electronics

MDPI

Article

A Novel Deep-Learning-Based Enhanced Texture Transformer Network for Reference Image Super-Resolution

Changhong Liu [1,*], Hongyin Li [2], Zhongwei Liang [1], Yongjun Zhang [3], Yier Yan [2], Ray Y. Zhong [4] and Shaohu Peng [2,*]

[1] School of Mechanical and Electrical Engineering, Guangzhou University, Guangzhou 510006, China
[2] School of Electronics and Communication Engineering, Guangzhou University, Guangzhou 510006, China
[3] School of Electromechanical Engineering, Guangdong University of Technology, Guangzhou 510006, China
[4] Department of Industrial and Manufacturing Systems Engineering, The University of Hong Kong, Hong Kong 999077, China
* Correspondence: lch@gzhu.edu.cn (C.L.); pengsh@gzhu.edu.cn (S.P.)

Abstract: The study explored a deep learning image super-resolution approach which is commonly used in face recognition, video perception and other fields. These generative adversarial networks usually have high-frequency texture details. The relevant textures of high-resolution images could be transferred as reference images to low-resolution images. The latest existing methods use transformer ideas to transfer related textures to low-resolution images, but there are still some problems with channel learning and detailed textures. Therefore, the study proposed an enhanced texture transformer network (ETTN) to improve the channel learning ability and details of the texture. It could learn the corresponding structural information of high-resolution texture images and convert it into low-resolution texture images. Through this, finding the feature map can change the exact feature of images and improve the learning ability between channels. We then used multi-scale feature integration (MSFI) to further enhance the effect of fusion and achieved different degrees of texture restoration. The experimental results show that the model has a good resolution enhancement effect on texture transformers. In different datasets, the peak signal to noise ratio (PSNR) and structural similarity (SSIM) were improved by 0.1–0.5 dB and 0.02, respectively.

Keywords: deep learning; texture transformer; generative adversarial network; super-resolution; attention mechanism

Citation: Liu, C.; Li, H.; Liang, Z.; Zhang, Y.; Yan, Y.; Zhong, R.Y.; Peng, S. A Novel Deep-Learning-Based Enhanced Texture Transformer Network for Reference Image Super-Resolution. *Electronics* **2022**, *11*, 3038. https://doi.org/10.3390/electronics11193038

Academic Editor: Donghyeon Cho

Received: 10 August 2022
Accepted: 20 September 2022
Published: 24 September 2022

1. Introduction

In the field of images, super-resolution has a wide range of practical uses and application scenarios. For example, digital imaging technology [1], deep space satellite remote sensing technology [2], target recognition analysis technology [3] and medical image analysis technology [4] have a far-reaching influence. At the same time, super-resolution technology can also make up for the limitations of hardware implementation, and has the advantages of a low cost and a short cycle, so it has become a hot topic in the field of image processing.

For traditional single-image super-resolution (SISR), the super-resolution of an image is a process of recovery from low resolution (LR) to high resolution (HR) [5]. Recently, it has been possible to take an image's content information to another level with good visual resolution. For example, convolution neural networks [6,7] were introduced to improve the performance of SISR. However, they had some problems with single images, as the existing methods for super-resolution (SR) of images still produce some blurry results, especially for 2× and 4× scale images. In recent years, reconstruction loss [8] has been inadequate for solving these issues. Thus, a high-resolution texture is too challenging to restored through the degradation process, which will result in a blurred effect [9]. Therefore, some researchers have proposed adversarial loss [10,11], which may be a solution for these

problems and produce good visual quality for scholars who have analyzed the adversarial loss function in the field of super-resolution [11,12]. It was found that one kind of loss alone cannot produce good visual quality [13,14]. Finally, a combination of these loss functions can produce different effects for better visual perception [15,16]. In a recent study, a super-resolution-based algorithm was proposed that required two images for the texture transformer [17], namely one single image for super-resolution and another reference image for super-resolution [18,19]. Currently, the most state-of-the-art (SOTA) method is to use one image as the reference, then another as the LR input. The texture of the reference image is then used to provide a low resolution.

Therefore, this can lead to unsatisfactory SR images. Zhang et al. [19] adopted a feature space determined by a pre-trained classification model to search for and transfer textures between the LR and reference images. Nevertheless, high-level semantic features have textures with different presentations [20,21]. To solve the problem of high-level semantic features, Yang et al. [18] introduced some closely related associated modules optimized for image generation tasks. Through the attention mechanism module, the HR features in the reference images are transformed and fused into LR features extracted from the trunk image through the attention image.

The design encourages the development of a more accurate method to search for and transfer related textures from the reference image to low-resolution images. At the same time, the number of research parameters in this project is insignificant. Under the same hardware configuration, it can successfully shrink the number of parameters and the storage space of the model, meaning that the application of the model in the industrial field can settle the storage problem of SISR in industrial practice [22].

This study proposed a multi-scale feature fusion network, in which the main research goal was to apply super-resolution in the manufacturing field so that the super-resolution could ultimately meet the application requirements of the manufacturing field. The features were learned at different scales ($1\times$, $2\times$ and $4\times$) to permit more representation and the final output was a $4\times$ image. In prior research work, the texture transformation relative to the field-learning ability of the feature mapping network was not powerful and the ability to transmit low-frequency information between channels was insufficient. Therefore, this study proposed a texture transformation network of super-resolution algorithms, which was able to search for and transfer reference images related to low-resolution texture images. Finally, compared with current methods, the main contributions of this study are as follows.

Firstly, the various spatial pyramid pools were used to capture the images' information on multiple scales, for which the sampling performance was enhanced.

Secondly, the block module for receptive fields was intended to help the network improve the learning ability of large receptive fields and to balance a small amount of computation between channels.

Finally, multi-residual channel attention blocks were applied to the network, which enhanced the interdependent learning ability of the feature channels and enriched the backward propagation of low-frequency information through identity mapping, guaranteeing a favorable flow of information and accelerating the training of neural networks.

2. Materials and Methods

2.1. Single-Image Super-Resolution

In recent years, compared with other traditional methods, deep-learning-based methods have significantly improved in PSNR/SSIM. In the field of GAN, images' texture information can be restored into high-resolution images through Wasserstein loss [10,11]. At the same time, the recently emerging transformer network structure can also restore images' texture information very well [19]. In addition, the reconstruction of high-resolution pixels by sampling the local distribution and the corresponding plane coordinates can reasonably solve the problem of selection performance and efficiency [23,24].

Therefore, this study mainly focused on the transformer's structure. The application converts the reference images' textures sources and channels the information to low-resolution (LR) images, which substantially enhances the performance of the input images.

Dong et al. [6] first proposed the idea of using a convolution neural network (CNN) for super-resolution. This includes three parts, namely patch extraction, non-linear mapping and reconstruction, and thoroughly embedded convolution within the super-resolution field. Experiments have proven its validity in terms of deep learning, and enhanced performance has been achieved. Later, Wang et al. [25] combined deep learning with sparse expression. At the same time, Chao et al. [5] integrated sparse expression into CNNs by decomposing the input globally, then proposing two specific components; ultimately, the fusion of the features improved the performance [26]. In order to apply deep learning CNNs to SR fields, Lim et al. [7] directly extracted the features from low-resolution images, which also achieved better performance than the enlarged LR images processed by bicubic interpolation. Meng et al. [27] found that the residual attention network applied in the field of super-resolution had good results for local image recovery. Kim et al. [8] proposed DRRN and Kim et al. proposed VDSR [28] to further improve the performance of SRCNN [6].

In general, there is a mean square error (MSE) loss between SR and HR images, but this may not always be consistent with human evaluations. In recent years, many researchers have conducted research on improving the visual quality of perceived loss. For example, John et al. [16] introduced a perception loss function into a SR project, and Ledig et al. [10] first introduced the idea of GAN into SR fields(SRGAN) [29,30]. The adversarial network was used to minimize the perceived correlation distance between SR and HR. Sajjadi et al. [31] further integrated the matching loss of a texture based on the idea of transferring the style to the texture in super-resolution ESRGAN [32], which is a type of SRGAN [10], by proposing RRDB [33]. The recently proposed RSRGAN [34] trains a sequence and uses the content loss to optimize the perceived quality, achieving more advanced visual results.

