Mathematical Biosciences and Engineering

http://www.aimspress.com/journal/MBE

MBE, 18(5): 6652–6671.
DOI: 10.3934/mbe.2021330
Received: 23 June 2021
Accepted: 27 July 2021
Published: 06 August 2021

Research article

Image super-resolution reconstruction for secure data transmission in Internet of Things environment

Hongan Li[1], Qiaoxue Zheng[1], Wenjing Yan[2,*], Ruolin Tao[1,*], Xin Qi[3] and Zheng Wen[4]

[1] College of Computer Science and Technology, Xi'an University of Science and Technology, Xi'an 710054, China

[2] Department of Information Management, School of E-business and Logistics, Beijing Technology and Business University, Beijing 100048, China

[3] Global Information and Telecommunication Institute, Waseda University, Shinjuku, Tokyo 169-8050, Japan

[4] School of Fundamental Science and Engineering, Waseda University, Tokyo 169-8050, Japan

* **Correspondence:** Email: yanwenjing@btbu.edu.cn, ruolintao@stu.xust.edu.cn.

Abstract: The image super-resolution reconstruction method can improve the image quality in the Internet of Things (IoT). It improves the data transmission efficiency, and is of great significance to data transmission encryption. Aiming at the problem of low image quality in image super-resolution using neural networks, a self-attention-based image reconstruction method is proposed for secure data transmission in IoT environment. The network model is improved, and the residual network structure and sub-pixel convolution are used to extract the feature of the image. The self-attention module is used extract detailed information in the image. Using generative confrontation method and image feature perception method to improve the image reconstruction effect. The experimental results on the public data set show that the improved network model improves the quality of the reconstructed image and can effectively restore the details of the image.

Keywords: image super-resolution; self-attention; generative adversarial networks; data encryption; Internet of Things

1. Introduction

With the improvement of people's living standard and quality, the use of smart devices in the Internet of Things [1–4] has become more common, and it has a very broad range of application scenarios in smart grids, telemedicine, and smart transportation [5, 6, 37]. The intelligent service of the IoT needs data as the basis. How to effectively obtain different types of data resources and analyze them becomes

particularly important. In addition, data transmission and data encryption among IoT devices have also become a major challenge for IoT security [7–9, 11]. Image as an important carrier of IoT data, can more clearly and intuitively reflect the information content. Image super-resolution reconstruction can turn a low-resolution image into a high-resolution image [12], which improves the transmission efficiency of data in the IoT and reduces the huge overhead of data encryption and key exchange during data transmission [10, 13, 22, 23].

In recent years, the rapid development of global deep learning has made the research on image super-resolution reconstruction methods become the focus of the public, which promotes the development of image super-resolution reconstruction [58, 59]. In the research dealing with computer vision, many tasks can not only be solved by deep learning, but the effects of deep learning-based methods are also better than traditional methods [35]. The neural network is applied to image super-resolution reconstruction, and the mapping relationship between input low-quality images and output high-quality images is established through learning, so as to improve the image quality. This method has great potential value in the application of Internet of things, such as medical images, high-resolution images can help more accurate judgment of the disease, and can also improve the user's visual experience in daily life.

Low resolution images limit the accuracy of information transmission to some extent, thus reducing the efficiency of information transmission [14, 38]. Image super-resolution reconstruction technology can improve the resolution of the image [15], thereby obtaining richer image details. The existing image super-resolution reconstruction methods are roughly divided into three categories : interpolation-based methods, reconstruction-based methods and learning-based methods [16].

Image super-resolution method based on interpolation. Interpolation-based method is mainly based on the image interpolation technology, also known as image scaling, which is to infer the value of unknown pixels by using the known pixels through the interpolation function or interpolation kernel. In this method, each known pixel on the image is regarded as a point on the image plane, and the value of unknown pixels to be interpolated is determined by using each known data for interpolation calculation. Finally, the digital image restoration technology is used to remove blur and reduce image noise. Commonly used interpolation methods mainly include nearest neighbor interpolation [19, 20, 25], bilinear interpolation [27] and bicubic interpolation [28, 51].

Although the method based on interpolation is simple and fast, there are also some obvious technical defects. The processing effect of image edge (i.e. pixel mutation) is poor, and it is prone to sawtooth and block effect, which can not reach the ideal super-resolution.

Image super-resolution method based on reconstruction. The imaging process of the image is inverted through the reconstruction method, and the local or global prior model of the image is introduced to establish the observation model between the LR image and the HR image. Finally, the optimization model of data fidelity term and image prior regularization is established to solve it. Tsai [29] proposed a super-resolution reconstruction method for recovering additional high frequency information from multiple images in the Fourier transform domain for the first time. However, this method assumes that there is no motion blur and observation noise in the image, and ignores the point spread function of the optical system, so it is only suitable for the ideal image degradation model. The improved method uses recursive least squares, spatial blur, relative motion, discrete DCT transform and wavelet transform to remove the image observation noise, which effectively improves the visual effect of the reconstructed image [30, 31].

The spatial domain method model the spatial factors (such as optical blur, motion blur, etc.) that affect the image imaging effect, so it is easier to approach the actual application. Common spatial super-resolution reconstruction methods mainly include non-uniform sampling interpolation (NUI), maximum a posteriori probability (MAP), iterative back projection, convex set projection [12,32,34]. The method based on reconfiguration has relatively good effect on image processing, but requiring the image to have better prior knowledge, so it is not fully applicable to directly reconstruct the images of the IoT with large magnification.

Image super-resolution method based on learning. Since machine learning has good application effect in the field of computer vision, the learning-based image super-resolution reconstruction method has obtained fruitful research results. Dong et al. [33] applied the convolutional neural network to super-resolution reconstruction of a single image, and its reconstruction effect was far superior to traditional methods. Kim et al. [36] proposed a VDSR method, which effectively improves reconstruction effect by expanding the network depth. Talab et al. [42] proposed ESPCN, which enlarges a low-resolution image into a high-resolution image through an interpolation method, and then uses a convolutional network to get a high-resolution image to make the texture of the generated image clearer. For making the reconstructed image meet the signal-to-noise ratio requirements, Ledig et al. [43] proposed an SRGAN method that uses perceptual loss to enhance the effect of network training, thereby recovering rich high-frequency detail information. Lim et al. [45] improved the EDSR method from the perspective of data processing, replacing the batch normalization in the residual module, making the network more versatile [39]. A single-frame character image super-resolution reconstruction method based on wavelet neural network is proposed, which improves the anti-interference performance of the model [48].

Although the image super-resolution method based on deep learning can improve the quality of restored images, it can not meet the quality requirements of data transmission in the Internet of Things. For actual IoT usage scenarios, domain knowledge, prior knowledge, and deep learning frameworks are not fully utilized, and they cannot complement each other in practical applications, so it can not pay more attention to the useful domain prior knowledge in image feature extraction, which leads to the poor quality of high-resolution image. Therefore, this paper proposes an image super-resolution method with self-attention. The main contributions of the work presented in this paper are as follows:

1) Introducing the self-attention into an image super-resolution model, which improves the model's attention to detail in IoT images.

