An Image Reconstruction Algorithm Based on Frequency Domain for Deep Subcooling of Melt Drops

Keqing Ning[1,2], Ze Su[2], Zhihao Zhang[2], Gwang-jun Kim[3]

[1] Institute of Photoelectronics Technology, School of Science Beijing Jiaotong University, China
[2] School of Computer Science and Technology, North China University of Technology, China
[3] Dept. of Computer Engineering, Chonnam National University, Korea
ningkq@ncut.edu.cn, su_ze28@163.com, mr_zhangzhihao@163.com, kgj@chonnam.ac.kr

Abstract

In a space environment, liquid alloys are in a thermodynamically metastable state, which facilitates research on the material structure and thermophysical properties of deep undercooling melt. Limited by the cost and technology of performing experiments in space, researchers developed electrostatic levitation that uses a drop pipe device to simulate the space environment. A high-speed camera was used to capture the falling image of the deep undercooling melt to study the melting and solidification process. Due to the exposure time and hardware limitations of the image acquisition equipment, the image resolution of the deep undercooling melt is low, which is not conducive to studying the thermophysical properties and solidification interface of the melt. Software design methods, such as super-resolution reconstruction, can more accurately reconstruct image contour information and effectively improve the image resolution. The most current deep learning-based super-resolution reconstruction algorithms directly perform Y-channel or Y, Cb, and Cr three-channel learning on the reconstructed image. This is insufficient in terms of providing more prior information to solve the super-resolution reconstruction. In this study, a single-frame image super resolution reconstruction network that is based on frequency-domain feature learning is proposed. It builds a time–frequency transformation layer at the front end of the neural network and uses the frequency to realize the neural network in the frequency domain. To evaluate the super-resolution reconstruction performance, the proposed algorithm is compared with the current mainstream interpolation, sparse coding, super resolution convolutional neural network, and enhanced single-image super-resolution deep residual algorithms. The proposed algorithm achieves good reconstruction effects on deep undercooled melt images in terms of the objective evaluation and visual perception. At the same time, the peak signal-to-noise ratio and structural similarity index measure achieved results that exceed the aforementioned comparison algorithms.

1 Introduction

The degree of undercooling refers to the difference between the theoretical crystallization temperature of a substance and the actual crystallization temperature. In materials science and the related fields, researchers often use methods such as inhibiting nucleation to keep the alloy liquid in a liquid state under extreme undercooling. This is called deep undercooling. The application of deep undercooling technology can lead to the development of new materials with excellent properties that are difficult to obtain under general solidification conditions. Liquid metal deep undercooling technology is an important means of modern solidification theory research. At the same time, because the liquid alloy is in a thermodynamically metastable state in the space environment, such an environment is very helpful to study the material structure and thermophysical properties of the deep undercooled melt. Further research on the microstructure of liquid alloys is of great significance for the deep understanding of their macroscopic properties, preparation, liquid–solid phase transition, and crystal growth; it is also an important research topic in space materials science [1-3]. To study the characteristics of the melt in the deep undercooled state, researchers simulated the microgravity space environment using a vacuum drop tube on the ground. A levitation device was installed at the top of the drop tube. When the melted sample freefalls in a 50 m drop tube and reaches the bottom of the drop tube, its falling speed is 31.3 m/s. The experimental equipment consists of a high-speed camera with a high frame rate of 106 frames/s and a low resolution for imaging. The number of pixels that are imaged during each exposure time is usually in the range 48–120, and the image

quality is low. Therefore, it is impossible to accurately measure the edge of the image and the texture information inside the droplet. As a result, image super-resolution technology is needed to reconstruct the low-imaging melt image into a high-resolution melt image; subsequently, the diameter of the droplet can be calculated. Changes in the physical parameters such as the sample density, specific heat, surface tension, and viscosity have also been reported [4].

In recent years, with the development of computer technology and the evolution of related algorithms, super-resolution reconstruction algorithms that are based on deep learning have become mainstream and they have achieved better results than traditional algorithms [5]. The main concept is to form a training set that is based on a large number of low–high-resolution image pairs. To obtain super-resolution reconstructed images, we designed a neural network model to characterize the mapping relationship between the low- and high-resolution images. In 2014, Dong et al. [6] used convolutional neural networks to construct an end-to-end mapping system for the first time; they combined sparse coding knowledge to explain this. They also obtained the leading reconstruction effect at the time and called their model the super resolution convolutional neural network (SRCNN). SRCNN needs to upsample the low-resolution images to the target size for feature learning. In 2016, Shi et al. [7] proposed the addition of a sub-pixel convolutional layer to the network. With this addition, the algorithm avoids the problem of the SRCNN's high computational complexity and improves its computational efficiency. Kim et al. [8] considered the network structure and introduced a residual network into the model; moreover, they increased the network depth to 20 layers. At the same time, because the image features that were obtained by residual learning are relatively sparse, the convergence speed of the model training improved. A generative adversarial network (GAN) is an important branch of convolutional neural networks (CNNs). Based on GAN, Ledig et al. proposed SRGAN [9]; they added perceptual and adversarial losses to the model test. The addition improved the realism of the reconstructed images. The results showed that SRGAN can generate more textural details and achieve better results in terms of visual effects. At the same time, in the field of video super-resolution reconstruction, Caballero et al. proposed VEPCN [10], which is based on the spatio-temporal information of multi-frame images; the model satisfied the real-time requirements of the reconstruction.