2.2. Super-Resolution of Reference Images

Unlike SISR using a single LR image as the input, the reference images can align or patch-match images to obtain more accurate details. In general, some existing reference super-resolution (RefSR) methods [18,19,35] chose to align the LR and reference (Ref) images, but the reference image needs to have a texture and a content structure similar to those of the LR image. Huanjing et al. [21] solved the problem by global repair to align the reference image and LR image. Zhang et al. [17] proposed the CrossNet streamer method to align LR images with the reference image in proportion and fuse them into the corresponding layer of the decoder. However, these methods [17,36] still have some limitations and require good alignment of the low-resolution and reference images. Meanwhile, the methods [21] require extensive resources and they are not conducive to practical applications.

Recently, with existing reference image methods, SRNTT [19] textured between the VGG [37] features of LR and reference images to produce the final output. Nevertheless, it ignored the correlation between the original features and the switching features. All switching features were inputted into the main network equally. Therefore, Yang et al. [18] combined the transformer with super-resolution for the first time, which could solve the problems of SRNTT. Nevertheless, there are still some issues, such as the inadequate extraction of spatial information and the unbalanced channel computation during fusion.

In recent years, RefSR methods have tended to ignore potentially large differences in the distribution, which has affected the effectiveness of the information utilized. The MASA [38] network was proposed, in which the current modules were designed to address these issues. Next, in order to enhance the ability to learn the significant details of the features of Ref images, the interference of noisy information was attenuated by introducing a multi-attentive mechanism using dual-view supervision to motivate the network to learn more accurate feature representations, and a DSMA [36] network was proposed. In a study

on SRTNN [19], the reference image was transferred to the LR input when restored to high-resolution detail, and patches were applied to match VGG features between the LR and Ref images by exchanging similar texture features. However, the SRNTT [19] network ignored the correlation between the original and swapped features, meaning that all the swapped features entered the main network equally. To address these issues, a texture transfer network (TTSR) [18] was developed, which enabled our method to search for and transfer correlated textures from Ref to LR images. Ignoring the fact that the LR space still contains valuable high-frequency details, TTSR [18] could not effectively fuse the two independently extracted features or extracting the finer features in LR space, so a new fusion module to combine LR and Ref features more effectively is needed, for which the DPSR network has been proposed. Thus, the ability to capture the contextual information of images and channel learning has not been fully achieved.

To solve these problems, the receptive field block can balance a small amount of computation and enlarge the receptive field [39], atrous spatial pyramid pools can capture a larger proportion of the contextual information of images, and the residual channel attention block can enhance the learning ability of the channel. In this study, a multi-scale fusion texture transformer network was proposed to further improve the performance.

3. The Proposed ETTN

In this section, we introduce the proposed ETTN to enhance the texture transformer network. It mainly consists of four parts. A multi-scale fusion integration (MSFI) method was introduced to further boost the model's performance. Its transformer texture comes from the backbone transformer *F* to achieve MSFI. It includes a receptive field block, a residual channel attention block and atrous spatial pyramid pools. This group of loss functions for optimizing the proposed network was clarified. Finally, the hardware environment was also considered.

3.1. The Enhanced Texture Transformer Network

In this study, low-resolution images and reference images needed to be prepared in the network. The first image was a low-resolution image to be restored. We then needed to prepare reference images related to the low-resolution images. Inspired by [23], the bicubic interpolation method achieved a better performance, so the bicubic interpolation method was selected as the main interpolation method for upsampling/downsampling in this network. The texture features need three images in the networks as the input feature vectors *K*, *V* and *Q*, so we needed to find the input image and the reference image. After four rounds of bicubic interpolation, the obtained reference image was expressed as Ref ↓↑ and the LR image was expressed as LR↑. Meanwhile, the reference image needed to be represented as Ref in order to input it into the texture transformer network. The corresponding features of the LR images were outputted by the backbone, and the output feature maps were then used to generate super-resolution images with the three feature vectors through feature extraction and fusion. The texture converter's structure is illustrated in Figure 1.

Feature extraction: In the reference image, texture extraction is essential because accurate and proper texture information will assist the generation of super-resolution images. Instead of the semantic features extracted by a pre-trained classification model such as VGG19 [37], the design feature are extracted, for which the parameters are updated during end-to-end training. The high-resolution reference image is transformed into a network.

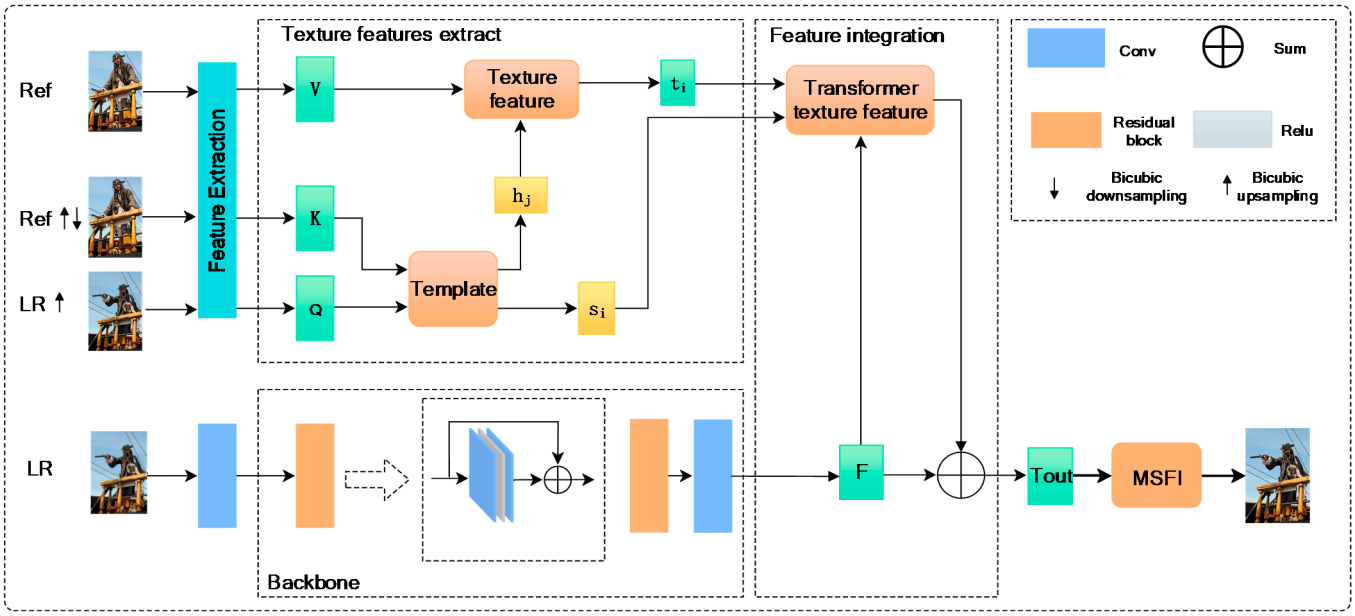


Figure 1. The architecture of the proposed enhanced texture transformer network (ETTN).

Here, Q is the LR sampling and K is the bicubic Ref image. They are the input images fed into the feature extraction module and to extract the features Q, K and the mapping to the template.

$$\begin{aligned} Q &= f_{FE}(Bicubic\ LR) \\ K &= f_{FE}(Bicubic\ Ref) \\ V &= f_{FE}(Ref) \end{aligned} \tag{1}$$

V is the feature extraction of the reference image, and then the output part of the template is combined with V in the texture feature module. The extracted texture features, namely Q (query), K (key) and V (value), indicate three basic elements of the attention mechanism inside a transformer and are also used in our template module.