2) Improving network model, extracting feature information from IoT images using residual network structure and subpixel convolution.

The rest of this paper is organized as follows. The section 2 briefly introduces the related work of neural networks and loss functions. Network structure and problem formulation is discussed in section 3. The section 4 displays and analyzes the experimental results. The paper is concluded in section 5.

2. Related work

2.1. Generative adversarial networks

For the high magnification factor, the image super-resolution reconstruction method is ill-posed [53]. The reconstructed super-resolution (SR) images have blurred texture details, and the

reconstruction effect is not ideal. Generative Adversarial Networks (GAN) [50] has strong image generation capabilities, and well solve the above ill-posed problems. Therefore, the image super-resolution method based on GAN is proposed [51]. This method learns the mapping relationship of the network through large-scale IoT image training [15], and the GAN structure is shown in Figure 1.

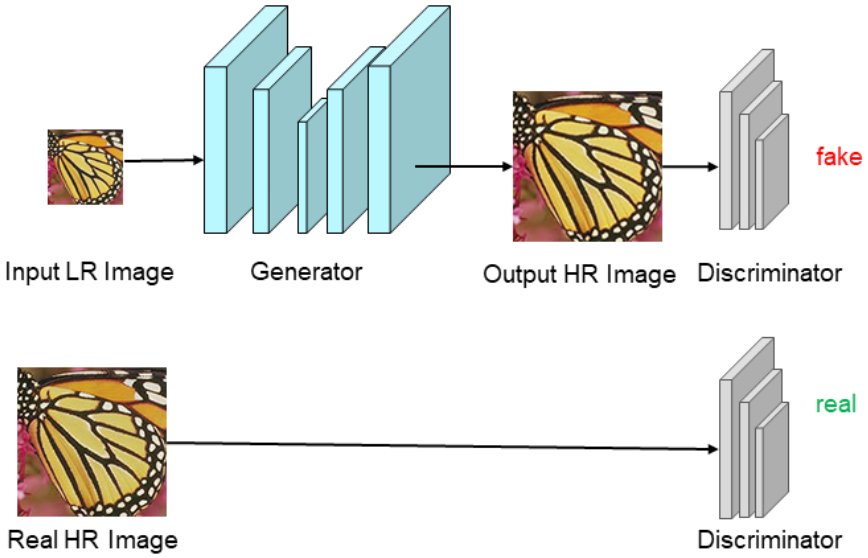


Figure 1. Structure of GAN.

GAN is composed of generator network G and discriminant network D. G is responsible for generating the corresponding high-resolution image by feeding the low-resolution image x. D is responsible for identifying its input image, judging from real data as true, judging from generated data from generator G as false. Therefore, the image super-resolution model based on GAN can be expressed as:

$$L(G, D) = E_{x,y}\left[\ln D(x, y)\right] + E_{x,y}\left[\ln(1 - D(y, G(x)))\right]. \tag{1}$$

Here, x represents the input low-resolution image, y represents the real high-resolution image. By minimizing the generator network loss and maximizing the discriminator loss, the overall network model is continuously trained through the game process between the G and the D. When GAN is successfully trained and converged, the Nash equilibrium between the generator and the discriminator is achieved, and the ability of the model to generate high-resolution images is improved.

Ledig et al. [43] introduced GAN into the field of image super-resolution and proposed SRGAN. Ledig combined the deep residual network with the sub-pixel convolution model, and used the residual network to extract the image feature and eliminate its noise. Up-sampling the extracted image features using the sub-pixel convolution model to gradually restore the detailed information in the image [48]. SRGAN makes the whole network pay more attention to the semantic feature difference between the generated super-resolution image and the real image, rather than the color brightness difference between pixels. However, this method performs super-resolution restoration of the image from a global perspective, ignoring the details of the image. Therefore, we proposed an improved method based on its work, adding self-attention to produce high-quality super-resolution image.

2.2. Subpixel convolution

Subpixel convolution is a super-resolution method that uses image features for image up sampling [41, 49, 68]. It is composed of a set of expansion filters, which continuously adjust the parameters in the filter through learning. By establishing an image mapping relationship, the low-resolution input image is filtered to output an enlarged high-resolution image. For each feature image, a more complex amplification filter is trained and a sub-pixel convolutional layer is formed. Using sub-pixel convolutional layers to replace commonly used image scale expansion methods (such as bilinear interpolation and deconvolution operations, etc.) can improve the accuracy of image pixel amplification while reducing the computational complexity in reconstruction operations. Compared with the traditional methods, the parameters of the image magnification filter in subpixel convolution can be changed adaptively through training and learning. Therefore, using this method to enlarge the image can greatly reduce the risk and get a high-quality image with clearer structure and texture. The sub-pixel convolution process is shown in the Figure 2.

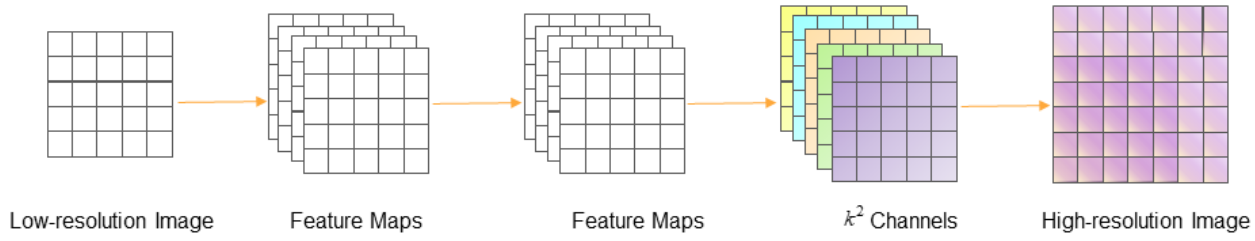


Figure 2. The structure of subpixel convolution.

Set the zoom factor of the image to k. Subpixel convolutional layer receives low-resolution feature images, and after convolution of two hidden layers, the feature image with the same size as the input image is obtained. The number of feature channels is k^2. Then rearrange all pixels in the channel features. By combining the single pixels on the multi-channel feature map into the pixel values on a feature map, the image is rearranged into one image, that is, the high-resolution image is a linear combination of k^2 feature pixels.

2.3. Loss function

The loss function is also known as the objective function, it's often used to evaluate the inconsistency between the actual output and the expected output of the neural network [17, 18]. In the process of training, the neural network optimizes the model by minimizing the loss function, narrowing the gap between the predicted image and the real image and improving the ability of model generation.

2.3.1. Perceptual loss function

When optimizing the parameters of the neural network, loss functions such as absolute value loss and mean square error loss are usually used to calculate the target difference value. However, these loss functions are difficult to deal with the uncertainty of high-frequency details such as textures, which leads to over-smooth reconstruction images and cannot produce clear image boundaries.