Presently, most researchers that specialize in deep learning have focused on the intermediate structure of the network; they have increased the depth and width of the network, or have changed the network structure [7, 11-12]. Although this improves the image reconstruction effect, it also increases the complexity of the model and requires considerable computations and time. However, because the super-resolution reconstruction algorithm is an uncertain problem, obtaining more prior information is a key issue. Owing to the end-to-end characteristics of the neural network, the input often consists of a three-channel RGB/YCbCr. When we start to train the network model, the characteristic information that is accepted by the network is limited; a priori data cannot be extracted from the existing data information. These restrictions make it difficult to reconstruct the images accurately. Therefore, this study considered the two ends of the neural network, especially the input end of the research and transformation. The literature [13] used frequency as the input data to realize image classification and semantic segmentation. This inspired us to extract more frequency domain features to realize single-frame image super resolution (SR) reconstruction. Thus, this study proposes a single-frame image SR reconstruction network that is based on frequency-domain features.

First, the network with the frequency domain feature extraction module performs discrete cosine transform (DCT) operations on the image to realize the time–frequency domain conversion of the image. Second, the network arranges and reorganizes the acquired frequency-domain features to refine the shallow features in model learning. Subsequently, the multi-channel feature selection module adopts an adaptive learning method for weighting and matching the features of each channel to characterize the corresponding effects of the different feature channels on the image composition and the relationship between the features of the same layer. At the same time, long and short hop connections are used in the network to deepen the network depth.

The contributions of this investigation are as follows.

1. This study uses the frequency domain analysis method to perform a DCT transformation on the reconstructed image to realize the time–frequency domain conversion. The image feature channels were reorganized and reshaped. In comparison to the time-domain analysis method, the input feature channel was added at the front of the network. The extraction of shallow features is more refined, which improves the utilization of the shallow features.

2. This study constructs a dynamic feature selection model and provides different coefficients to features that are at the same level through adaptive learning. It characterizes the relationship between the features at the same level and the importance of the image composition.

3. The algorithm that is adopted in this study incorporates a long–short-hop connection. Every ten feature selection modules can create short jump connections. It makes a long jump connection before and after the feature selection module at the beginning and the end. This can effectively increase the depth of the network so that the model can use more abstract deep features to characterize the mapping relationship

between the images.

2 Frequency Domain Feature Learning Algorithm

This study proposes a super-resolution reconstruction algorithm that is based on frequency-domain feature learning from the perspective of the image frequency domain. It provides more shallow features at the front end of the network through a time-frequency domain transformation and enhances the constraints. Simultaneously, a multi-channel feature extraction and selection model was built in the middle of the network. Finally, a super-resolution reconstructed image was obtained after upsampling and reconstruction. The block diagram of the network structure of this algorithm is shown in Figure 1.
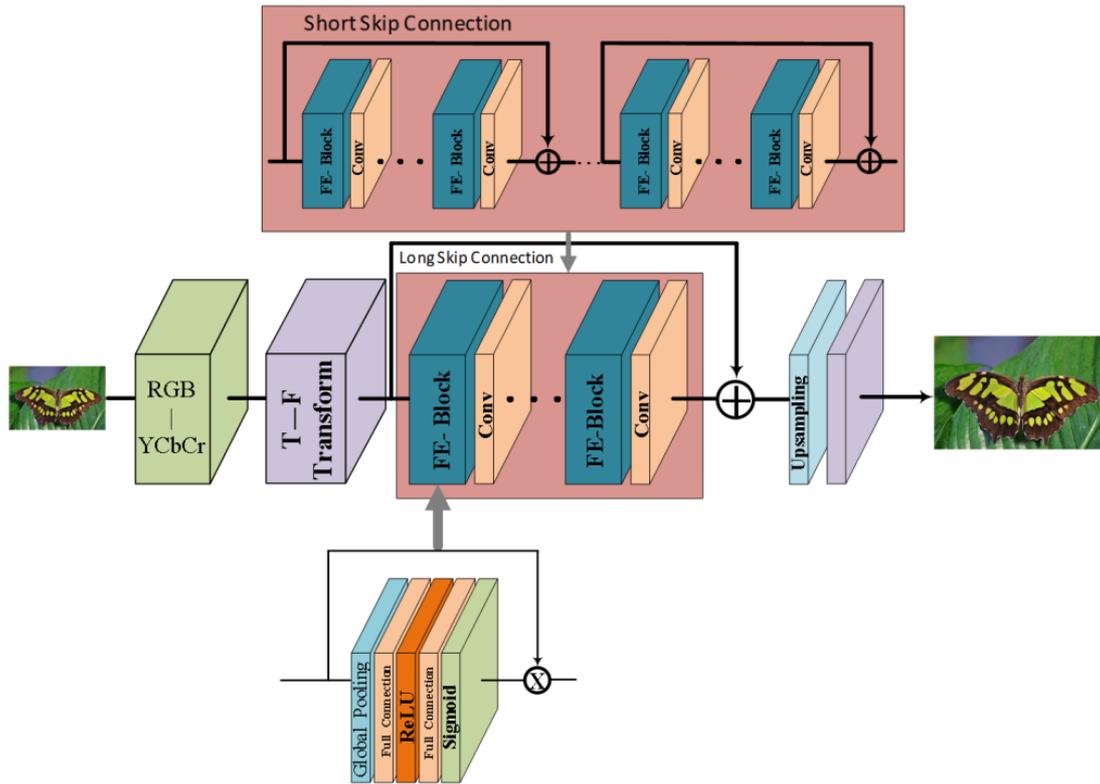

Figure 1. Super-resolution reconstruction algorithm that is based on frequency domain feature learning

2.1 Extraction of the Frequency Features

Li et al. [14] proposed a method that combines the image transform domain with a CNN. They converted the input image to the frequency domain using a Fourier transform; the Fourier coefficients were used as the input of the next layer of the network. They believed that the convolution of the network filter and the spatial image is equivalent to the multiplication of the filter and the Fourier coefficients of the corresponding image.