The attention relationship is the correlation between the LR image and the Ref image embedded within the transformation of features through the similarity between Q and K.

$$Attention(V, K, Q) = V \times \left(\frac{Q}{||Q||} \cdot \frac{K}{||K||} \right) \tag{2}$$

This is also a major component of attention and takes three vector parameters: Q, K, and V. The similarity between Q and K is calculated, then the calculated result is multiplied by the weighted coefficient of the corresponding V. Finally, the weighted sum is used to obtain the attention value.

Template: We propose a relevance attention module to transfer the HR K for each query Q. However, such an operation may cause a blurring effect which lacks the ability to transfer the features of the HR texture. For this, Q unfolds to patch $\{q_1, \ldots, q_i\}$, and K expands to patch $\{k_1, \ldots, k_i\}$. Next, for each patch of Q, we find its most relevant patch in K. Finally, we perform dense patch-matching on the unrolled patches of Q and K. For example, for the ith patch q_i, we calculate the cosine similarity of each q and k patch as:

$$\eta = \left\langle \frac{q_i}{||q||}, \frac{k_j}{||k||} \right\rangle \tag{3}$$

This module is also used to obtain the template and the maps of extracted attention.

Texture features: In the experiment, we used the T_i HR texture feature V from the Ref image. However, this option may cause a blurred effect which lacks the power to transform

the HR texture features. Therefore, we first extracted more texture features to calculate a template map transformer h_i, in which the ith element was calculated from the relevance.

$$h_i = argmaxh_{i,j} \tag{4}$$

where h_i is the maximum index position, which is the best related position of the Ref image. We also needed to convert the texture of Ref T to LR so that we could select the non-collapsed patch V as our template map, where t_i denotes the value of t in the ith position, which was selected from position h_i of V. Therefore, the low resolution changes from the reference images to the high-resolution feature representation T.

Transformed texture features: In this study, we proposed the combination of an attention extraction module and the LR texture feature F extracted by the backbone network. This helped to enhance the attention texture's transformation effect, and reduced the transfer of unrelated textures. Therefore, the experiment needed to calculate $\eta_{i,j}$ in the template and obtain the maximum mapping value S, which represents the confidence degree T of the enhanced and transformed texture.

$$s_i = \max_j \eta_{i,j} \tag{5}$$

where S_i represents the position of confidence. The main purpose was to integrate the HR texture feature T with the texture F proposed by backbone through the concat method to improve the detailed texture of LR images. Such fused features were further multiplied element-wise by the attention map S and added back to F to obtain the final output of the texture transformer. This operation can be represented as

$$T_{out} = F + concat(F, T) \cdot S \tag{6}$$

where T_{out} represents the fusion output, and *concat* represents the operation of convolution, which expresses multiplication between two elements. The values of the corresponding image size of T_{out} are half and quarter of the original image. Finally, multiplication of the elements with the confidence S plus the original F is carried out.

3.2. Multi-Scale Feature Integration

We proposed a method of multi-scale feature integration, which is mainly divided into three parts: RFB, RCAB and ASPP.

Receptive field block (RFB): Extreme super-segmentation is needed to resolve the texture's details. The super-partition was introduced to balance the small amount of computation and enlarge the receptive field, which could extract very detailed features. Its structure is shown in Figure 2. RFB has proven to be powerful in target detection and image recognition [40].

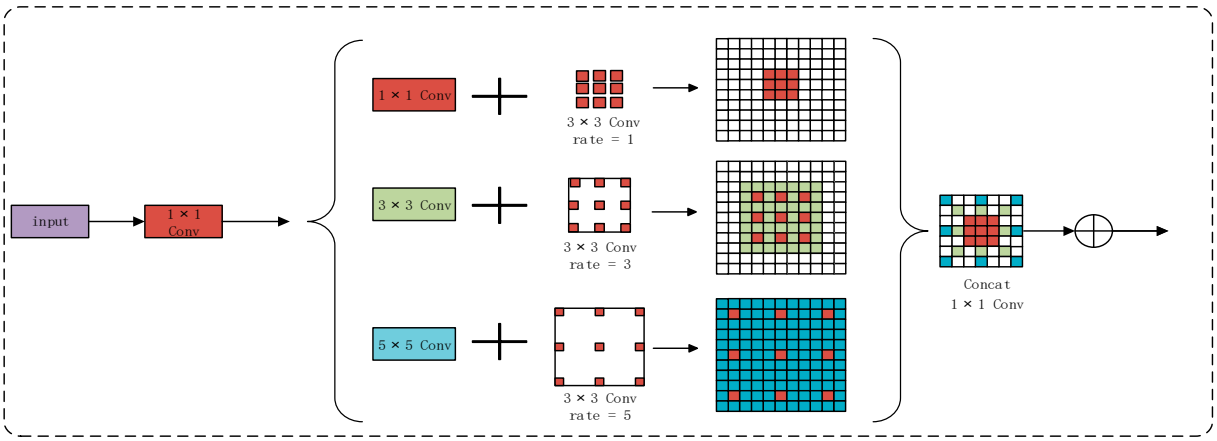


Figure 2. Receptive field block.

Atrous spatial pyramid pool (ASPP): We used this model to enhance the computer channels of the super-resolution field. Parallel sampling of the dilated convolution with different sampling rates for a given input image is equivalent to capturing the context information of the image at multiple scales [41]. In this study, the information was mainly placed on Lv2, Lv3 and Lv4, which were able to capture low-frequency information and expand the acceptance domain. Compared with the previous method [42], there were fewer pooling layers, because this can lead to a decrease in the spatial resolution, thus affecting the performance. As shown in Figure 3, we placed this information on the last layer to help extract the high-frequency information, thus enhancing the resolution.

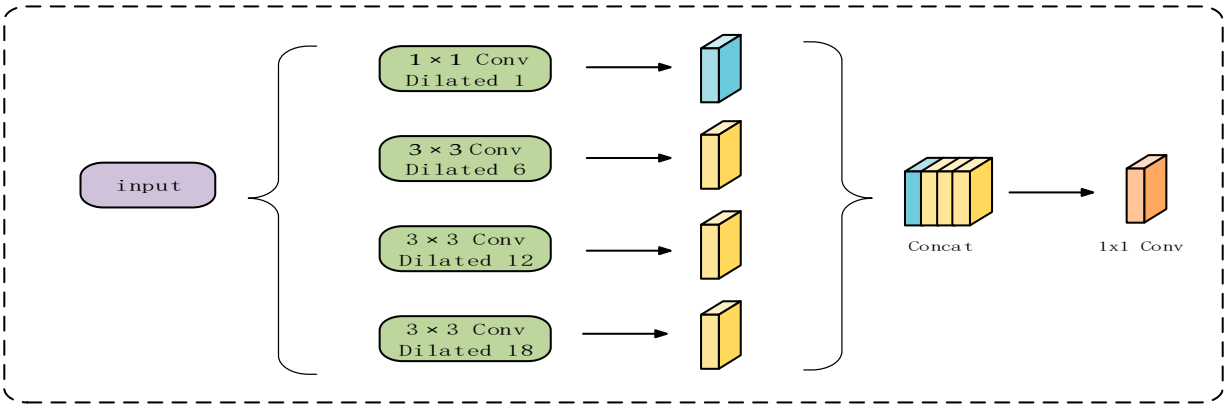


Figure 3. Atrous spatial pyramid pool.

Multi-residual channel attention block (MRCAB): Modeling the interdependence between feature channels is possible through adaptive rescaling of each channel [43]. The upsamping process allowed the network to focus on enhancement of the more useful channel. The transient inside the residual allowed rich low-resolution mapping between the channel and the signal behind the channels, which accelerated the network's training [44,45].

The shortcut inside the residual allowed a large number of low-frequency channels to pass through. Consequently, the method selected for this experiment was the MRCAB composed of four to eight RCABs, which could achieve a better channel transmission effect. The network structure of this experiment is shown in Figure 4.