Perception loss can be judged from the perspective of overall image information to increase the global perception of the model. Mualfah et al. [54] use gray level co-occurrence matrix to extract image feature information, and encrypt the asymmetric key through a single layer perceptron. Johnson et al. [55] used the perceptual loss function to train the feedforward network of image conversion task to generate high-quality images. Leading et al. [43] added the perceptual loss function in the production resolution model, making the entire network pay more attention to differences in high-level semantic information. The generated image and real image are input into the pre-training network model, and the perceptual loss of the two images is obtained by extracting the image feature information and calculating [44]:

$$L_p = \|\varphi(\hat{y}) - \varphi(y)\|^2 . \tag{2}$$

Here, ϕ represents the pre-training network, $\hat{y}$ and y epresent the generated high-resolution and real high-resolution images, respectively. By calculating the perceptual loss between images, the expression ability of network model on semantic information can be improved.

2.3.2. Content loss

The image super-resolution model maps the image according to the feature information provided, and the restoration result of the model depends on the texture detail information of the image [21]. The content of the image includes brightness information, color information, structure information and texture information. By comparing the content information of the generated and real high-resolution images, the network can be better optimized and the quality of image generation can be improved. Cheon et al. [57] used discrete cosine transform coefficient loss and differential content loss to solve the trade-off problem. Lee et al. [60] used the loss of new content generated by the network to overcome the loss of detailed information caused by the pooling operation. By calculating the mean square error per pixel of fake image $G(x)$ and real image y, the difference between the two images is compared. The content loss formula is as follows:

$$L_c = \frac{1}{WH} \sum_{i=1}^{W} \sum_{j=1}^{H} (y_{i,j} - G(x)_{i,j})^2. \tag{3}$$

Here, W and H respectively represent the width and height of images, (i, j) represents a pixel in images [24], x represents the input low-resolution image, y represents the ground truth, and G represents the generator network.

3. Image super-resolution based on self-attention

3.1. Self-attention

Image super-resolution reconstruction can be seen as a process of restoring high-frequency information of low-resolution images, that's why so much low-frequency information can be transmitted to the final output high-resolution image directly [40, 46, 47]. In the process of feature extraction, low-frequency information and high-frequency information are not distinguished while learning, which results in the network can not make full use of the high-frequency information in LR images. By adding the self-attention, he network model can focus more on the important information in the image and ignore the irrelevant information, which can improve these problems effectively.

Attention in neural networks has the same function as the way the human eye observes objects, and information with different characteristics is treated in the form of a network module differently. The self-attention structure is visualized as shown in Figure 3. The self-attention consists of three parts: K, Q, V [44]. Q represents the input query of the model, K represents the reference with high similarity to Q, and V represents the output content information corresponding to K. According to different processing tasks, the information stored in each part is different. In image processing tasks, the inherent information of image features is used for attention interaction, and the original feature map is usually mapped to Q, K, V. We calculate the correlation weight coefficient between Q and K and normalizing, and finally superimposed the coefficient into V.

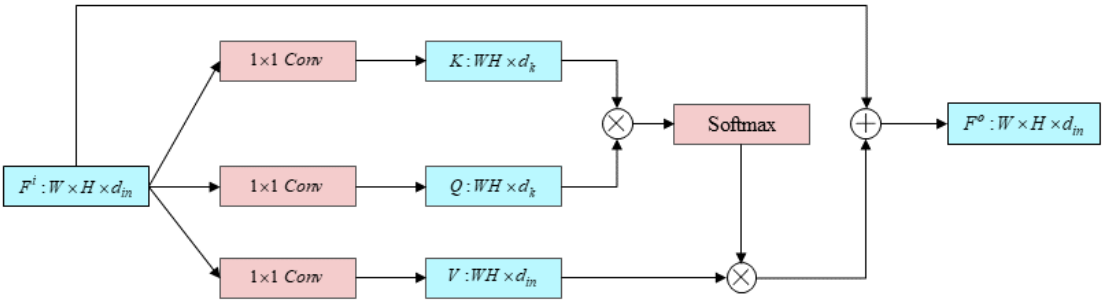


Figure 3. Structure of self-attention.

Most of the research work on the combination of deep learning [65–67] and visual attention focuses on the use of masks to form the attention [61]. However, this method cannot dynamically adjust the network's attention to feature images, so we use the self-attention to filter image feature information, and add a larger attention coefficient to the information of interest. The self-attention is also called the internal attention. It is mainly an attention that calculates the correlation between different positions in a single sequence. Vaswani et al. [62] use the full attention on machine translation tasks, it proves that the network structure based on the attention can run in parallel and requires less training time. The self-attention mechanism is used to pay attention to the different information in the image, and the feature map is convolved to obtain Q, K, V. The self-attention formula is as follows:

$$Attention(Q, K, V) = soft\max(\frac{QK^T}{\sqrt{d_k}})V. \tag{4}$$

Here, d_k represents the output dimension of the feature image.Multiplying the self-attention feature map with the dynamic adjustment coefficient and superimposing it on the original feature input can realize the amplification of the detailed information of interest and suppress the irrelevant noise information in the image.

3.2. Network structure

The self-attention-based image super-resolution network model proposed in this paper includes two parts: a generator network and a discriminator network. The generator network consists of 16 deep residual modules, 3 convolution modules, and 2 sub-pixel convolutions, and 2 self-attention modules. The specific network structure is shown in Figure 4.

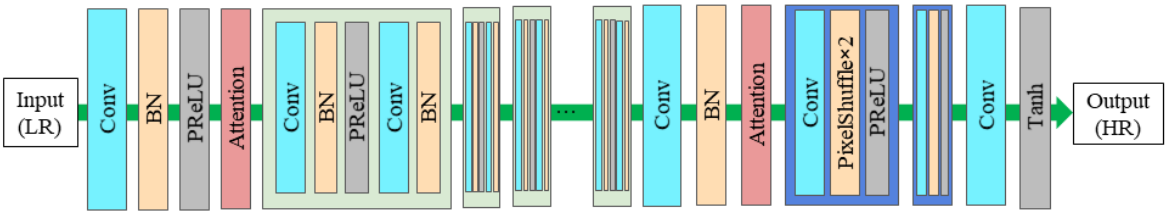


Figure 4. Network structure of our method.

The low-resolution image is input into the network, the features of the image are extracted through the convolution operation, and the high-resolution image is finally output. In order to achieve the purpose of enabling the generator to capture the content information of the image greatly, this paper uses 16 depth residual modules to extract the feature information. Each residual block consists of 2 convolutional layers, 2 batch normalization layers and 1 activation layer. The activation layer uses the PReLU to perform the activation operation, so that during the network training process, the slope of the negative part can be dynamically adjusted according to the input image data, so as to improve the generalization and avoid over-fitting of the model. In order to enable the network to extract detailed information in the image, this paper adds self-attention after the first and second convolution blocks. Self attention is used to adjust the information of interest in the feature map to improve the attention of texture details and image content, and reduce the interference of noise data and irrelevant information. The feature map after self-attention calculation is input into the sub-pixel convolution layer, and the resolution of the image is enlarged by 4 times by using two up-sampling convolutions, and finally the high-resolution image is obtained through the convolution operation. All parameters of the network are described in detail in Table 1.

Table 1. Generator network parameters.