This accelerates the training and inference processes of the network. However, the experimental results show that the aforementioned operation has not achieved optimal spatial learning method results. In fact, commonly used neural networks generally process the RGB images, and most of these images are obtained through JPEG encoding. From this perspective, this study uses the JPEG encoding process to perform DCT [15-16], complete the time–frequency domain transformation of the image, and then extract the frequency domain features of the image.

In general, when considering super-resolution reconstruction algorithms that are based on deep learning, there are two commonly used spatial learning methods. One is to feed the RGB image directly into the model to reconstruct the image. The other is to transform the image into the RGB–YCbCr color space to form three sets of characteristic channels of Y, Cb, and Cr, which increases the prior information and enhances the constraint conditions. In the second reconstruction method, the channel separation operation of the image is one of the steps of JPEG encoding the RGB image (the general JPEG encoding step is shown in Figure 2), and the further operation is the DCT transformation that can obtain the frequency coefficient. During image composition, different frequency information play different roles. The main portion of the image, the basic gray part, comprises low-frequency information; it has a relatively small decisive effect on the image structure. The intermediate frequency information of the image reflects the basic

structure of the image, especially the edge structure. The high-frequency information of the image mainly reflects the texture and details of the image, and it has a further strengthening effect on the image content compared to the intermediate frequency information. This study utilizes the DCT transform to perform a time–frequency domain conversion on the image. After obtaining more sets of frequency features, it directly performs model learning on them. It not only reflects the important role of frequency in image composition, but also obtains more characteristic channels; thus, it increases the constraint conditions and solves the extremely ill-posed problem of SR reconstruction.
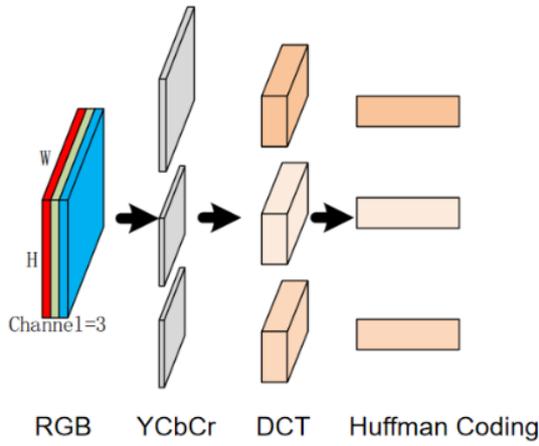
Figure 2. JPEG encoding process

For any image I, H represents the length and W represents the width. When DCT is performed on the image in sequence, blocks that have a size of $N \times N$ along with $\frac{H}{N} \times \frac{W}{N}$ DCT coefficient matrices can be obtained. It is assumed here that H and W are integers of N (the value of N in JPEG coding is usually eight). For the (m, n)th block, the DCT coefficient $C_{m,n}(k_1,k_2)$ can be calculated using Equation (1).

$$C_{m,n}(k_1,k_2) = \sum_{n_2=0}^{N-1}\sum_{n_1=0}^{N-1} I_{m,n}(n_1,n_2) \times w_{k_1,k_2}^{DCT}(n_1,n_2) \quad (1)$$

Here, k_1 and k_2 indicate that the DCT coefficients are in the k_1 th row and k_2 th column of the block, respectively. $w_{k_1,k_2}^{DCT}(n_1,n_2)$ is the basis function of the DCT transform. In particular, the basis function of the two-dimensional DCT transform can be expressed as follows.

$$w_{k_1,k_2}^{DCT}(n_1,n_2) = L_{k_1,k_2}\cos\cos\left[\frac{\pi}{N}(n_1+\frac{1}{2})k_1\right]$$
$$\cos\cos\left[\frac{\pi}{N}(n_2+\frac{1}{2})k_2\right] \quad (2)$$

Here, $L_{k_1,k_2} = \frac{\sqrt{1+\delta_{k_1}}\sqrt{1+\delta_{k_2}}}{N}$. The classic DCT transform is usually applied to orthogonal images, and the basis functions are pairwise orthogonal, which constitutes the orthogonal basis family. Figure 3(a) shows a grayscale image of the orthogonal base family after the aforementioned DCT transformation when $N=8$; an orthogonal dictionary heat map is shown in Figure 3(b).
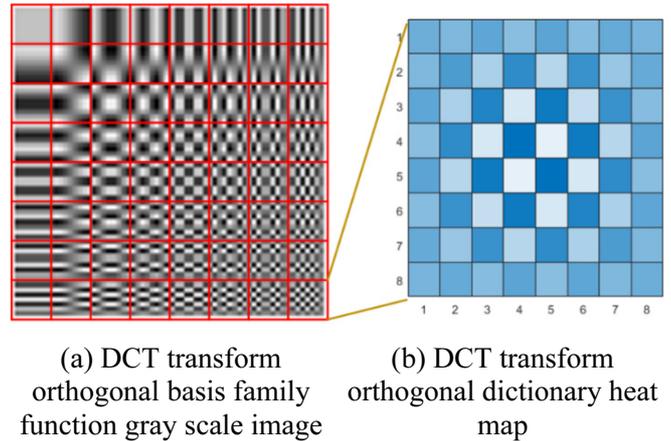
(a) DCT transform orthogonal basis family function gray scale image

(b) DCT transform orthogonal dictionary heat map

Figure 3. DCT conversion Correspondingly, the inverse DCT can be expressed as follows.

$$C_{m,n}(n_1,n_2) = \sum_{n_2=0}^{N-1}\sum_{n_1=0}^{N-1} w_{k_1,k_2}^{DCT}(n_1,n_2) \quad (3)$$

In the specific operation, we first use Equation (4) (k_r, k_g, and k_b is the weighting factor, that is defined in the ITU-R BT.601 standard where $k_r = 0.299$, $k_g = 0.587$, and $k_b = 0.114$) to transform the image into the YCbCr space.

$$Y = k_r * R + k_g * G + k_b * BCb$$
$$= B - Y\, Cr = R - Y\, k_r + k_g + k_b = 1 \quad (4)$$

Then, we can use Equation (1) to perform the DCT transformation on the images that belong to the three channels. Thus, we completed the time–frequency domain conversion of the image and obtained a feature map that characterizes the image with the DCT frequency coefficients. As shown in the next section, we rearranged and reorganized these features.

