scale factor **s** can be selected from 1,2,3,4. For fuzzy cores, we fix the kernel size at 25 × 25. For the noise level, we set its range to [0, 25], the patch size of the HR image of USRNet is set to 96 × 96, the learning rate of USRNet is 0.0001 and in terms of loss function, USRNet uses the L1 loss function.

To optimize the parameters of USRNet, we use the Adam optimizer with a mini batch size of 128. It is worth pointing out that due to the infeasibility of parallel computing for different scale factors, each mini batch involves only one random scale factor.

### *4.3. Evaluation Criteria*

In this experiment, the Peak signal-to-noise ratio (PSNR) and Structural similarity (SSIM) are used to evaluate the quality of the image.

**Figure 8.** The pictures in the test set Set12 and Set5.

4.3.1. PSNR

PSNR is a commonly used evaluation index in the field of image super resolution. It is mainly used to indicate the quality of heavy images. Using L2 as the loss function, calculating the difference between the reconstructed image and the labeled image, PSNR is negatively correlated with MSE. The smaller the MSE value, the larger the PSNR value, and the better the quality of the reconstructed image. For the monochrome high-definition original image and the super-resolution image obtained by MSE, H and W are the height and width of the image, respectively.

$$MSE = \frac{1}{H \times W} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [X(i,j) - Y(i,j)]^2 \tag{8}$$

For the three-color high-definition original image and the super-resolution image, each pixel has three channels, so the formula is:

$$MSE = \frac{1}{3WH} \sum_{i=0}^{W-1} \sum_{j=0}^{H-1} [X(i,j)_{R,G,B} - Y(i,j)_{R,G,B}]^2 \tag{9}$$

$$PSNR = 10 \log_{10} \left( \frac{\left(2^k - 1\right)^2}{MSE} \right) \tag{10}$$

Among them, $MSE$ represents the mean square error of the current image $X$ and the $Y$ reference image. The unit of PSNR is dB, and the larger the value, the smaller the distortion.

PSNR is the most common and widely used image objective evaluation index, but it is based on the error between corresponding pixels; that is, it is based on error-sensitive image quality evaluation. Since the visual characteristics of the human eye are not considered (the human eye is more sensitive to contrast differences with lower spatial frequencies, and the human eye is more sensitive to luminance contrast differences than chromaticity; the human eye's perception of an area is affected by the surrounding area, etc.), so the evaluation results are often inconsistent with people's subjective feelings.

4.3.2. SSIM

Since the PSNR calculates the error of the pixels at the corresponding positions of the reconstructed image and the label image, it is an image quality evaluation method based on the difference of gray values, and does not take into account the human visual perception.

Sometimes, the PSNR value is high, but the reconstructed image quality is unsatisfactory. This is because the human visual system is also sensitive to the position information of the gray value. Therefore, another objective evaluation index SSIM needs to be introduced.

Compared with PSNR, SSIM is more suitable for the human visual perception system, considering the brightness information, contrast information and structural information of the image. The SSIM value is positively correlated with the quality of the reconstructed image, with a minimum value of zero and a maximum value of one. The closer the SSIM value is to one, the higher the quality of the reconstructed image $X$ will be, and the closer it is to the label image $Y$. The formula of structural similarity SSIM is based on the mutual measurement of three indicators between $X$ and $Y$ under the sample: $\updownarrow$ (luminance),⌋(contrast)and $\int$ (structure). The formula is:

$$l(X,Y) = \frac{(2\mu_X\mu_Y + C_1)}{(\mu_X^2 + \mu_Y^2 + C_1)} \tag{11}$$

$$c(X,Y) = \frac{(2\sigma_X\sigma_Y + C_2)}{(\sigma_X^2 + \sigma_Y^2 + C_2)} \tag{12}$$

$$s(X,Y) = \frac{\sigma_{XY} + C_3}{\sigma_X\sigma_Y + C_3} \tag{13}$$

where $\mu_X$ and $\mu_Y$ represent the mean values of images $X$ and $Y$, respectively, $\sigma_X$ and $\sigma_Y$represent the variances of images $X$ and $Y$, respectively, and $\sigma_{XY}$ represent the covariances of images $X$ and $Y$, namely:

$$\mu_X = \frac{1}{H \times W}\sum_{i=1}^{H}\sum_{j=1}^{W} X(i,j) \tag{14}$$

$$\sigma_X^2 = \frac{1}{H \times W - 1}\sum_{i=1}^{H}\sum_{j=1}^{W}((i,j) - \mu_X)^2 \tag{15}$$

$$\sigma_{XY} = \frac{1}{H \times W - 1}\sum_{i=1}^{H}\sum_{j=1}^{W}((X(i,j) - \mu_X)(Y(i,j) - \mu_Y)) \tag{16}$$