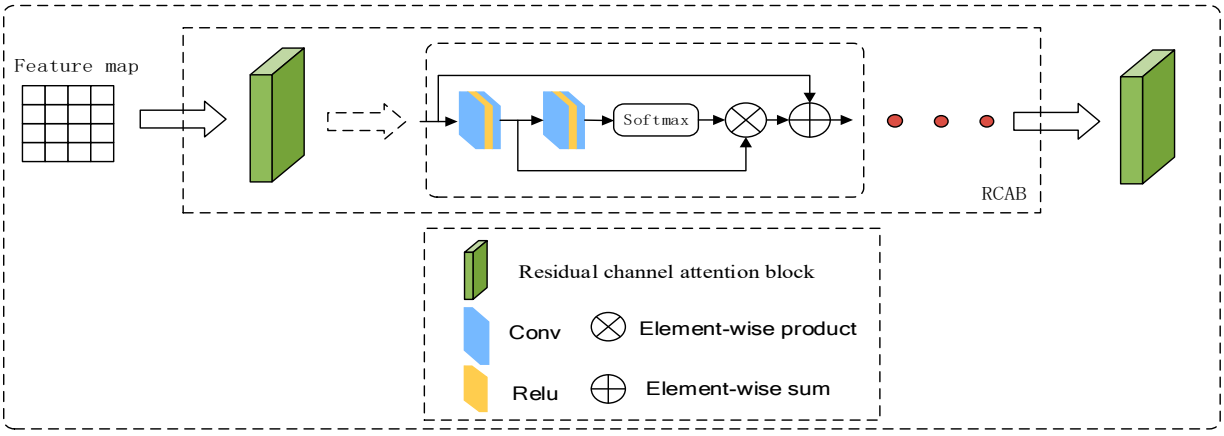


Figure 4. Multi-residual channel attention block.

Meanwhile, three jumps were added to the MSFI module to transmit some low-frequency information. The high-frequency information was improved through the characteristics of the fusion process. Finally, Figure 5 shows the architecture of the multi-scale feature integration process for a better analysis of the low frequencies.

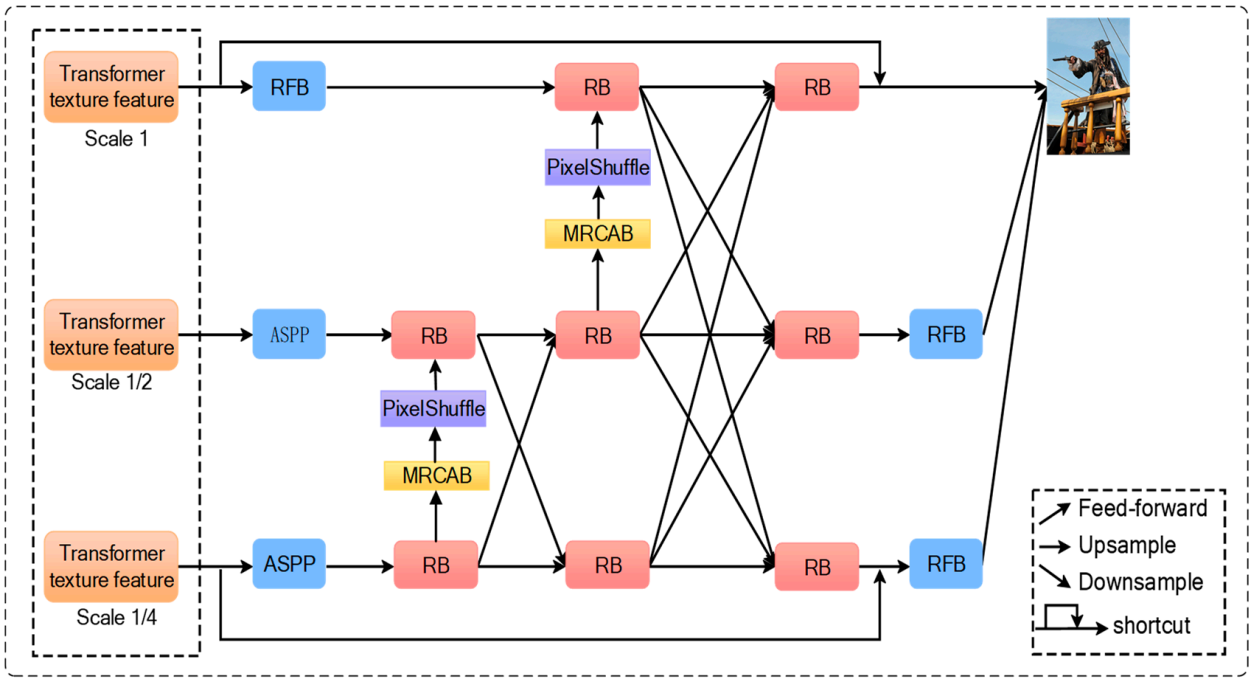


Figure 5. The architecture of the multi-scale feature integration process.

3.3. Objective Function and Evaluation Metrics

To preserve the spatial structure of the LR images, improve the visual quality of the SR images and take advantage of the rich texture from the reference images, the objective function combines reconstruction loss L_{rec}, perceptual loss L_{per} and adversarial loss L_{adv}. The reconstruction loss is adopted in most super-resolution, whereas perceptual and adversarial loss improve the visual quality. The texture loss is specific to the reference super-resolution image.

Reconstruction loss: The overall loss can be interpreted as:

$$L_{rec} = \left|\left| I^{HR} - I^{SR} \right|\right|_1 \tag{7}$$

The use of L_1 loss has been shown to be sharper than L_2 loss, with an increased ease of convergence in performance. Perceptual loss has been proven useful for improving visual quality [11]. The key idea of perceptual loss is to enhance the similarity in the feature space between the prediction image and the target image. The perceptual loss includes two parts

$$\begin{aligned} L_{per} &= \frac{1}{V}\left|\left|\varphi_i^{vgg}(I^{SR}) - \varphi_i^{vgg}(I^{HR})\right|\right|_2^2 \\ &+ \frac{1}{V}\left|\left|\varphi_j^{FE}(I^{SR}) - T\right|\right|_2^2 \end{aligned} \tag{8}$$

where the first part is the traditional perceptual loss, in which $\varphi_i^{vgg}(\cdot)$ denotes the ith layer's feature map of VGG19 and V represents the shape of the feature map of size (C ∗ H ∗ W) at that layer. I_{SR} is the predicted super-resolution image, and the information extracted by feature extraction is constrained by the transformed texture feature T, which is more beneficial for texture transformation of Ref images.

Adversarial loss: Wang et al. [25] found that the adversarial loss can significantly enhance the sharpness of synthesized images. Although the loss model is powerful, it has unstable training results. Therefore, Gulrajani et al. [46] used WGAN to optimize the cost function so as to reduce the gradient and maintain the relative stability of training [13].

Therefore, the use of WGAN in this experiment enhanced the visual effects, for which the expression is as follows:

$$L_{\mathrm{G}} = -E_{\tilde{x}-P_g}[D(\tilde{x})] \tag{9}$$

$$L_{adv} = -E_{\tilde{x}-P_g}[D(\tilde{x})] - L_{\mathrm{G}} + \lambda E_{\hat{x}\sim P_x}[(||\nabla_{\hat{x}} D(\hat{x})||_2 - 1)] \tag{10}$$

Object loss: The experiment in this study consisted of the three loss functions above, and each coefficient of the constraint condition can be expressed as:

$$L_{\mathrm{total}} = \lambda_{rec} L_{rec} + \lambda_{per} L_{per} + \lambda_{adv} L_{adv} \tag{11}$$

Metrics Evaluation: The peak signal to noise ratio (PSNR), the structural similarity (SSIM) and the mean square error (MSE) were deployed as quantity assessment metrics. Specifically, the PSNR of an $M \times N$ ground truth image g relative to the SR image r was calculated as follows

$$PNSR(r,g) = 10\log_{10}(\frac{1}{MSE(r,g)}) \tag{12}$$

where the mean square error (MSE) function was defined as

$$MSE(r,g) = \frac{1}{MN}\sum_{i=1}^{M}\sum_{j=1}^{N}(r_{ij} - g_{ij})^2 \tag{13}$$

The SSIM thus can be represented as

$$SSIM(r,\mathrm{g}) = \frac{(2\mu_r\mu_g + C_1)(2\sigma_{RG} + C_2)}{(\mu_r^2 + \mu_g^2 + C_1)(\sigma_r^2 + \sigma_r^2 + C_2)} \tag{14}$$

where μ_r, σ_r and σ_{rg} are the mean, standard deviation and cross-correlation between the two images r and g, respectively. C_1 and C_2 are positive stabilizing constants.