2.2 Feature Channel Arrangement and Reorganization

Because different frequencies have different effects on the composition of the image structure, further processing is needed for the frequency domain characteristics of the image that are obtained after the preliminary DCT transformation (specifically in the form of two-dimensional DCT grouping coefficients). This is done to reflect the importance of the different

frequency characteristics on the image composition.

The standard JPEG image encoding utilizes a DCT transform base function that is $N \times N$ in the YCbCr color space. Thus, we also applied the same block size in the operation of the previous section. As shown in Figure 4, the specific operation assigns frequency coefficients with the same relative position in all the $N \times N$ blocks to the same feature channel; a frequency feature of the image is represented by this, while maintaining the spatial relationship of each frequency coefficient. Thus, each color space of Y, Cb, and Cr can provide $N \times N$ feature channels, and there are a total of $N \times N \times C$ feature channels in the frequency domain. Assuming that the shape of the original RGB input image is $H \times W \times C$, where $C = 3$ represents the image color space, H represents the length of the image, and W represents the width of the image. After the frequency domain conversion and the feature channel arrangement and recombination, the overall image input feature shape becomes a three-dimensional cube that has a size of $\dfrac{H}{N} \times \dfrac{W}{N} \times C \times N \times N$. This shows that in comparison to the method that is based on spatial learning, after the different DCT coefficients are regrouped and arranged, the number of feature maps is increased from C to $C \times N \times N$. This makes the preliminary shallow feature vector classification more refined and enhances the constraints on the image reconstruction. Because this operation is a rearrangement of the frequency characteristic information, the input data volume maintains the same size as that which is based on spatial domain learning, and the utilization of the image information improves.
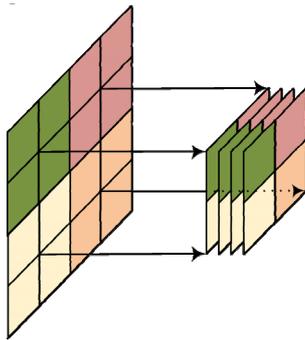

Figure 4. Schematic diagram of the channel recombination of the characteristic coefficients

2.3 Feature Channel Selection Model

After arranging and reorganizing the frequency coefficients in the previous section, we obtained a set of feature maps with different frequency vectors as the distinguishing standard. The previous article explained that the high, medium, and low frequency spectrums constitute the various structural parts of the image, and their importance is different. At the same time, it has been demonstrated in the literature [13] that in

computer vision tasks such as image classification, target recognition, and semantic segmentation, some frequency characteristic channels can only provide less information for the subsequent learning model. By contrast, it increases the number of calculations of the model operation. Therefore, constructing an effective selection mechanism for these characteristic channels is helpful for building the reconstruction model and reducing the number of model calculations. The feature channel selection mechanism has the following two approaches: static and dynamic selection mechanisms.

The static selection method utilizes mathematical methods to calculate the statistical feature information of each feature channel, it manually sets a gating function, and each feature channel is compared with the threshold set by the gating function. When the overall feature of the feature channel is greater than the threshold set by the gating function, the weight that is obtained by this feature channel is 1; otherwise, it is assigned to be 0, which can be expressed as follows.

$$F_n^{out} = \{F_n^{in}, M \geq G\ 0, M < G\} \tag{5}$$

$$M = C_M(F_n^{in}) \tag{6}$$

Here, G represents the threshold set by the gating function, M represents the mathematical statistical feature of the feature channel, F_n^{out} represents the input feature channel, and F_n^{in} represents the output feature channel.

The purpose of the dynamic selection mechanism is to dynamically adjust the weight coefficients of each feature channel during the model training process through the front and back feedback mechanisms. This is done to reflect the importance of the different frequency features in the image structure. Based on the aforementioned concept, this study proposes a learning-based feature channel selection module that adaptively learns and adjusts the weight coefficients that are obtained by each frequency feature. Figure 5 shows a schematic diagram of the feature channel selection module based on the learning strategy that is proposed in this study.
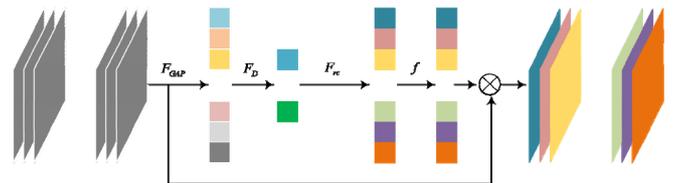

Figure 5. Frequency domain feature selection module

As shown in Figure 6, for n groups of feature maps, if the image feature map is marked as $X = \{x_1, x_2, \ldots, x_{n-1}, x_n\}$, where n represents the number of frequency feature channels, we can calculate the statistical features of the n feature maps (the statistical feature that is applied in this study is the

average value) y_k, as shown in Equation (7). The statistical features of each feature map were then obtained. To show the relationship between each feature map, the dimensionality reduction operation F_D is used to concentrate the statistical features from n y_k to n* y_k^*. Based on the purpose of assigning weight coefficients to each feature channel, we performed a dimension expansion operation F_{rc} on all the statistical features y_k^* to restore the number of coefficients to n, that is, the weight coefficient Raf of each feature channel. The weight coefficient Ra is multiplied by the corresponding feature channels to reflect the relationship between each feature channel and dynamically adjust the degree of influence of each feature channel on the image reconstruction.