$C_1$, $C_2$ and $C_3$ are constants. In order to avoid the situation where the denominator is zero, we usually take $C_1 = (K_1 * L)^2, C_2 = (K_2 * L)^2, C_3 = C_2/2$; generally, $K_1 = 0.01$, $K_2 = 0.03$, $L = 255$. Combining the three calculation formulas, the calculation formula of SSIM is as follows:

$$\mathrm{SSIM}(X,Y) = [l(X,Y)]^{\alpha} \cdot [c(X,Y)]^{\beta} \cdot [s(X,Y)]^{\gamma} \tag{17}$$

where $\alpha = \beta = \gamma = 1$, the value range of SSIM is [0, 1]. The larger the value, the smaller the image distortion. In practical applications, a sliding window can be used to divide the image into blocks, so that the total number of blocks is N. Considering the influence of the shape of the window on the blocks, Gaussian weighting is used to calculate the mean, variance, and covariance of each window, and then the structural similarity SSIM of the corresponding block is calculated. Finally, the average value is used as the structural similarity measure of the two images; that is, the average structural similarity MSSIM:

$$\mathrm{MSSIM}(X,Y) = \frac{1}{\mathrm{N}}\sum_{\mathrm{k}=1}^{\mathrm{N}} \mathrm{SSIM}(X_k, Y_k) \tag{18}$$

### *4.4. Experimental Results and Analysis*

This paper presents the super-resolution capability of the model in two aspects: objective evaluation results and subjective evaluation results. In addition to the longitudinal

comparison between the algorithm model integrated into the channel attention mechanism and the USRNet model, the comparison experiments are carried out on the SRMD [37], PAN [38], SRCNN [17], RCAN [39] algorithm models in this chapter, in order to fully verify the effectiveness of the super-resolution reconstruction of the image integrated into the attention mechanism.

#### 4.4.1. Comparison of Model Algorithms

Through the trained and improved AttentionUSRNet model, the reconstruction renderings after 10,000 iterations of training are obtained, as shown in Figure 9. From left to right, we show the original low-resolution images and the reconstructed high-resolution images of the improved USRNet model. In this paper, PSNR and SSIM are selected as quantitative measurement indicators of image reconstruction effect.As can be seen from Table 1 The PSNR and SSIM of the high-resolution images reconstructed by the improved USRNet model on the Set12 dataset are 29.72 dB and 0.8488, respectively. It is concluded that integrating the channel attention mechanism makes the neural network deeper, and it can improve the performance of the model and get better results.

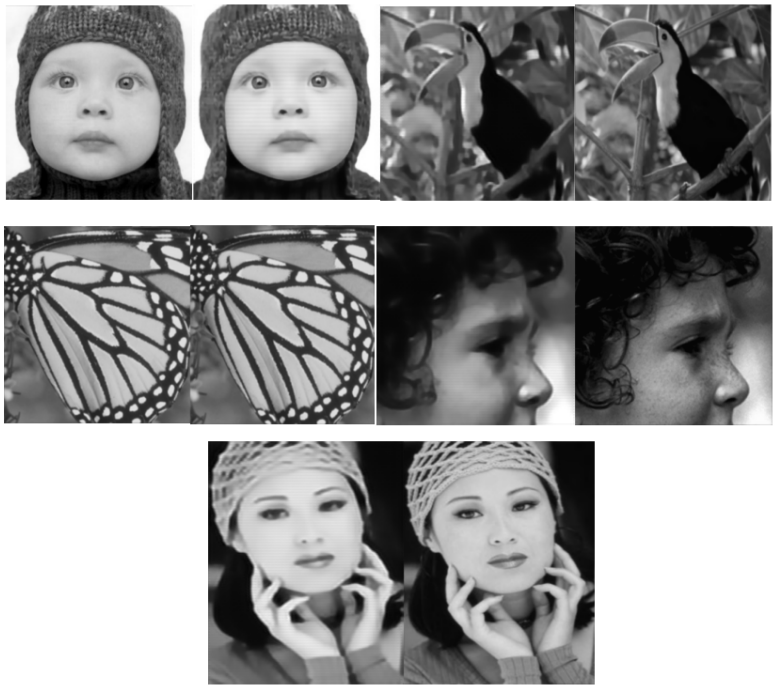


**Figure 9.** Comparison of original low-resolution images and high-resolution images reconstructed by the improved USRNet model.