Layer name	Output size	Parameter	Numbers
Conv_1	24 × 24	[9 × 9, 64]	× 1
SA_1	24 × 24	[1 × 1, 8] [1×1, 8] [1×1, 64]	× 1
ResConv	24 × 24	[3×3, 64] [3×3, 64]	× 16
Conv_2	24 × 24	[3 × 3, 64]	× 1
SA_2	24 × 24	[1 × 1, 8] [1 × 1, 8] [1 × 1, 64]	× 1
SubpixelConv	96 × 96	[3 × 3, 256]	× 2
Conv_3	96 × 96	[9 × 9, 3]	× 1

In order to improve the ability of the network to recognize the global features of the image, a larger convolution kernel is used in the input and output parts of the generator to improve the receptive field of vision. At the same time, in order to reduce the number of parameters and prevent the model from being too large, a smaller convolution kernel is used in the model for feature extraction. The discriminator model judges the real high-resolution images and the generated high-resolution images [56]. Its essence is a classifier model, which judges the real images as true and the generated images as false.The discriminator network structure is shown in Figure 5.

The discriminator model is composed of 8 convolutional blocks, which are used to extract the deep

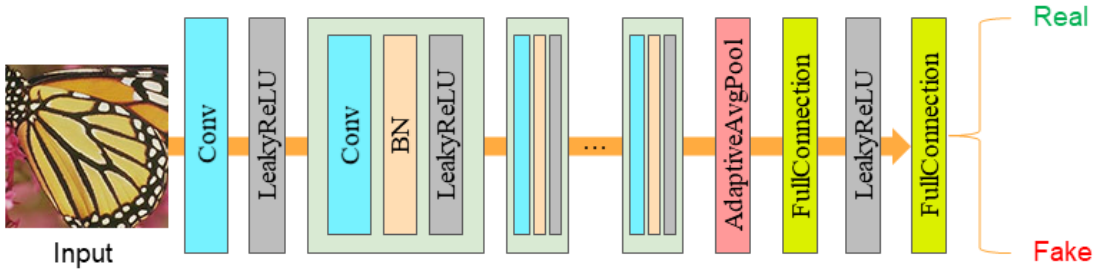


Figure 5. Discriminator structure of our method.

feature information of the image, and judge the generated image and the real image on the high-level semantic information. To adapting to different input image sizes, an adaptive average pooling operation is used to pool the obtained feature images, and finally the classification results are output through a multilayer linear perceptron. The network parameters of the discriminator are described in Table 2.

Table 2. Discriminator network parameters.

Layer name	Output size	Parameter
Conv_1	96×96	$[3 \times 3, 64]$
Conv_2	48×48	$[3 \times 3, 64]$
Conv_3	48×48	$[3 \times 3, 128]$
Conv_4	24×24	$[3 \times 3, 128]$
Conv_5	24×24	$[3 \times 3, 256]$
Conv_6	12×12	$[3 \times 3, 256]$
Conv_7	12×12	$[3 \times 3, 512]$
Conv_8	6×6	$[3 \times 3, 512]$
AdaptiveAvgPool	6×6	–
FC_1	1024	–
FC_2	1	–

The image super-resolution reconstruction model designed in this paper can reconstruct the image of any resolution. When the discriminator is used to judge the image, the adaptive pooling operation is used to adapt to different resolutions. The convoluted feature image is pooled to get a uniform output size, and finally the true and false images are distinguished.

3.3. Overall loss

Use loss function to optimize image super-resolution network based on neural network. The quality of the image super-resolution model depends largely on the overall optimization goal. In this paper, the loss function of the self-attention-based image super-resolution reconstruction model is composed of three parts, namely, generative adversarial loss, perceptual loss, and content loss. In order to improve the overall perception ability of the model, this paper uses the pre-trained VGG-19 network to calculate the perception loss of the model. Input the real high-resolution image and generated high-resolution image to the VGG-19 network respectively, obtain the corresponding feature map through feature

extraction, and obtain the perceptual loss by calculating the difference between the feature maps. The specific calculation is as follows:

$$L_p = \frac{1}{C}\sum_{j=1}^{C}\left\|\varphi_j(y) - \varphi_j(G(x))\right\|^2. \tag{5}$$

Here, C represents the number of channels of the feature map, φ_j represents the feature image at the j level, y represents the ground turth image, x represents the low-resolution image, and G represents the generator. The generative adversarial loss is used to improve the generation ability and discrimination ability of generators and discriminators. The content loss function mainly constrains the content information and compares content differences. The overall loss of the model can be expressed as

$$L_{total} = \beta L(G, D) + L_p + L_c. \tag{6}$$

Here, β represents the generative adversarial loss coefficient , and the perceptual loss and content loss are respectively represented as L_p and L_c . The perceptual loss function and content loss function are used to judge the generated image and the real image. Figure 6 describes all the loss function calculation processes.

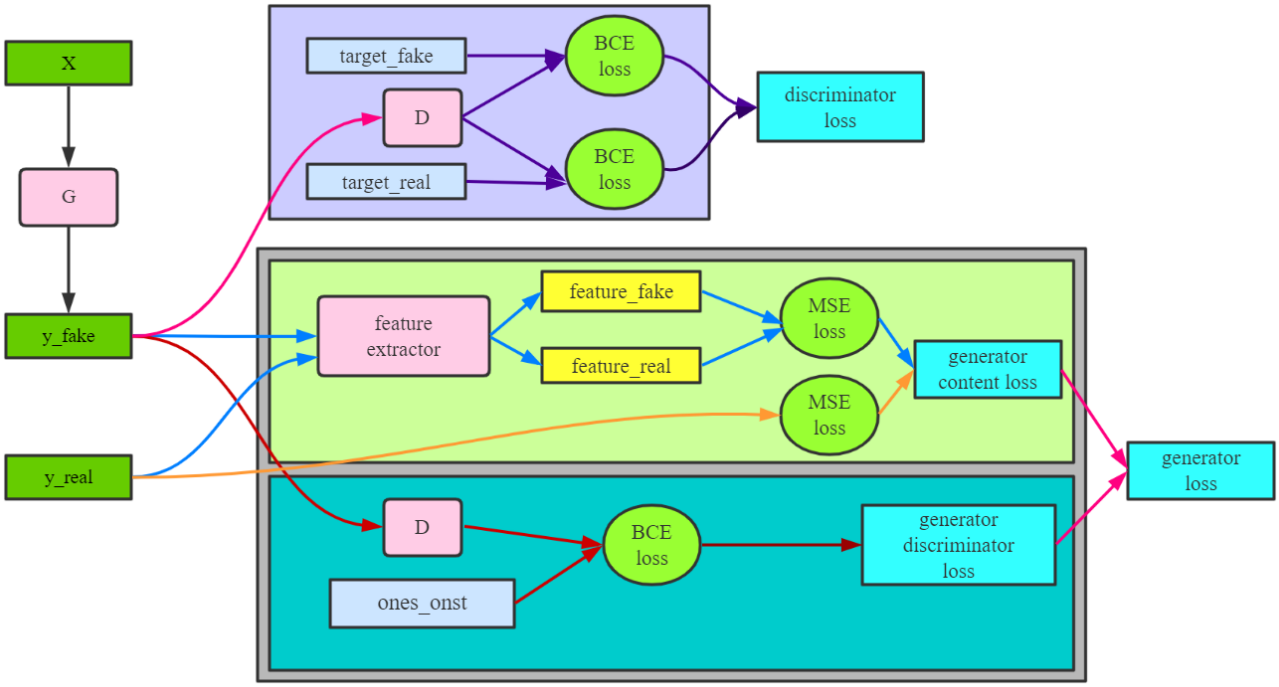


Figure 6. Visualization of loss function calculation process.