$$y_k = F_{GAP}(x_k) = \frac{1}{H'*W'} \sum_{i=1}^{H'} \sum_{j=1}^{W'} x_k(i, j), k \leq n \quad \textbf{(7)}$$

In a convolutional neural network, the weight coefficient can be obtained by the following operation.

$$Ra = f(F_{rc}(F_D(y_k))) = Sigmoid(Z_2 \cdot Relu \cdot Z_1 \cdot y_k) \quad \textbf{(8)}$$

Here, y_k can be obtained by the global average pooling layer, and Z_1 and Z_2 represent the dimensionality reduction and dimensionality upgrade operations, which can be obtained by the fully connected layers with different sizes. Finally, the *Sigmoid* function is used to normalize the value range of each coefficient that is between 0 to 1.
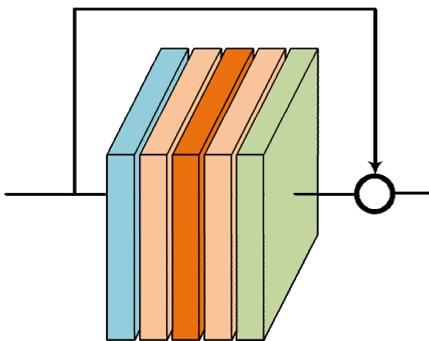

Figure 6. Implementation of the feature selection in the convolutional neural network

2.4 Extraction of Deep Feature Information

The feature channel selection module represents the relationship between the feature maps in the same layer of the network and the difference in the image composition. However, these feature maps only represent shallow feature information, which is insufficient to abstractly express the more accurate nonlinear mapping relationship between the high-resolution images. Increasing the network depth is an important way to solve this problem. However, simply increasing the depth of the network can easily cause

the gradient to disappear and explode. At the same time, if too many regularization layers are added to the network, it will cause model degradation, which indicates that there is oversaturation of the training model. He et al. [17] proposed the idea of residual learning and successfully solved the aforementione problems using local jump connections.

Based on this, the study used the long–short-skip connection method in the Unet [18] network to connect the feature selection modules in a sequence. As shown in Figure 7, "long-skip" refers to the part between the front end of the first feature selection module and the back end of the last feature selection module. "Short-skip" refers to the use of a skip connection for every e feature selection module, so that E feature selection modules form a group of "short-skip" connections where g = E/e (E is an integer multiple of e). After the long–short-skip connection operation, the feature selection module can be effectively superimposed, which increases the network depth and extracts more abstract deep-level feature information.
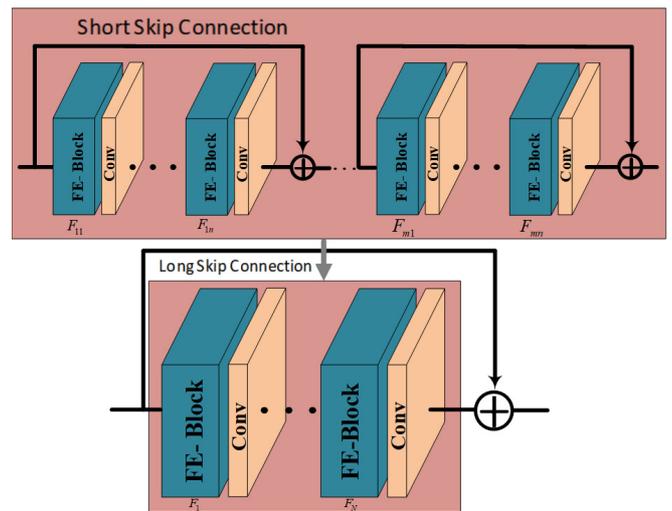

Figure 7. Network structure diagram of the long skip connection and short skip connection between the feature selection modules

3 Experimental Results and Analysis

3.1 Deep Undercooled Melt Datasets

The datasets in this study are taken as a high-resolution image of the melting process of a deep undercooled melt in a vacuum levitation device. The vacuum levitation device is shown in Figure 8(a), and the deep undercooled melt image is illustrated in Figure 8(b). The image is sequentially down-sampled two, three, and four times to simulate the image degradation process; the resulting images are used as low-resolution images. Subsequently, deep undercooled melt image datasets were constructed. The overall flow of the algorithm is illustrated in Figure 9.
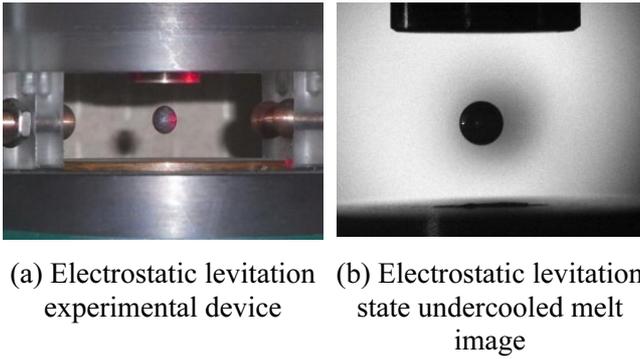
(a) Electrostatic levitation experimental device

(b) Electrostatic levitation state undercooled melt image

Figure 8. Electrostatic levitation device and the "static" undercooled melt image

---

Super-resolution reconstruction algorithm based on frequency domain feature learning

Input: Image to be reconstructed, learning rate = 0.0001, batch size = 16, optimizer: ADAM, loss: L1 norm1

1. **Train:**
2. **For** epoch ≤ 300 **do**
3. Image to be reconstructed: I_l RGB→YCbCr converted: I_l'
4. Use the DCT transform to complete the time-frequency domain conversion, feature shaping, and recombination
5. Go through the feature selection module that is connected by long and short hops, upsampling, and reconstruction layer in turn, and perform forward propagation
6. The reconstructed image and the high-resolution image use the L1 norm to calculate the error
7. Backpropagation to update the network weights
8. **end for**
9. **End train**