**Table 1.** Comparative test index analysis of different data sets.

| Data Set | USRNet PSNR SSIM | AttentionUSRNet PSNR SSIM |
|---|---|---|
| Set5 | 28.20/0.8117 | 32.23/0.8832 |
| Set12 | 25.93/0.7574 | 29.72/0.8488 |
| real_faces | 30.79/0.8701 | 32.53/0.8924 |

#### 4.4.2. Objective Evaluation Results

This section will compare the differences between each algorithm model and the algorithm in this paper on the dataset Set12 of SRMD [36], PAN [37], SRCNN [38], and RCAN [17]. The Set12 dataset is a low-complexity single-image dataset for super-resolution research based on non-negative domain embeddings. That is, based on low-resolution

images, they are reconstructed through deep learning algorithms to obtain high-resolution images. This dataset was released by Bell Labs in France in 2012 and is widely used in super-resolution research. The Table 2 shows comparison test index analysis for different algorithms models.

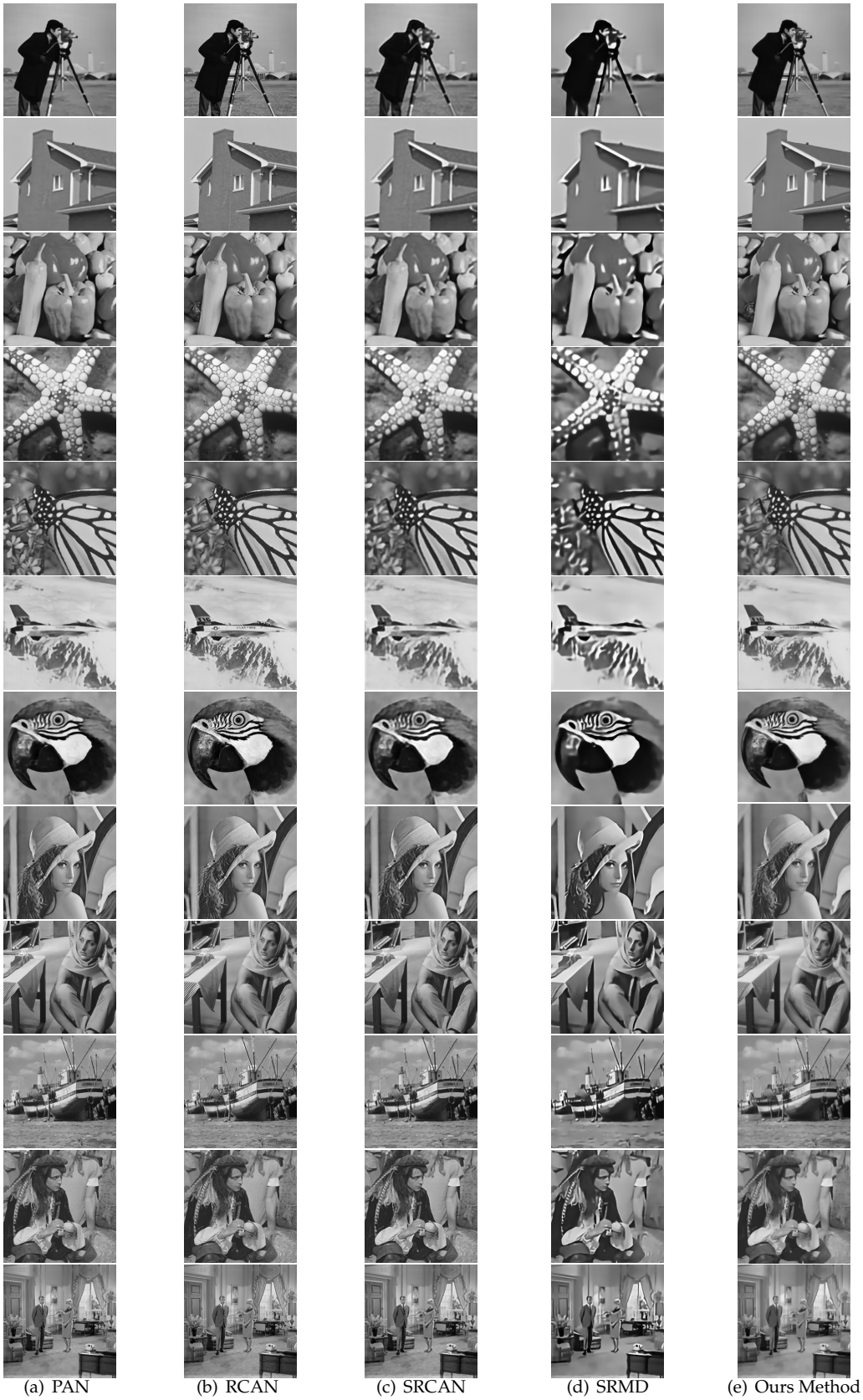
As can be seen from Table 2, the proposed method has achieved good results on both PSNR and SSIM indicators, and the data results show that the PSNR and SSIM values of the proposed method are the largest, which means that the reconstruction result is the best. From the comparison of PSNR values, it is obvious that the proposed method has an improvement of more than 1 dB on different test sets compared with the SRMD method, the SRCNN method and the PAN method. Compared with the PAN method, the proposed method improves the PSNR result by two times on the Set12 test set by 2.6 dB, and the SSIM result with two times magnification is improved by 0.03dB, which indicates that the proposed model has better reconstruction results on different datasets. At the same time, the results of the proposed method are also slightly improved compared with the RCAN method, which indicates that the integration channel attention mechanism USRNet has a good improvement in network performance.