3.4. Evaluation metric

In order to evaluate the superiority of the method more comprehensively and accurately, subjective evaluation indicators and objective evaluation indicators are used to evaluate and analyze the reconstruction effect of the self-attention-based image super-resolution model proposed in this paper.

The subjective evaluation mainly evaluates and scores the satisfaction of the reconstructed image from the perspective of visual effects, while the objective evaluation uses quantitative evaluation indicators to compare the similarity between the reconstructed image and the real one.

3.4.1. Mean opinion score (MOS)

Average subjective opinion score usually evaluates the image quality based on the subjective feelings [63]. In this paper, MOS is used as a subjective evaluation metric. The quality of the image is scored by a sufficient number of observers. MOS evaluation metric are shown in Table 3.

Table 3. MOS reference.

Leval	Inpairment	Score
1	Very annoying	1–2
2	Annoying	3–4
3	Slightly annoying	5–6
4	Perceptible,but not annoying	7–8
5	Imperceptible	9–10

MOS evaluation metric shows the human visual evaluation of the image is good or bad, with simple operation, intuitive results. The evaluation method based on the user's subjective intentions has a total of 50 participants, and used a random algorithm to select 4 pairs of images from the Bicubic model, SRGAN model and the high-resolution images generated by the model in this article to experiment. Each participant is shown reconstructed images generated by the three models respectively for the same input image within 5 seconds. All participants are not informed of the image generation model after making their own judgment. However, it is difficult to ensure the objective fairness of data, because of the evaluation of image quality depends entirely on the aesthetic preference.

3.4.2. Objective evaluation

Different from the MOS, the objective evaluation uses a series of method formulas to analyze the image more rigorously. PSNR and SSIM are the most basic indicators for measuring the quality of compressed reconstructed images in image super-resolution tasks [64]. Therefore, we use these two indicators to evaluate the improved method proposed in this article and compare it with other models. The larger the value of PSNR and SSIM, the better the quality of the image, that is, the better the performance of the image super-resolution method [52].

PSNR. PSNR is the ratio of the maximum power of the signal to the noise power of the signal [56]. It is used to measure the quality of the reconstructed image and is usually expressed in decibels (dB). The higher the score obtained by the PSNR indicator, the smaller the pixel difference between the generated image and the real image. Its calculation formula is expressed as follows:

$$MSE = \frac{1}{H \times W} \sum_{i=1}^{H} \sum_{j=1}^{W} [X(i, j) - Y(i, j)]^2. \tag{7}$$

$$PSNR = 10\log_{10} \frac{(2^n - 1)^2}{MSE}. \tag{8}$$

Here, X and Y respectively denote the generated super-resolution images and the real super-resolution images, W and H respectively denotes the width and height of images, (i, j) representing each pixel, which n is the number of bits of the pixel.

SSIM. PSNR also has its limitations, it cannot fully reflect the consistency of the image quality and the human visual effect, so SSIM is used for further comparison. SSIM is structural similarity, which is an evaluation index for evaluating the similarity of images in structure. Its calculation formula is as follows:

$$SSIM = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)}. \tag{9}$$

Here, μ_x and μ_y respectively represents the average value of the real image and the generated image, σ_x^2 and σ_y^2 respectively represents the variance of the real image and the generated image, σ_{xy} represents the covariance of the real image and the generated image. $c_1 = (k_1, L)^2$, $c_2 = (k_2, L)^2$ are constants to maintain stability and L is the dynamic range of pixel values. $k_1 = 0.01$, $k_2 = 0.03$.

4. Experiments and discussion

4.1. Experimental data and environment

All the experiments in this paper are done on the same lab environment, and the network model proposed in this article is implemented using Python 3.6 and Pytorch 1.4.0 software development environment. It runs on a 64-bit Windows 10 operating system, the processor is Intel(R) Core(TM) i9-10900X CPU @ 3.70GHz 3.70 GHz, and the graphics card is NVIDIA GeForce RTX 2080 Ti and CUDA 10.2.

The experiment uses COCO2014 data to train the network model proposed in this paper. In order to restore the complex image environment in the data transmission of Internet of things, SET5, SET14 and BSD1003 public data sets are used for model testing. From the BSD100 dataset, different kinds of images such as airplane, vase, racing car, human and animal are selected for testing; Different types of image data, such as bird, head, baby, butterfly and woman, are selected in SET5 data set; Zebra, pepper, flower, bridge and man are selected in SET14 dataset. Some of the test images in the dataset are shown in Figure 7.

4.2. Experimental procedures

Firstly, we pre-processed the training dataset. The HR images are scaled using the bilinear down-sampling method, and the resolution of the images are reduced by 4 times to obtain the LR images. In order to reduce the interference in the image information, so we de-averaging all LR images, and normalizing all pixels to [−1,1].

In the training phase, the pre-processed LR images are input into the generative network, and the HR images are used as references to improve the ability of the generation through adversarial iteration. After 50 times of training, the final image super-resolution reconstruction model is obtained.

In the testing phase, the trained network model is used for testing. We use bilinear interpolation and SRGAN model as a comparisons.

Figure 7. Some examples in the test image.