Output: Super-resolution reconstructed image

Figure 9. Block diagram of the overall network model that is based on the frequency domain feature learning algorithm

3.2 Comparison and Analysis of the Results

This section presents an analysis of the experimental results. For comparison with existing models, the selected ideas and corresponding algorithms include the following: the bicubic interpolation algorithm that was one of the earliest proposed models (BICUBIC), the traditional field theory development, which is relatively complete, the classic sparse coding algorithm [16] (ScSR), the pioneering work of the super-resolution reconstruction of the images that use deep learning methods, SRCNN [6], and the currently enhanced single-image super-resolution deep residual network algorithm (EDSR) [12] with improved reconstruction effects. At the same time, to facilitate the subsequent display and comparison, the super-resolution reconstruction algorithm that is based on frequency domain feature learning that is proposed in this study is called FSSR.

The comparative experimental analysis is divided into an objective index evaluation, subjective feeling, and task-based evaluation [19-20]. In the objective index evaluation, the PSNR and SSIM [21-22] of each reconstructed image are calculated, and then the average value is taken to characterize the reconstruction effect of the algorithm on the entire data set. In the subjective perception part, one is selected from the image that is reconstructed two, three, and four times, and the sensory evaluation of the same image from each algorithm is compared. Based on the task-based evaluation, pixel scanning was used to calculate the error between the reconstructed image of each algorithm and the high-resolution image, and the relative error reduction of each algorithm.

Table 1. PSNR (dB) value of the deep undercooled melt image that is reconstructed two, three, and four times

PSNR	Scale	BICUBIC	ScSR	SRCNN	EDSR	FSSR
Deep undercooled melts datasets	X2	35.94	37.76	38.03	39.24	**39.82**
	X3	32.27	33.36	34.75	35.61	**36.14**
	X4	30.79	32.37	32.84	33.46	**34.05**

Table 1 shows the PSNR (dB) values that are reconstructed two, three, and four times for the deep undercooled melt image when using different algorithms. The best results are expressed in the bold font (the same below). It can be observed from the table that the FSSR algorithms that are proposed in this study achieved the best results. Moreover, the reconstruction results that use deep learning methods (SRCNN, EDSR, FSSR) surpassed the traditional algorithms. In particular, as far as this algorithm is concerned, it exceeds the second-place EDSR algorithm by 0.58 dB when it is reconstructed two times; it exceeds the second-place EDSR algorithm by 0.53 dB when it is reconstructed three times; and it exceeds the second-place EDSR algorithm by 0.59 dB when it is reconstructed four times. It verifies that the reconstruction effect of this algorithm is better in terms of comparing the pixel-level.

Table 2. SSIM value of the deep undercooled melt image after being reconstructed two, three, and four times

| SSIM	Scale	BICUBIC	ScSR	SRCNN	EDSR	FSSR
Deep undercooled melts datasets	X2	0.9303	0.9562	0.9624	0.9647	**0.9657**
	X3	0.8671	0.8774	0.9127	0.9158	**0.9173**
	X4	0.8246	0.8359	0.8841	0.8964	**0.9053**

Table 2 shows the SSIM values for the deep undercooled melt image that is reconstructed two, three, and four times when different algorithms are used. It can be observed from the table that the FSSR algorithms achieved the best results. In comparison to the second-ranked EDSR algorithm, the results of this algorithm improved by 0.001, 0.0015, and 0.0089 when it was reconstructed two, three, and four times, respectively. This shows that the reconstructed image that uses this algorithm is more similar to the original image structure, and the image reconstruction quality is higher.

The relationship between the multiple reconstructions and the reconstruction algorithm is shown in Figure 10. It can be observed that regardless of the PSNR or SSIM, each algorithm achieved the best results when it was reconstructed twice. At the same time, as the reconstruction multiple increases, the reconstruction effect decreases to varying degrees, which indicates that the reconstruction multiple increases. The difficulty of reconstruction also increases, which also tests the stability of the algorithm model. Among them, the BICUBIC algorithm exhibits the most dramatic changes, and the FSSR algorithm has the best performance. The PSNR and SSIM that were obtained by this algorithm after four reconstructions of the deep undercooled melt image exceeded the value when the BICUBIC algorithm was used three times. This shows that the algorithm model greatly improved.

Figure 11, Figure 12, and Figure 13 compare the overall effect and partial magnification n of the reconstructed image along the two-to-four times reconstructed scale. As shown in Figure 11, each

reconstruction algorithm has a better sensory effect at the two-times scale, and the image edges are more natural after the magnification. In Figure 12, the two spots in the figure are used to compare the magnification. In comparison to the other algorithms, where the size and brightness of the spots in this algorithm are closer to the original image, and the other algorithms have different degrees of magnification and they blur or even disappear. In Figure 13, the main part of the local magnification is the upper right edge of the melt; BICUBIC, ScSR, and SRCNN all have ripple spots. The proposed algorithm and the EDSR algorithm remove most of the noise, which is helpful for observing the edge of the image.