**Table 2.** Comparison of test indicators for different algorithm models.

| Methods | SRMD PSNR SSIM | PAN PSNR SSIM | SRCNN PSNR SSIM | RCAN PSNR SSIM | AttentionUSRNet PSNR SSIM |
|---|---|---|---|---|---|
| 1 | 20.78/0.6880 | 24.73/0.8012 | 23.31/0.7659 | 27.88/0.8904 | 27.38/0.6728 |
| 2 | 23.51/0.7482 | 30.43/0.8473 | 29.52/0.8364 | 33.22/0.9010 | 32.73/0.7287 |
| 3 | 20.92/0.7008 | 27.55/0.8698 | 26.43/0.8405 | 26.17/0.8605 | 31.19/0.7707 |
| 4 | 19.57/0.6301 | 26.86/0.8528 | 24.84/0.7826 | 31.10/0.9282 | 29.88/0.7085 |
| 5 | 19.71/0.7163 | 26.89/0.8951 | 24.41/0.8321 | 31.25/0.9532 | 30.48/0.7812 |
| 6 | 20.07/0.7007 | 25.53/0.8262 | 24.14/0.7872 | 27.22/0.8880 | 27.82/0.6177 |
| 7 | 19.63/0.6643 | 24.41/0.8167 | 22.53/0.7811 | 28.32/0.9047 | 27.12/0.7038 |
| 8 | 23.54/0.7288 | 30.28/0.8610 | 29.42/0.8461 | 33.06/0.9154 | 33.30/0.7502 |
| 9 | 20.79/0.5795 | 23.55/0.6812 | 23.60/0.6819 | 24.39/0.7660 | 24.62/0.6238 |
| 10 | 21.95/0.5981 | 27.26/0.7753 | 26.46/0.7467 | 30.29/0.8701 | 30.00/0.6401 |
| 11 | 22.84/0.6328 | 28.88/0.8167 | 27.75/0.7839 | 32.05/0.9091 | 30.61/0.6902 |
| 12 | 21.81/0.5686 | 26.50/0.7350 | 26.16/0.7269 | 30.41/0.8800 | 29.63/0.6289 |
| Menu | 21.26/0.6630 | 26.91/0.8149 | 25.71/0.7843 | 29.61/0.8889 | 29.72/0.8488 |