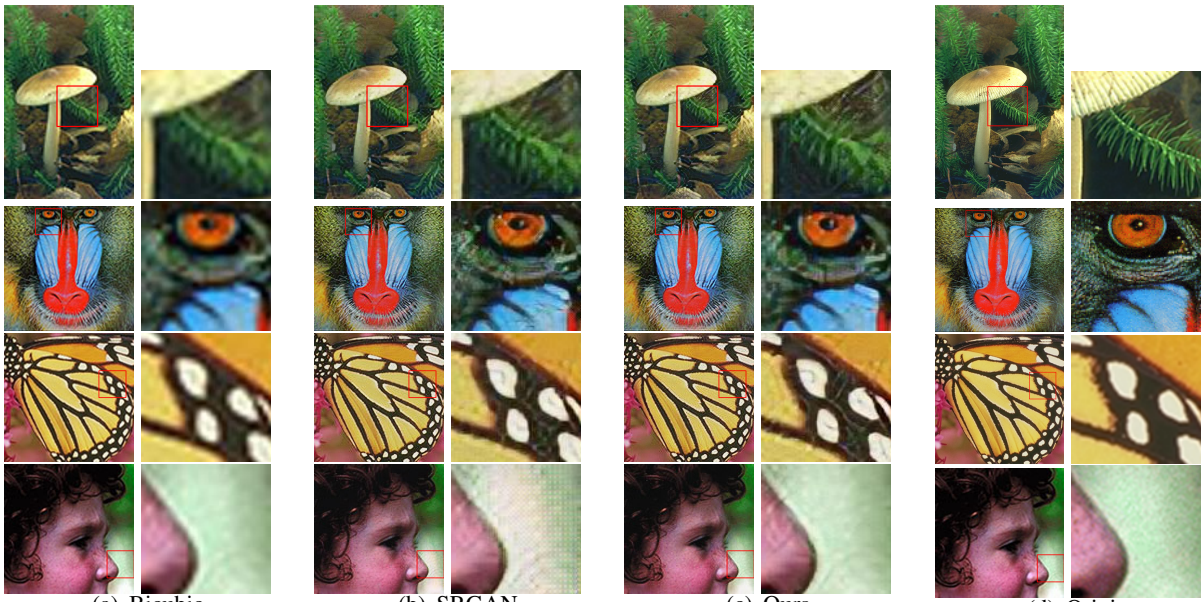
4.3. Experimental results and analysis

Under the same experimental conditions, we compared the reconstruction effects of the bicubic method, the super-resolution method utilizing the generative adversarial network, and our improved method integrated with the self-attention. The experimental results of the three models are shown in Figure 8.

All the low-resolution images used in this paper are obtained by down sampling. The reconstructed images obtained by traditional methods and deep learning methods can not recover the same details as the original image, but in contrast, the reconstruction effect obtained by our method is better. It can be seen from the Figure 3 that the reconstructed images of Bicubic can only retain the high-frequency information such as the contour and color of the images, but reconstruction of details and texture is poor. Compared with the super-resolution image obtained by Bicubic, the effect of the SRGAN model has been significantly improved, and there are better results in terms of texture and contour. Because self-attention improves the attention ability of the network, it better retains the detailed feature information. It can be seen from Figure 3(c) that the improved method proposed in this paper produces higher image quality and reconstructs the details of the image. The color of the reconstructed image is richer and more natural than other methods.

4.4. Quantitative evaluation

In order to verify the effectiveness of this method, we quantitatively evaluate the super-resolution images generated by different methods. For the 50 participants who participated in the subjective



(a) Bicubic (b) SRGAN (c) Ours (d) Origin

Figure 8. Image super-resolution reconstruction results of different methods.

evaluation, without telling them the corresponding method of each reconstructed image, each group of images was scored and the average value was calculated. MOS is shown in Table 4.

Table 4. MOS Result of Figure 8.

Method	Group1	Group2	Group3	Group4
Bicubic	4.200	4.230	4.166	4.067
SRGAN	6.833	7.166	6.833	6.710
SA+SRGAN(Ours)	**7.066**	**7.533**	**7.250**	**6.900**

It shows that the Bicubic interpolation method has the lowest recognition for the effect of image reconstruction, because the interpolation method only enlarges the LR image, and there are a lot of zero-filling operations in the reconstruction process. There is no effective extraction of image feature information. The SRGAN model and the image super-resolution reconstruction method with the self-attention get higher scores. Compared with the reconstruction method that only uses the generative confrontation network, because our method pays attention to the detailed features , it has a better effect on the boundary and detail processing of images while retaining the basic texture features. The best evaluation result was obtained on the group's score. The results show that the method proposed in this article is superior than the other two methods.

In order to compare the effects of different methods more objectively, we further used quantitative indicators such as PSNR and SSIM to quantitatively compare Set5, Set14 and BSD100. The quantitative evaluation results are shown in Table 5.

The results are summarized in Table 4 confirmed that when the scaling factor is 4, the super-resolution reconstruction method based on self-attention we proposed has a certain improvement in PSNR and SSIM compared with other methods. On the Set5 data set , PSNR and SSIM increased by 0.673 and 0.011; on the Set14 data set, the PSNR and SSIM increased by 0.599

Table 5. Quantitative results of different methods in Set5, Set14 and BSD100 data sets.

Method	Set5		Set14		BSD100	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	27.067	0.749	24.502	0.682	23.924	0.626
SRGAN	28.072	0.843	26.057	0.731	25.490	0.686
SA+SRGAN(Ours)	**28.745**	**0.854**	**26.656**	**0.745**	**26.037**	**0.697**

and 0.014, respectively; on the BSD100 data set, the PSNR increased by 0.547, and the SSIM increased by 0.011. The objective evaluation indicators show that our method can better recover the detailed information and improve the quality of the super-resolution.

5. Conclusions

In order to improve the transmission efficiency of image data in the Internet and reduce the huge overhead of data encryption during data transmission, we have propoed an image super-resolution method based on the self-attention. By introducing the attention, the details of the images are preserved during the reconstruction , and the images are more photo-realistic. Our method can effectively reduce the huge overhead of image transmission and data encryption of the Internet , and improve the efficiency of data transmission and the security of data encryption.

At present, most super-resolution reconstruction methods simulate the image degradation process under natural conditions by down-sampling high-resolution images to obtain low-resolution images. However, low-resolution images in practical applications usually have many problems such as motion distortion, optical blur, and noise pollution. How to solve the problem of image degradation caused by down-sampling is also a direction worth exploring for the image super-resolution method of the Internet of Things.

Acknowledgments

This work was supported in part by the Japan Society for the Promotion of Science (JSPS) Grants-in-Aid for Scientific Research (KAKENHI) under Grant 21K17737.

Conflict of interest

All authors declare no conflict of interest in this paper.

References

1. J. Zhang, K. Yu, Z. Wen, X. Qi, A. K. Paul, 3D Reconstruction for Motion Blurred Images Using Deep Learning-Based Intelligent Systems, *CMC Comput. Mater. Continua*, **66** (2021), 2087–2104.
2. W. Wang, H. Xu, M. Alazab, T. R. Gadekallu, Z. Han, C. Su, Blockchain-Based Reliable and Efficient Certificateless Signature for IIoT Devices, *IEEE Trans. Ind. Inf.*, 2021. Available from: https://ieeexplore.ieee.org/document/9444140.