3.4. Histogram Analysis

To further verify the method, this study used two models to verify the generalization of the image trained by the performance indicators, including the Babbittacharyya distance and Chi-square analysis.

The Bhattacharyya coefficient is an approximate measure of the amount of overlap between two statistical samples. It therefore can be used to evaluate the degree of similarity between two images. Moreover, it can be applied to the similarity calculation of a histogram, which can be used to evaluate the best effect obtained by the Babbittacharyya distance.

$$D(I_{HR}, I_{SR}) = \sqrt{1 - \frac{1}{\sqrt{\bar{I}_{HR}\bar{I}_{SR}N^2}}\sum_i \sqrt{I_{HR}(i)\cdot I_{SR}(i)}} \tag{15}$$

where I_{hr} and I_{sr} are, respectively, the histogram data of the high-resolution images and low-resolution images. N is the coefficient. When the calculated result is 0, the two images are perfectly correlated.

Chi-square: Chi-square comparison involves the degree of deviation between the actual value of a high-resolution pixel and the value of a super-resolution pixel. It determines the Chi-square value and can also be used to calculate the correlation.

$$D(I_{HR}, I_{SR}) = \sum_i \frac{(I_{HR}(i) - I_{SR}(i))^2}{I_1(i)} \tag{16}$$

When the pixels of the two images are the most similar, the minimum value is D (I_{HR}, I_{SR}) = 0. Meanwhile, the smaller the Chi-square, the better the similarity, which also implies that better correlation between the two images.

4. Experiments

4.1. Datasets and Evaluation Metrics

To verify the superior performance of the model, tests were carried out on the CUFED5 proposed by Zhang et al. [19], in which the datasets consisted of 11,871 pairs of input images and reference images. The test consists of five similar images from the CUFED5 datasets. Accordingly, the generalization and adaptability of the model proposed in this study was verified using other datasets from Urban100 [47] and Sun80 [48]. Firstly, to introduce Urban100, an LR image was selected and put into the network as the reference image. Since these datasets are composed of architectural images, which have high textural similarity, the model has outstanding search and texture transfer effects. Second, Sun80 contains natural landscape images that match the reference images. Nevertheless, prior experience used Urban100 for testing. Its datasets are composed of curves and planes. Therefore, we use HR images as the reference images so that the low-resolution images were used as input images. The model finally used the RGB channel to evaluate the results of PSNR and SSIM.

4.2. Implementation Details

In this study, except for the convolution layer marked by the parameters, the size of the convolution kernel was 3, the number of channels was 64 and the batch size was 18. The Adam optimized the network using an optimizer with $\beta_1 = 0.9$ and $\beta_2 = 0.999$. The learning rate for both the generator and the discriminator was set to 10^{-4}. The loss was used to train the texture converter for 60 epochs, and the initial learning rate was 10^{-4}. The weights of λ_{rec}, λ_{per} and λ_{adv} were 1, 10^{-4} and 10^{-4}, respectively. In addition, the experimental environments were Ubuntu 18.04 and GPU Tesla T4.

4.3. Comparison of Super-Resolution

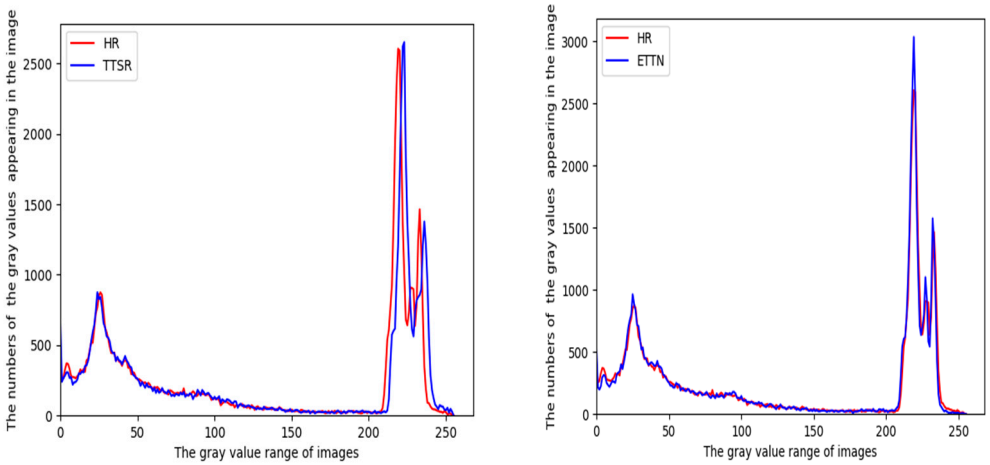
To compare the generalization of the model, we conducted quantitative analysis on the latest models, namely TTSR [18] and SRNTT [19], and the model proposed in this study. The comparative performance mainly relied on the Bhattacharyya distance and Chi-square distance for the comparative analysis. The study selected one image of the CUFED5 datasets and Urban100 datasets as the output image for analysis. The results are shown in Table 1. It can be seen that the model proposed in this study outperformed the others with the addition of higher perception quality.

Table 1. Performance comparison: Bhattacharyya distance and Chi-Square distance.

CUFED im006	Bhattacharyya Distance	Chi-Square Distance	Urban100 im008	Bhattacharyya Distance	Chi-Square Distance
SRNTT	0.219	95,860	SRNTT	0.231	231,325
TTSR	0.187	96,554	TTSR	0.171	393,684
Ours	0.108	13,646	Ours	0.142	196,324

To verify the degree coincidence between the proposed method and the recent TTSR method, the super-resolution images and the original images were used to compute the histograms of a series of arrays in this study. In Figure 6, the y-axis represents the number of different gray values, each of which appears in the whole image. The x-axis is the range of pixels. Counting the degree of conformity of the histograms by the method of counting the histograms, it was found that the higher the degree of conformity, the more identical the images.

The study principally compared the degree coincidence of the images from the CUFED datasets for testing. As shown in the results, the TTSR model still had a higher degree of conformity and still performs better than the ENNT model when the pixels (Figure 6a) were in the range of 50 or so. Though the gray value was around 50–230 with the ENNT, providing a good result (Figure 6b), the ENNT model had a stronger effect on the pixels' range.



(**a**) (**b**)

Figure 6. Histograms of the model images. (**a**) HR images of the TTSR model; (**b**) HR images of our model.

4.4. Ablation Study

In order to prove the impact of the reference images' correlation on the model, we conducted performance tests on the reference images. As shown in Table 2, results with similar levels were obtained, in which L1 to L5 are the test results from the CUFED5 datasets, where L1 is the most correlated level and L5 is the least correlated level with the best results. "LR" represents the low-resolution images of the CUFED5 datasets as a reference. The test sets had a total of 126 samples, each consisting of one person in the HR image and five reference points, namely LR, L1, L2, L3 and L4, which was very convenient for the CUFED5 RefSR study and provided a benchmark for a fair comparison.

Table 2. Ablation study on reference images with different levels of similarity.