#### 4.4.3. Subjective Evaluation Results

In the experiment, SRMD, PAN, SRCNN, RCAN and the improved model of this paper are selected for the comparison of the reconstructed visual effects. As can be seen from Figure 10, among the 12 grayscale images, the details recovered in this paper on the character scene images are also better than the first four models, so this paper has certain advantages in the recovery of high-frequency feature details, and the recovered images are closer to the original pictures. As can be seen from Figure 11, although the PSNR value and SSIM value of the picture reconstructed by RCAN have good performance, the lock on the chest of the girl image is blurred, and the visual effect is not as good as the image reconstructed in this article. Compared with the pictures obtained in this model, RCAN pictures are smoother, and the visual effect is not as good as the details recovered by this model. As can be seen from Figure 12, in the character scene image, the details recovered in this paper on the character scene image are also better than the first four models, so it can be concluded that this model has a certain advantage in the recovery of high-frequency feature details, and the recovered image is more similar to the original picture.



(a) PAN (b) RCAN (c) SRCAN (d) SRMD (e) Ours Method

**Figure 10.** Comparison of reconstruction effects of various algorithms.

In this experiment, a convolutional neural network structure fused with channel attention mechanism is designed to perform super-resolution reconstruction of a single

image, and the core idea is to strengthen the use of feature maps and strengthen the robustness of the network to image size. Through the learning method, the channel attention mechanism ensures the feature map can no longer be treated as equal to the traditional network, and the computer computing power is allocated more efficiently. The experimental results show that the reconstruction method in this paper has good effect and outstanding performance in various indicators and subjective vision.


**Figure 11.** Comparison of girl's lock.

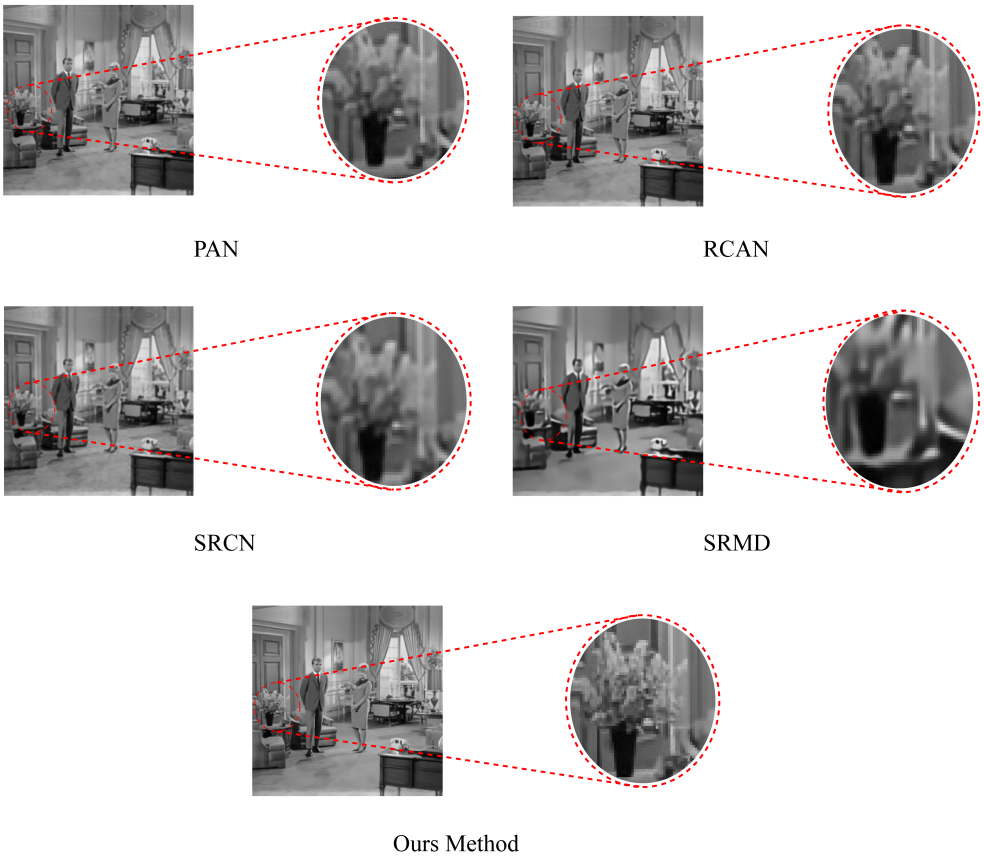



**Figure 12.** Comparison of girl's lock.

## 5. Conclusions

In this paper, the image super-resolution algorithm integrated into the channel attention mechanism is optimized based on the super-resolution network of USRNet, which is prone to problems such as artifacts in high-frequency details, ignoring the reconstructed image texture structure and easy overfitting of network training. At the same time, because the deep learning convolutional neural network is too deep, there are problems such as a long training time and low super-resolution efficiency. Furthermore, the unnecessary batch normalization operation convolutional layers are removed from the generative model. On