3. L. Tan, H. Xiao, K. Yu, M. Aloqaily, Y. Jararweh, A blockchain-empowered crowdsourcing system for 5g-enabled smart cities, *Comput. Stand. Interfaces*, **76** (2021), 103517.
4. L. Zhen, Y. Zhang, K. Yu, N. Kumar, A. Barnawi, Y. Xie, Early Collision Detection for Massive Random Access in Satellite-Based Internet of Things, *IEEE Trans. Veh. Technol.*, **70** (2021), 5184–5189.
5. B. C. Chifor, I. Bica, V. V. Patriciu, F. Pop, A security authorization scheme for smart home Internet of Things devices, *Future Gener. Comput. Syst.*, **86** (2018), 740–749.
6. L. Zhang, Z. Zhang, W. Wang, Z. Jin, Y. Su, H. Chen, Research on a covert communication model realized by using smart contracts in blockchain environment, *IEEE Syst. J.*, **2021** (2021), 1–12.
7. B. B. Zarpelão, R. S. Miani, S. Rodrigo, C. T. Kawakani, Miani, S. C. de Alvarenga, A survey of intrusion detection in Internet of Things, *J. Network Comput. Appl.*, **84** (2017), 25–37.
8. L. Zhen, A. K. Bashir, K. Yu, Y. D. Al-Otaibi, C. H. Foh, P. Xiao, Energy-efficient random access for LEO satellite-assisted 6G Internet of remote things, *IEEE Internet Things J.*, **8** (2020), 5114–5128.
9. C. Feng, K. Yu, A. K. Bashir, Y. D. Al-Otaibi, Y. Lu, S. Chen, D. Zhang, Efficient and secure data sharing for 5G flying drones: a blockchain-enabled approach, *IEEE Network*, **35** (2021), 130–137.
10. C. Feng, K. Yu, M. Aloqaily, M. Alazab, Z. Lv, S. Mumtaz, Attribute-based encryption with parallel outsourced decryption for edge intelligent IoV, *IEEE Trans. Veh. Technol.*, **69** (2020), 213784–13795.
11. H. Li, K. Yu, B. Liu, C. Feng, Z. Qin and G. Srivastava, An Efficient ciphertext-policy weighted attribute-based encryption for the internet of health things, *IEEE J. Biomed. Health Inf.*, 2021. Available from: https://ieeexplore.ieee.org/document/9416735.
12. D. Qiu, L. Zheng, S. Zhang, Y. Liu, An Image Super-resolution Reconstruction Method by Using of Deep Learning, in *2019 IEEE 4th International Conference on Image, Vision and Computing (ICIVC)*, (2019), 213–217.
13. Y. Yang, H. Cai, Z. Wei, H. Lu, K. K. R. Choo, Towards lightweight anonymous entity authentication for IoT applications, in *Australasian conference on information security and privacy*, Springer, Cham, (2016), 265–280.
14. C. Sun, J. Lv, J. Li, R. Qiu, A rapid and accurate infrared image super-resolution method based on zoom mechanism, *Infrared Phys. Technol.*, **88** (2018), 228–238.
15. X. Feng, J. Li, Z. Hua, Guided filter-based multi-scale super-resolution reconstruction, *CAAI Trans. Intell. Technol.*, **5** (2020), 128–140.
16. Z. Huang, C. Jing, Super-resolution reconstruction method of remote sensing image based on multi-feature fusion, *IEEE Access*, **8** (2020), 18764–18771.
17. N. Shi, L. Tan, W. Li, X. Qi, K. Yu, A Blockchain-Empowered AAA Scheme in the Large-Scale HetNet, *Digital Commun. Networks*, 2021. Available from: https://doi.org/10.1016/j.dcan.2020.10.002.
18. Z. Guo, A. K. Bashir, K. Yu, J. C. Lin, Y. Shen, Graph Embedding-based Intelligent Industrial Decision for Complex Sewage Treatment Processes, *Int. J. Intell. Syst.*, 2020. Available from: https://doi.org/10.1002/int.22540.

19. G. T. Reddy, M. P. K. Reddy, K. Lakshmanna, R. Kaluri, D. S. Rajput, G. Srivastava, T. Baker, SAnalysis of dimensionality reduction techniques on big data, *IEEE Access*, **8** (2020), 54776–54788.
20. L. Zhang, Y. Zou, W. Wang, Z. Jin, Y. Su, H. Chen, Resource allocation and trust computing for blockchain-enabled edge computing system, *Comput. Secur.*, **105** (2021), 102249.
21. X. Yao, Q. Wu, P. Zhang, F. X. Bao, Adaptive rational fractal interpolation function for image super-resolution via local fractal analysis, *Image Vision Comput.*, **82** (2019), 39–49.
22. J. Song, Q. Zhong, W. Wang, C. Su, Z. Tan, Y. Liu, FPDP: Flexible privacy-preserving data publishing scheme for smart agriculture, *IEEE Sens. J.*, 2020. Available from: https://ieeexplore.ieee.org/document/9170612.
23. W. Wang, H. Huang, L. Zhang, C. Su, Secure and efficient mutual authentication protocol for smart grid under blockchain, *Peer Peer Networking Appl.*, **2020** (2020), 1–13.
24. L. Wang, S. Yang, J. Jia, A super-resolution reconstruction algorithm based on feature fusion, *2020 39th Chinese Control Conference (CCC)*, (2020), 3060–30605.
25. R. R. Schultz, R. L. Stevenson, A Bayesian approach to image expansion for improved definition, *IEEE Trans. Image Process.*, **3** (1994), 233–242.
26. M. Yu, H. Wang, M. Liu, P. Li, Overview of Research on Image Super-Resolution Reconstruction, *2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE)*, (2011), 131–135.
27. K. T. Gribbon, D. G. Bailey, A novel approach to real-time bilinear interpolation, in *Proceedings. DELTA 2004. Second IEEE International Workshop on Electronic Design, Test and Applications* , (2004), 126–131.
28. R. Keys, Cubic convolution interpolation for digital image processing, *IEEE Trans. Acoust. Speech Signal Process.*, **29** (1981), 1153–1160.
29. R. Tsai, Multiframe image restoration and registration, *Adv. Comput. Visual Image Process.*, **1** (1984), 317–339.
30. Y. Abe, Y. J. Iiguni, Image restoration from a downsampled image by using the DCT, *Signal Process.*, **87** (2007), 2370–2380.
31. P. Liu, H. Zhang, K. Zhang, L. Lin, W. Zuo, Multi-level wavelet-CNN for image restoration, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, (2018), 773–782.
32. S. W. Jung, T. H. Kim, S. J. Ko, A novel multiple image deblurring technique using fuzzy projection onto convex sets, *IEEE Signal Process. Lett.*, **16** (2009), 192–195.
33. C. Dong, C. C. G. Loy, K. M. He, X. O. Tang, Learning a deep convolutional network for image super-resolution, *European conference on computer vision*, (2014), 184–199.
34. D. Sun, Q. Gao, Y. Lu, Z. Huang, T. Li, A novel image denoising algorithm using linear Bayesian MAP estimation based on sparse representation, *Signal Process.*, **100** (2014), 132–145.
35. H. Li, Q. Zheng, J. Zhang, Z. Du, Z. Li, B. Kang, Pix2Pix-Based Grayscale Image Coloring Method, *J. Comput.-Aided Comput. Graphics*, **33** (2021), 929–938.