Level	L1	L2	L3	L4	L5	LR
TTSR	25.53/0.765	25.30/0.760	25.17/0.750	25.17/0.750	25.23/0.751	25.31/0.756
Ours	25.68/0.768	25.44/0.758	25.34/0.756	25.31/0.755	25.31/0.754	25.46/0.761

The transformer model in this study consists mainly of four parts, including feature extraction of the images, a template, the texture features for the texture transformer and the MSFI module for feature fusion. As shown in Table 3, the study was mainly based on TTSR and then gradually added the effects of ASPP, RFB and RCAB to the experiment. From Table 3, although the data in ASPP have decreased, the main reason was to add the balance of the channel, leading to a performance of 25.30. However, adding RFB and using the residual error of the channel function was good enough to transform the data to the other part. The performance (PSNR) rose to 25.50.

Table 3. Ablation study on the transformer model.

Method	ASPP	RFB	RCAB	Param(M)	PSNR/SSIM
Base + ASPP	√			7.01	25.30/0.749
Base + ASPP + RFB	√	√		7.10	25.50/0.761
Base + ASPP + MRCAB(4/8) + RFB	√	√	√	7.56 (8.22)	25.60/0.767 (25.65/0.768)

The RCAB was added to make the feature channels come from the adaptive scaling of each channel during the upsampling process. The network allowed more channels to enhance the ability to learn. Finally, the features were fully extracted, and the performance

(PSNR) was improved by 0.05. After adding MRCAB, the residual function of the channel could be used to extract more low-frequency information from each channel, which gradually improved the performance. As shown in Table 3, the experiments showed that the effect of using eight residual channels was better.

4.5. Quantitative Evaluation

To evaluate the effectiveness of TTSR [18], this section compared the model with other SISR and super-resolution methods of reference images. The SOTA performance of ENet [49] for both PSNR and SSIM was achieved. RSRGAN [34] was considered to achieve SOTA visual quality. Some algorithms utilize adversarial loss, such as SRGAN [10] ESRGAN [32] and RSRGAN [34], which can retain the features' details as often as possible, hence conserving the texture's information nicely.

In recent years, the latest reference image algorithms MASA [38] and DSMA [36] have allowed the network features to learn more accurate features and convert their textures better. Their performance is also superior, with significantly better performance than previous methods. The algorithms DPSR [35] and TTSR [18], which transform the texture of the reference image to a low-division image to make it a high-resolution image, also have excellent quantitative performance compared with the previous algorithms. All experiments were performed with a scale factor of 4× between the low-resolution images and the original reference images. As shown in Table 4, the performance is ranked according to the PNSR from high to low, compared with the combined effect of the above three losses.

Table 4. PSNR/SSIM/MSE comparison of different super-resolution methods on three datasets. Bold numbers denote the highest scores. A lower MSE means better performance.

Algorithm	CUFED5 [19]	Sun80 [48]	Urban100 [47]
ESRGAN [32]	21.90/0.633/0.00646	24.18/0.651/0.00382	20.91/0.620/0.00810
RSRGAN [34]	22.31/0.635/0.00587	25.60/0.667/0.00275	21.47/0.624/0.00712
SelfEx [47]	23.22/0.680/0.00476	27.03/0.756/0.00198	24.67/0.749/0.00341
Bicubic	24.18/0.648/0.00382	27.26/0.739/0.00188	23.14/0.674/0.00485
ENet [49]	24.24/0.695/0.00377	26.24/0.702/0.00237	23.63/0.711/0.00433
SPSR [50]	24.39/0.714/0.00364	27.94/0.744/0.00160	24.29/0.729/0.00372
SRGAN [10]	24.40/0.695/0.00363	26.76/0.729/0.00211	23.63/0.711/0.00433
Landmark [21]	24.91/0.718/0.00323	27.68/0.776/0.00170	—
MASA [38]	24.92/0.729/0.00322	27.12/0.708/0.00194	23.78/0.712/0.00419
SRCNN [6]	25.33/0.745/0.00293	28.26/0.781/0.00149	24.41/0.738/0.00362
DPSR [35]	25.23/0.808/0.00281	28.42/0.762/0.00144	24.35/0.734/0.00367
SCN [51]	25.45/0.743/0.00285	27.93/0.786/0.00161	24.52/0.741/0.00535
TTSR [18]	25.53/0.765/0.00280	28.59/0.774/0.00138	24.62/0.747/0.00345
SRNTT [19]	25.61/0.764/0.00275	27.59/0.756/0.00174	**25.09/0.774/0.00310**
DSMA [36]	25.61/0.758/0.00275	-	24.55/0.733/0.00354
Ours	**25.67/0.769/0.00271**	**28.80/0.786/0.00131**	**25.09/0.768/0.00310**

Therefore, we trained the model on the CUFED5 dataset and tested it on the CUFED5, Sun80 and Urban100 datasets. As can be seen in Table 4, SRGAN [10] and ESRGAN [32] first provided competitive training performance results. Moreover, most SR algorithms minimized the MSE, such as SCN [51] and SelfEx [47], which could only improve the PNSR and SSIM metrics accordingly, without fine visual quality. In contrast, the latest methods, such as DPSR [35] and DSMA [36], yielded the best texture effects after using the constraints of the three losses, and the proposed method outperformed the SOTA method by 0.06 dB with the CUFED5 test datasets. Furthermore, we achieved improvements with the Sun80 and Urban100 datasets.

The perceptual loss enhanced the similarity between the image and the target image in the feature space. It has been demonstrated that the mutual constraint of these three loss functions can provide excellent results in TTSR and SRNTT. The comparison results showed that good performance indicators were achieved for the three datasets in the test. As shown

in Table 4, our method was distinctly better than the SOTA reference images based on the super-resolution methods TTSR [18] and SRNTT [19]. Although the performance with Urban100 was slightly worse than that of SRNTT [19], as shown in Figure 7, the visual effect was much better than that of the previous methods [17,19].