the discriminant model, the original discriminator of the USRNet network is used to guide the training of the super-resolution model. The L1 loss function is used on the loss function to further improve the visual effect of image super-resolution. On the Set5, Set12 and real faces datasets, the PSNR has been improved by 0.72 dB, 0.60 dB and 1.42 dB respectively, which is a certain improvement compared with the original model reconstruction results of USRNet. A series of experiments have verified the effectiveness of the method proposed in this section.The algorithm in this paper has some room for improvement in reconstructing the effect on some images with extremely complex textures. In addition, with the development of deep learning, the operation speed of deep networks also has great potential, and future research on SR can pay more attention to texture extraction and design new network frameworks to improve reconstruction efficiency.

**Author Contributions:** Writing—original draft preparation, L.C., P.M. and Z.M.; writing—review and editing, L.C., S.Z., G.D. and H.W. All authors have read and agreed to the published version of the manuscript.

**Funding:** This work is partially supported by the Young Talent Project of Science and Technology Plan of Hubei Education Departmen (Q20191514), Scientific Research Fund Project of Wuhan Institute of Technology (16QD25, 20QD32) , and the Graduate Education Innovation Fund of Wuhan Institute of Technology (CX2021271).

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Informed consent was obtained from all subjects involved in the study.

**Data Availability Statement:** Not applicable.

**Acknowledgments:** The authors also thank everyone involved in the project.

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

## References

1. Niu, X. An overview of image super-resolution reconstruction algorithm. In Proceedings of the 2018 11th International Symposium on Computational Intelligence and Design (ISCID), Hangzhou, China, 8–9 December 2018; Volume 2, pp. 16–18.
2. Timofte, R.; De, V.; Gool, L.V. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. In Proceedings of the 2013 IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 1920–1927.
3. Timofte, R.; De Smet, V.; Van Gool, L. A+: Adjusted anchored neighborhood regression for fast super-resolution. In Proceedings of the Asian Conference on Computer Vision, Singapore, 1–5 November 2014; pp. 111–126.
4. Huang, J.B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the IEEE conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 5197–5206.
5. Brown, L.G. A survey of image registration techniques. *ACM Comput. Surv. (CSUR)* **1992**, *24*, 325–376. [CrossRef]
6. Yang, S.; Kim, Y.; Jeong, J. Fine edge-preserving technique for display devices. *IEEE Trans. Consum. Electron.* **2008**, *54*, 1761–1769. [CrossRef]
7. Duchon, C.E. Lanczos filtering in one and two dimensions. *J. Appl. Meteorol. Climatol.* **1979**, *18*, 1016–1022. [CrossRef]
8. 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.
9. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
10. Kim, J.; Lee, J.K.; Lee, K.M. Deeply-recursive convolutional network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1637–1645.
11. Li, J.; Fang, F.; Mei, K.; Zhang, G. Multi-scale residual network for image super-resolution. In Proceedings of the European conference on computer vision (ECCV), Munich, Germany, 17–24 May 2018; pp. 517–532.
12. Tian, C.; Xu, Y.; Zuo, W.; Zhang, B.; Fei, L.; Lin, C.W. Coarse-to-fine CNN for image super-resolution. *IEEE Trans. Multimed.* **2020**, *23*, 1489–1502. [CrossRef]
13. Qiao, C.; Li, D.; Guo, Y.; Liu, C.; Jiang, T.; Dai, Q.; Li, D. Evaluation and development of deep neural networks for image super-resolution in optical microscopy. *Nat. Methods* **2021**, *18*, 194–202. [CrossRef] [PubMed]
14. Wang, X.; Girshick, R.; Gupta, A.; He, K. Non-local neural networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7794–7803.
15. Hu, J.; Shen, L.; Sun, G. Squeeze-and-excitation networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 7132–7141.