36. J. Kim, J. K. Lee, K. M. Lee, Accurate image super-resolution using very deep convolutional networks, *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), 1646–1654.
37. K. Yu, L. Lin, M. Alazab, L. Tan, B. Gu, Deep learning-based traffic safety solution for a mixture of autonomous and manual vehicles in a 5G-enabled intelligent transportation system, *IEEE Trans. Intell. Transp. Syst.*, **22** (2020), 4337–4347.
38. K. Yu, L. Tan, M. Aloqaily, H. Yang, Y. Jararweh, Blockchain-enhanced data sharing with traceable and direct revocation in IIoT, *IEEE Trans. Ind. Inf.*, **17** (2021), 7669–7678.
39. C. Y. Ma, J. W. Zhu, Y. J. Li, J. R. Li, Y. Jiang, X. Li, Single image super resolution via wavelet transform fusion and SRFeat network, *J. Ambient Intell. Hum. Comput.*, (2020), 1–9.
40. K. Yu, M. Arifuzzaman, Z. Wen, D. Zhang, T. Sato, A Key Management Scheme for Secure Communications of Information Centric Advanced Metering Infrastructure in Smart Grid, *IEEE Trans. Instrum. Meas.*, **64** (2015), 2072–2085.
41. L. Tan, K. Yu, A. K. Bashir, X. Cheng, F. Ming, L. Zhao, et al., Towards Real-time and Efficient Cardiovascular Monitoring for COVID-19 Patients by 5G-Enabled Wearable Medical Devices: A Deep Learning Approach, *Neural Compu. Appl.*, 2021. Available from: https://doi.org/10.1007/s00521-021-06219-9.
42. M. A. Talab, S. Awang, S. A. M. Najim, Super-low resolution face recognition using integrated efficient sub-pixel convolutional neural network (ESPCN) and convolutional neural network (CNN), *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, (2019), 331–335.
43. C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, et al., Photo-realistic single image super-resolution using a generative adversarial network, *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2017), 4681–4690.
44. A. Wang, Z. Fang, Y. Gao, X. Jiang, S. Ma, Depth estimation of video sequences with perceptual losses, *IEEE Access*, **6** (2018), 30536–30546.
45. B. Lim, S. Son, H. Kim, S. Nah, K. Mu Lee, Enhanced deep residual networks for single image super-resolution, *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, (2017), 136–144.
46. L. Tan, N. Shi, K. Yu, M. Aloqaily, Y. Jararweh, A Blockchain-Empowered Access Control Framework for Smart Devices in Green Internet of Things, *ACM Trans. Internet Technol.*, **21** (2021), 1–20.
47. Z. Guo, K. Yu, A. Jolfaei, A. K. Bashir, A. O. Almagrabi, N. Kumar, A Fuzzy Detection System for Rumors through Explainable Adaptive Learning, *IEEE Trans. Fuzzy Syst.*, 2021. Available from: https://doi.org/10.1109/TFUZZ.2021.3052109.
48. L. Guo, M. Woźniak, An image super-resolution reconstruction method with single frame character based on wavelet neural network in internet of things, *Mobile Networks Appl.*, **26** (2021), 390–403.

49. W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, et al., Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network, *Proceedings of the IEEE conference on computer vision and pattern recognition*, (2016), 1874–1883.
50. A. Creswell, T. White, V. Dumoulin, K. Arulkumaran, V. Sengupta, A. A. Bharath, Generative adversarial networks: An overview, *IEEE Signal Process. Mag.*, **35** (2018), 53–65.
51. M. Yu, H. Wang, M. Liu, P. Li, Overview of Research on Image Super-Resolution Reconstruction, *2021 IEEE International Conference on Information Communication and Software Engineering (ICICSE)*, (2021), 131–135.
52. S. Lei, X. Liao, Z. Tao, Content-aware Upsampling for Single Image Super-resolution, *2020 Asia-Pacific Conference on Image Processing, Electronics and Computers (IPEC)*, (2020), 213–217.
53. S. E. El-Khamy, M. M. Hadboud, M. I. Dessouky, B. M. Salam, F. E. A. El-Samie, A new super-resolution image reconstruction algorithm based on wavelet fusion, *Proceedings of the Twenty-Second National Radio Science Conference, 2005. NRSC 2005.*, (2005), 195–204.
54. D. Mualfah, Y. Fatma, R. Ramadhan, Anti-forensics: The image asymmetry key and single layer perceptron for digital data security, *Journal of Physics: Conference Series*, **1517** (2020), 012106.
55. J. Johnson, A. Alahi, L. Fei-Fei, Perceptual losses for real-time style transfer and super-resolution, *European conference on computer vision*, (2016), 694–711.
56. K. Fu, J Peng, H. Zhang, X. Wang, J. Frank, Image super-resolution based on generative adversarial networks: a brief review, *Comput. Mater. Continua*, **64** (2020), 1977–1997.
57. M. Heon, J. H. Kim, J. H. Choi, J. S. Lee, Generative adversarial network-based image super-resolution using perceptual content losses, *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*, 2018.
58. Z. Guo, L. Tang, T. Guo, K. Yu, M. Alazab, A. Shalaginov, Deep graph neural network-based spammer detection under the perspective of heterogeneous cyberspace, *Future Gener. Comput. Syst.*, **117** (2021), 205–218.
59. K. Yu, L. Tan, X. Shang, J. Huang, G. Srivastav, P. Chatterjee, Efficient and Privacy-Preserving Medical Research Support Platform Against COVID-19: A Blockchain-Based Approach, *IEEE Consum. Electron. Mag.*, **10** (2021), 111–120.
60. D. Lee, S. Lee, H. Lee, K. Lee, H. J. Lee, Resolution-preserving generative adversarial networks for image enhancement, *IEEE Access*, **7** (2019), 110344–110357.
61. C. F. Song, Y. Huang, W. L. Ouyang, L. Wang, Mask-guided contrastive attention model for person re-identification, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2018), 1179–1188.
62. A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, et al., Attention is all you need, in *Advances in neural information processing systems*, (2017), 5998–6008.
63. A. N. Moldovan, I. Ghergulescu, C. H. Muntean, A novel methodology for mapping objective video quality metrics to the subjective MOS scale, *2014 IEEE International Symposium on Broadband Multimedia Systems and Broadcasting*, (2014), 1–7.
64. Z. Wang, A. C. Bovik, H. R. Sheikh, E. P. Simoncelli, Image quality assessment: from error visibility to structural similarity, *IEEE Trans. Image Process.*, **13** (2004), 600–612.

65. K. Yu, L. Tan, L. Lin, X. Cheng, Z. Yi, T. Sato, Deep Learning Empowered Breast Cancer Auxiliary Diagnosis for 5GB Remote E-Health, *IEEE Wireless Commun.*, **28** (2021), 54–61.

66. L. Tan, K. Yu, F. Ming, X. Cheng, G. Srivastava, Secure and Resilient Artificial Intelligence of Things: a HoneyNet Approach for Threat Detection and Situational Awareness, *IEEE Consum. Electron. Mag.*, 2021. Available from: https://doi.org/10.1109/MCE.2021.3081874.

67. Z. Guo, K. Yu, Y. Li, G. Srivastava, J. C. W. Lin, Deep Learning-Embedded Social Internet of Things for Ambiguity-Aware Social Recommendations, *IEEE Trans. Network Sci. Eng.*, 2021. Available from: https://doi.org/10.1109/TNSE.2021.3049262.

68. K. Yu, Z. Guo, Y. Shen, W. Wang, J. C. Lin, T. Sato, Secure Artificial Intelligence of Things for Implicit Group Recommendations, *IEEE Internet Things J.*, 2021. Available from: http://dx.doi.org/10.1109/JIOT.2021.3079574.

AIMS Press

© 2021 the Author(s), licensee AIMS Press. This is an open access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0)