To further evaluate the algorithms to reflect the value of the subsequent applications, we compared the performance of the reconstruction algorithms that are based on the task-based evaluation concept. We observed the change in the melting interface, which is important for studying the specific heat of the deep undercooled melt and other image-specific thermal thermophysical parameters. The thermophysical parameters can be obtained by measuring the area of the deep undercooled melt image. We measured the image area in this study by binarizing the image (after the literature review and the related experimental tests, the threshold value that was selected in this study was 0.33) [23]; the pixel scanning method was then used to measure the image area. Figure 14 shows an example of the image binarization. Furthermore, the task-based algorithm evaluation was realized by comparing the regional absolute errors.
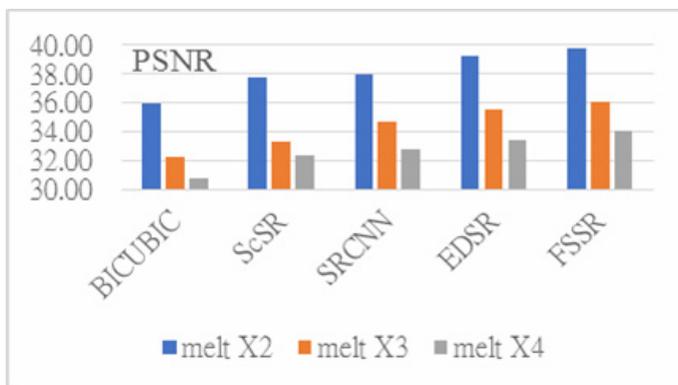
(a) PSNR of the different algorithms when reconstructed two, three, and four times
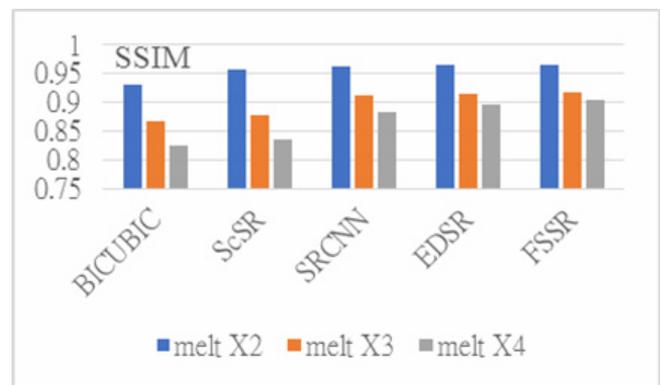
(b) SSIM of the different algorithms when reconstructed two, three, and four

Figure 10. PSNR and SSIM of the reconstructed image of the deep undercooled melt times
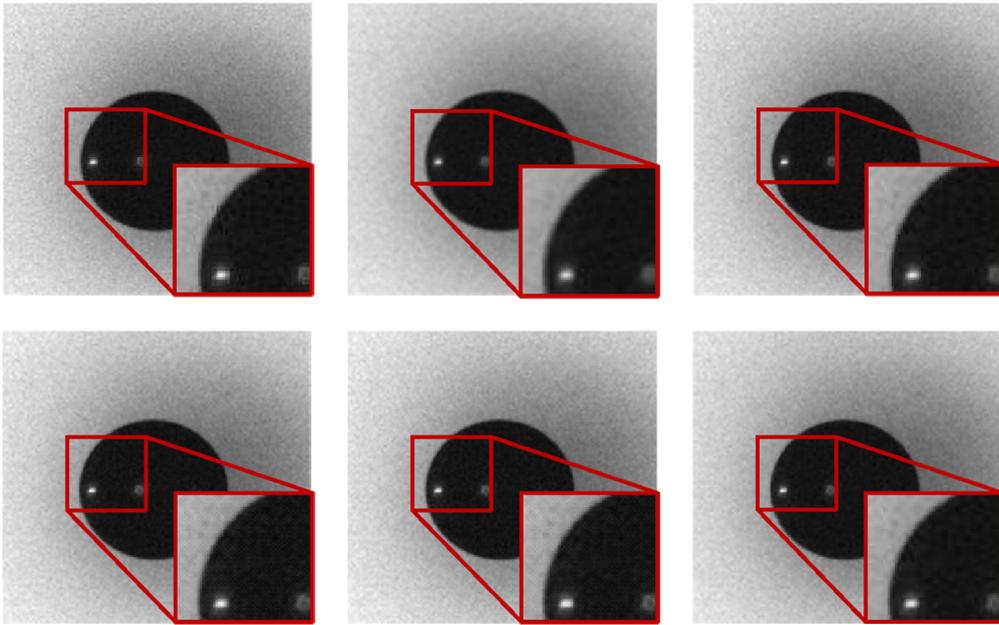
Figure 11. Comparison of the "melt350" image reconstruction effects at the two-times reconstruction scale. The first layer from left to right is: the original image, the BICUBIC reconstructed image, and the ScSR reconstructed image. The second layer from left to right is: the SRCNN reconstructed image, the EDSR reconstructed image, and the FSSR reconstructed image