16. Dai, T.; Cai, J.; Zhang, Y.; Xia, S.T.; Zhang, L. Second-order attention network for single image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 11065–11074.
17. 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.
18. Zhao, H.; Kong, X.; He, J.; Qiao, Y.; Dong, C. Efficient image super-resolution using pixel attention. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 56–72.
19. 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, Honolulu, HI, USA, 21–26 July 2017; pp. 4681–4690.
20. Gao, H.; Xiao, J.; Yin, Y.; Liu, T.; Shi, J. A Mutually Supervised Graph Attention Network for Few-Shot Segmentation: The Perspective of Fully Utilizing Limited Samples. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, 1–13. [CrossRef] [PubMed]
21. Gao, H.; Xu, K.; Cao, M.; Xiao, J.; Xu, Q.; Yin, Y. The Deep Features and Attention Mechanism-Based Method to Dish Healthcare Under Social IoT Systems: An Empirical Study With a Hand-Deep Local–Global Net. *IEEE Trans. Comput. Soc. Syst.* **2022**, *9*, 336–347. [CrossRef]
22. Efrat, N.; Glasner, D.; Apartsin, A.; Nadler, B.; Levin, A. Accurate blur models vs. image priors in single image super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Sydney, Australia, 1–8 December 2013; pp. 2832–2839.
23. 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.
24. Caselles, V.; Morel, J.M.; Sbert, C. An axiomatic approach to image interpolation. *IEEE Trans. Image Process.* **1998**, *7*, 376–386. [CrossRef] [PubMed]
25. 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 IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 624–632.
26. Lim, B.; Son, S.; Kim, H.; Nah, S.; Mu Lee, K. Enhanced deep residual networks for single image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops, Honolulu, HI, USA, 21–26 July 2017; pp. 136–144.
27. Hu, X.; Mu, H.; Zhang, X.; Wang, Z.; Tan, T.; Sun, J. Meta-SR: A magnification-arbitrary network for super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 15–20 June 2019; pp. 1575–1584.
28. Chen, Y.; Tai, Y.; Liu, X.; Shen, C.; Yang, J. FSRNet: End-to-end learning face super-resolution with facial priors. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 2492–2501.
29. Kim, J.; Lee, J.K.; Lee, K.M. Accurate image super-resolution using very deep convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1646–1654.
30. Ren, D.; Zhang, K.; Wang, Q.; Hu, Q.; Zuo, W. Neural blind deconvolution using deep priors. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3341–3350.
31. Zhang, K.; Gool, L.V.; Timofte, R. Deep unfolding network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 13–19 June 2020; pp. 3217–3226.
32. Yasarla, R.; Perazzi, F.; Patel, V.M. Deblurring face images using uncertainty guided multi-stream semantic networks. *IEEE Trans. Image Process.* **2020**, *29*, 6251–6263. [CrossRef] [PubMed]
33. Bluche, T. Joint line segmentation and transcription for end-to-end handwritten paragraph recognition. *Adv. Neural Inf. Process. Syst.* **2016**, *29*, 838–846.
34. Wang, F.; Jiang, M.; Qian, C.; Yang, S.; Li, C.; Zhang, H.; Wang, X.; Tang, X. Residual attention network for image classification. In Proceedings of the IEEE conference on computer vision and pattern recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 3156–3164.
35. Bevilacqua, M.; Roumy, A.; Guillemot, C.; Alberi-Morel, M.L. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. In Proceedings of the 23rd British Machine Vision Conference (BMVC), Surrey, UK, 3–7 September 2012.
36. Zeyde, R.; Elad, M.; Protter, M. On single image scale-up using sparse-representations. In Proceedings of the International Conference on Curves and Surfaces, Avignon, France, 24–30 June 2010; pp. 711–730.
37. Shi, W.; Caballero, J.; Ledig, C.; Zhuang, X.; Bai, W.; Bhatia, K.; Marvao, A.M.; Dawes, T.; O'Regan, D.; Rueckert, D. Cardiac image super-resolution with global correspondence using multi-atlas patchmatch. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Nagoya, Japan, 22–26 September 2013; pp. 9–16.
38. 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]
39. Srivastava, R.K.; Greff, K.; Schmidhuber, J. Training very deep networks. In Proceedings of the 28th Annual Conference on Neural Information Processing Systems (NIPS 2015), Montreal, QC, Canada, 11–12 December 2015.