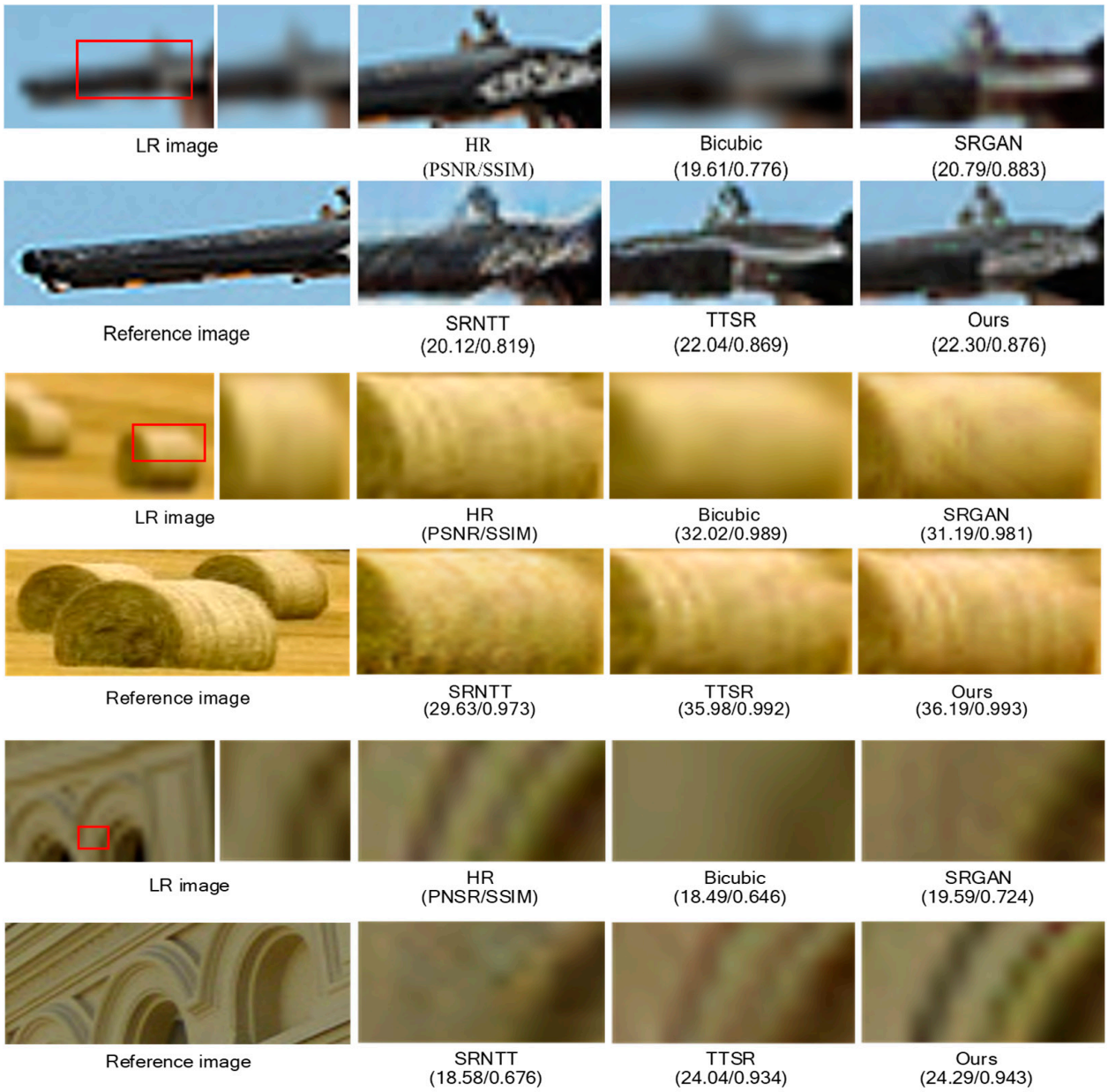


Figure 7. Visual comparison among different SR methods on the CUFED5 [19] testing set (top example), Sun80 [48], (the second example) and Urban100 [47] (the third example). To see the details of the image clearly, the area marked by the red frame is enlarged.

As shown in Table 5, TTSR and our method were faster than SRNTT in terms of execution time. Through comprehensively considering Table 4, it can be concluded that ETTN is slightly inferior to TTSR in terms of network parameters and execution time, but has the best performance in terms of PSNR and SSIM.

Table 5. Comparison of the number of network parameters and execution time. The methods used for comparison are all patch-based RefSR methods.

Method	Param (M)	Average Execution Time (ms)
SRNTT [19]	5.74	3811.18
TTSR [18]	6.73	198.59
Ours	8.22	338.39

4.6. Qualitative Evaluation

Compared with current methods such as SRGAN [10], SRNTT [19] and TTSR [18], as shown in Figure 7, the model in this study could convert the texture information of high-resolution images into reference images and achieved better results.

As shown in the first example in Figure 7, the test used CUFED5 datasets, and the output LR images and the reference image show that the effect of local information recovery was excellent. The second sample was from the Sun80 datasets, and used the same methods as the first example. The method obtained the local details, the effect of texture recovery was better, showing that the effect was excellent. The final example obviously showed the effects of image restoration, where the textural details were extracted very well, specifically highlighting the texture of strips. Therefore, this showed that the algorithm had a remarkable effect on more organized information. Finally, even if the high-resolution reference images and low-resolution input images are not globally correlated, our proposed model can extract extremely extensive parts through local extraction. At the same time, the information can be essentially converted into super-resolution results. Moreover, the proposed method is more suitable for real images than other reference super-resolution methods. Qualitative comparisons showed that the method could successfully integrate textural features with LR features. This is valuable for obtaining satisfactory super-resolution results as a reference.

5. Conclusions

In this study, we proposed a novel deep-learning-enhanced texture transformer network (ETTN) for super-resolution images, which improved performance by transferring textural information from the reference images to LR images. By combining three modules, the textural information weas converted to MSFI for feature fusion and representation to output the image through convolution layers. By using PSNR and SSIM for a performance evaluation of the output images, it was found that the performance of the proposed method has been significantly improved. Thus, the output images were of better quality. Meanwhile, the degree of coincidence of the histograms was used to verify that the model could make the HR images better. Finally, based on the results of the ablation study and the qualitative evaluation, we can draw the conclusion that the model proposed has superior performance for a single image.

Experiments have demonstrated that the ETTN model can be applied to the field of super-resolution. It provides good perceptual quality for the recovery of single images.

This network can be of great importance in the medical field and industry. Next, we will apply super-resolution to MR images in medicine. Considering the simple repetitive structure and distribution of such images, our model can handle the task of processing MR images more efficiently and precisely, in which it decreases the cost of materials to a certain extent and promotes research in medical MR.

For the field of detecting circuit board defects, it can be applied to derive high-resolution images that can better differentiate the location of the defects and facilitate the next steps so researchers can examine the causes. For this reason, the algorithms in this study are of importance for future research in these areas.

Author Contributions: Conceptualization, C.L.; data curation, C.L.; funding acquisition, H.L.; investigation, H.L.; methodology, C.L. and S.P.; project administration, C.L.; resources, C.L.; software, S.P.; supervision, Y.Y. and R.Y.Z.; validation, Z.L. and Y.Z.; writing—review and editing, C.L. All authors have read and agreed to the published version of the manuscript.

Funding: The authors acknowledge the funding of following science foundations: the Science and Technology Planning Project of Guangzhou, China (202102010392); the Science and Technology Planning Project of Guangdong Province, China (2020A1414050067); the National Natural Science Foundation of China, China (51975136, 52075109); and the Innovation and Entrepreneurship Education Project in Guangzhou Universities, China (2020PT104, 2022CXCYZX001).

Data Availability Statement: Publicly available datasets were analyzed in this study. Our training set CUFED5 can be obtained from: https://drive.google.com/drive/folders/1hGHy36XcmSZ1LtARWmGL5OK1IUdWJi3I (accessed on 20 April 2022). The test sets Sun80 and Urban100 are available online at: https://github.com/jbhuang0604/SelfExSR (accessed on 20 April 2022).

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

References

1. Chen, Q.; Song, H.; Yu, J.; Kim, K. Current development and applications of super-resolution ultrasound imaging. *Sensors* **2021**, *21*, 2417. [CrossRef] [PubMed]
2. Dong, X.; Xi, Z.; Sun, X.; Gao, L. Transferred multi-perception attention networks for remote sensing image super-resolution. *Remote Sens.* **2019**, *11*, 2857. [CrossRef]
3. Shermeyer, J.; Van Etten, A. The effects of super-resolution on object detection performance in satellite imagery. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Long Beach, CA, USA, 16–17 June 2019; pp. 1432–1441.
4. Wu, K.; Qiang, Y.; Song, K.; Ren, X.; Yang, W.; Zhang, W.; Hussain, A.; Cui, Y. Image synthesis in contrast MRI based on super resolution reconstruction with multi-refinement cycle-consistent generative adversarial networks. *J. Intell. Manuf.* **2020**, *31*, 1215–1228. [CrossRef]
5. Yang, J.; Wright, J.; Huang, T.; Ma, Y. Image super-resolution as sparse representation of raw image patches. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Anchorage, AK, USA, 23–28 June 2008; pp. 1–8.
6. Dong, C.; Loy, C.C.; He, K.; Tang, X. Image super-resolution using deep convolutional networks. *IEEE Trans. Pattern Anal. Mach. Intell. (PAMI)* **2015**, *38*, 295–307. Available online: https://arxiv.org/abs/1501.00092 (accessed on 19 September 2022). [CrossRef]
7. 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.
8. 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, 26 June 2016; pp. 1637–1645.
9. Freedman, G.; Fattal, R. Image and video upscaling from local self-examples. *ACM Trans. Graph. (TOG)* **2012**, *30*, 1–11. [CrossRef]
10. Ledig, C.; Theis, L.; Huszár, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.; Tejani, A.; Totz, J.; Wang, Z. 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.
11. Goodfellow, I.; Pouget-Abadie, J.; Mirza, M.; Xu, B.; Warde-Farley, D.; Ozair, S.; Courville, A.; Bengio, Y. Generative adversarial nets. *Adv. Neural Inf. Processing Syst.* **2014**, 27. [CrossRef]
12. 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; 2017; pp. 624–632.
13. Zhang, Z.; Song, Y.; Qi, H. Age progression/regression by conditional adversarial autoencoder. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Venice, Italy, 21–30 October 2017; pp. 5810–5818.
14. Lai, W.-S.; Huang, J.-B.; Ahuja, N.; Yang, M.-H. Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2018**, *41*, 2599–2613. [CrossRef]
15. 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.
16. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for real-time style transfer and super-resolution. In Proceedings of the IEEE European Conference on Computer Vision, Amsterdam, The Netherlands, 11–14 October 2016; pp. 694–711.
17. Zheng, H.; Ji, M.; Wang, H.; Liu, Y.; Fang, L. Crossnet: An end-to-end reference-based super resolution network using cross-scale warping. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 88–104.
18. Yang, F.; Yang, H.; Fu, J.; Lu, H.; Guo, B. Learning texture transformer network for image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 19–24 June 2020; pp. 5791–5800.
19. Zhang, Z.; Wang, Z.; Lin, Z.; Qi, H. Image super-resolution by neural texture transfer. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Long Beach, CA, USA, 16–20 June 2019; pp. 7982–7991.

20. Yao, C.; Zhang, S.; Yang, M.; Liu, M.; Qi, J. Depth super-resolution by texture-depth transformer. In Proceedings of the International Conference on Multimedia and Expo (ICME), Shenzhen, China, 5–9 July 2021; pp. 1–6.
21. Yue, H.; Sun, X.; Yang, J.; Wu, F. Landmark image super-resolution by retrieving web images. *IEEE Trans. Image Process.* **2013**, *22*, 4865–4878.
22. Meng, Z.; Zhang, J.; Li, X.; Zhang, L. Lightweight Image Super-Resolution Based on Local Interaction of Multi-Scale Features and Global Fusion. *Mathematics* **2022**, *10*, 1096. [CrossRef]
23. Wu, W.; Xu, W.; Zheng, B.; Huang, A.; Yan, C. Learning Local Distribution for Extremely Efficient Single-Image Super-Resolution. *Electronics* **2022**, *11*, 1348. [CrossRef]
24. Yun, J.-S.; Yoo, S.-B. Single image super-resolution with arbitrary magnification based on high-frequency attention network. *Mathematics* **2022**, *10*, 275. [CrossRef]
25. Wang, Z.; Liu, D.; Yang, J.; Han, W.; Huang, T. Deep networks for image super-resolution with sparse prior. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Santiago, Chile, 7–13 December 2015; pp. 370–378.
26. Xie, C.; Liu, Y.; Zeng, W.; Lu, X. An improved method for single image super-resolution based on deep learning. *Signal Image Video Processing* **2019**, *13*, 557–565. [CrossRef]
27. Zhu, M.; Luo, W. Closed-Loop Residual Attention Network for Single Image Super-Resolution. *Electronics* **2022**, *11*, 1112. [CrossRef]
28. 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, 26 June–1 July 2016; pp. 1646–1654.
29. Liu, A.; Li, S.; Chang, Y. Image super-resolution using progressive residual multi-dilated aggregation network. *Signal Image Video Processing* **2022**, *16*, 1271–1279. [CrossRef]
30. Xu, T.; Zhang, P.; Huang, Q.; Zhang, H.; Gan, Z.; Huang, X.; He, X.A. Fine-grained text to image generation with attentional generative adversarial networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA,, 18–23 June 2018.
31. Sajjadi, M.S.; Scholkopf, B.; Hirsch, M. Enhancenet: Single image super-resolution through automated texture synthesis. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4491–4500.
32. Wang, X.; Yu, K.; Wu, S.; Gu, J.; Liu, Y.; Dong, C.; Qiao, Y.; Change Loy, C. Esrgan: Enhanced super-resolution generative adversarial networks. In Proceedings of the European Conference on Computer Vision, Munich, Germany, 8–14 September 2018.
33. Zhang, Y.; Tian, Y.; Kong, Y.; Zhong, B.; Fu, Y. Residual dense network for image super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR), Salt Lake City, UT, USA, 18–23 June 2018; pp. 2472–2481.
34. Chudasama, V.; Upla, K. RSRGAN: Computationally efficient real-world single image super-resolution using generative adversarial network. *Mach. Vis. Appl.* **2021**, *32*, 3. [CrossRef]
35. Lin, R.; Xiao, N. Dual Projection Fusion for Reference-Based Image Super-Resolution. *Sensor* **2022**, *22*, 4119. [CrossRef]
36. Liu, X.; Li, J.; Duan, T.; Li, J.; Wang, Y. DSMA: Reference-Based Image Super-Resolution Method Based on Dual-View Supervised Learning and Multi-Attention Mechanism. *IEEE Access* **2022**, *10*, 54649–54659. [CrossRef]
37. Simonyan, K.; Zisserman, A. Very deep convolutional networks for large-scale image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Columbia, IN, USA, 23–28 June 2014.
38. Lu, L.; Li, W.; Tao, X.; Lu, J.; Jia, J. MASA-SR: Matching acceleration and spatial adaptation for reference-based image super-resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 6368–6377.
39. Georgescu, M.-I.; Ionescu, R.T.; Miron, A.-I.; Savencu, O.; Ristea, N.-C.; Verga, N.; Khan, F.S. Multimodal Multi-Head Convolutional Attention with Various Kernel Sizes for Medical Image Super-Resolution. *arXiv* **2022**, preprint. arXiv:2204.04218.
40. Shang, T.; Dai, Q.; Zhu, S.; Yang, T.; Guo, Y. Perceptual extreme super-resolution network with receptive field block. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 440–441.
41. Liu, H.; Cao, F.; Wen, C.; Zhang, Q. Lightweight multi-scale residual networks with attention for image super-resolution. *Knowl. Based Syst.* **2020**, *203*, 106103. [CrossRef]
42. Chen, L.-C.; Papandreou, G.; Kokkinos, I.; Murphy, K.; Yuille, A.L. Deeplab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs. *IEEE Trans. Pattern Anal. Mach. Intell.* **2017**, *40*, 834–848. [CrossRef]
43. Lin, R.; Xiao, N. Residual Channel Attention Connection Network for Reference-based Image Super-resolution. In Proceedings of the 2021 8th International Conference on Information, Cybernetics, and Computational Social Systems (ICCSS), Beijing, China, 10–12 December 2021; pp. 307–313.
44. Duan, C.; Xiao, N. Parallax-based spatial and channel attention for stereo image super-resolution. *IEEE Access* **2019**, *7*, 183672–183679. [CrossRef]
45. Zhang, Y.; Li, K.; Li, K.; Fu, Y. Mr image super-resolution with squeeze and excitation reasoning attention network. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Nashville, TN, USA, 20–25 June 2021; pp. 13425–13434.
46. Gulrajani, I.; Ahmed, F.; Arjovsky, M.; Dumoulin, V.; Courville, A.C. Improved training of wasserstein gans. *Adv. Neural Inf. Process. Syst.* **2017**, *30*. [CrossRef]
47. Huang, J.-B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 7–12 June 2015; pp. 5197–5206.

48. Sun, L.; Hays, J. Super-resolution from internet-scale scene matching. In Proceedings of the IEEE International Conference on Computational Photography (ICCP), Seattle, WA, USA, 28–29 April 2012; pp. 1–12.
49. 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, 26 June–1 July 2016; pp. 1874–1883.
50. Cheng, M.; Yongming, R.; Yean, C.; Ce, C.; Lu, J. Structure-Preserving Super Resolution with Gradient Guidance. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 19–24 June 2020; pp. 7769–7778.
51. Schulter, S.; Leistner, C.; Bischof, H. Fast and accurate image upscaling with super-resolution forests. In Proceedings of the Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3791–3799.