Figure 12. Comparison of the "melt1" image reconstruction effects at the three-times reconstruction scale. The first layer from left to right is: the original image, the BICUBIC reconstructed image, and the ScSR reconstructed image. The second layer from left to right is: the SRCNN reconstructed image, the EDSR reconstructed image, and the FSSR reconstructed image
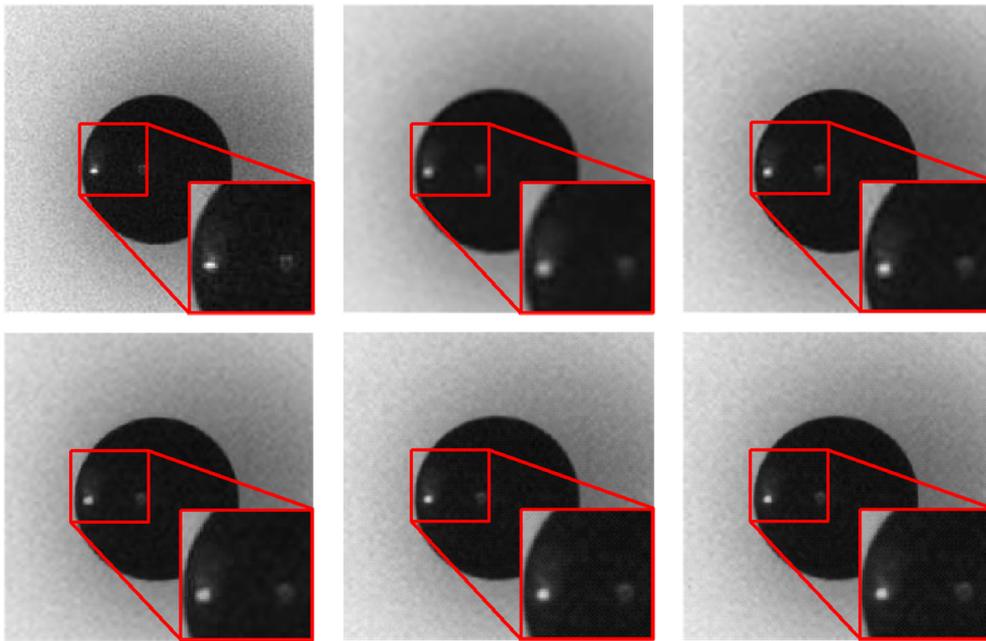
Figure 13. Comparison of the "melt396" image reconstruction effect at the four-times reconstruction scale. The first layer from left to right is: the original image, the BICUBIC reconstructed image, and the ScSR reconstructed image. The second layer from left to right is: the SRCNN reconstructed image, the EDSR reconstructed image, and the FSSR reconstructed image
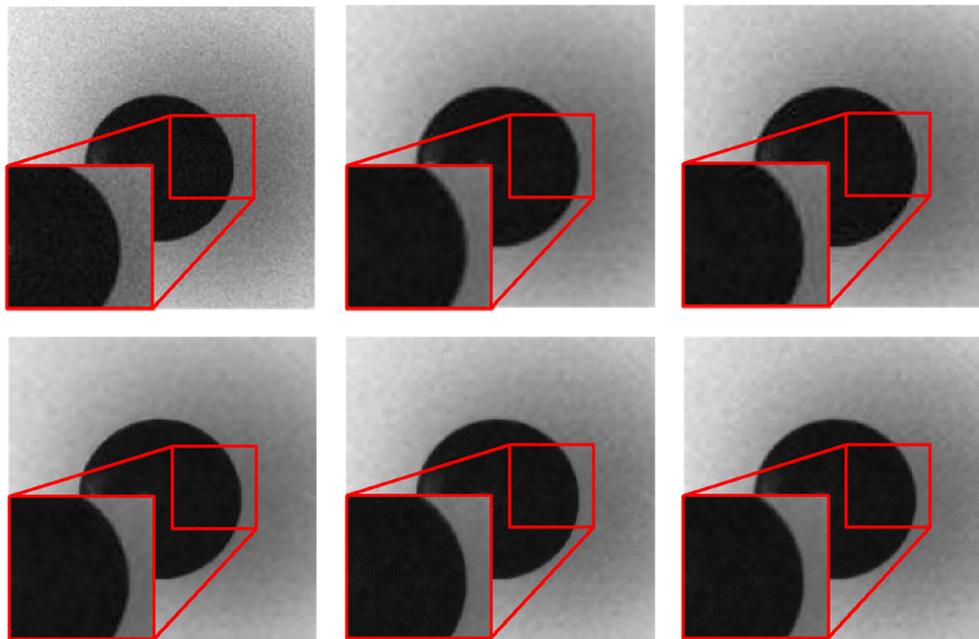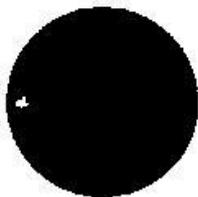

Figure 14. Example of image binarization

Table 3 lists the scanning area of the reconstructed image that was obtained with different algorithms and the absolute errors of the corresponding original image (HR). As shown in the table, the FSSR algorithm that is proposed in this study achieves the best results in the reconstructed images for various multiples, and the range of the absolute error is small for the different reconstruction multiples. This indicates that the stability of the algorithm is higher than the other algorithms.

Table 3. Scanning area and absolute error of the reconstructed image (mm^2)

	Reconstructed two-times		Reconstructed three-times		Reconstructed four-times	
	Scan area	Absolute error (HR=10509)	Scan area	Absolute error (HR=10721)	Scan area	Absolute error (HR=10707)
BICUBIC	10531	22	10854	133	10731	24
ScSR	10502	7	10807	86	10716	9
SRCNN	10502	7	10696	25	10688	19
EDSR	10499	10	10702	19	10682	25
FSSR	10504	**5**	10713	**8**	10703	**4**

At the same time, to further explain the performance difference between the various algorithms in the area error indicator, the error that was obtained by the BICUBIC algorithm can be used as the standard to compare the performance of each algorithm relative to the BICUBIC algorithm. The specific formula can be expressed as follows.

$$P_C = \left(\frac{e_{BI} - e_C}{e_{BI}} \right)\% \qquad (9)$$

Here, e_c is the absolute error of the comparison algorithms (ScSR, SRCNN, EDSR, FSSR), and e_{BI} is the absolute error of the BICUBIC algorithm.

Table 4 shows the performance data that was calculated using Equation (9). Figure 15 is drawn based on the data. It can be observed from these two graphs that the performance of the FSSR algorithm that is proposed in this study achieved the best results for the various reconstruction multiples, and the trend of the change was relatively stable. This indicates that the scope of the application of this algorithm for the reconstruction multiples is better than the other comparison algorithms.

Table 4. Comparison of the performance of each algorithm relative to the BICUBIC algorithm

Reconstruction Multiple	2	3	4
BICUBIC	——	——	——
ScSR	68.18%	35.34%	62.50%
SRCNN	68.18%	81.20%	20.83%
EDSR	54.55%	85.71%	-4.17%
FSSR	**77.27%**	**93.98%**	**83.33%**
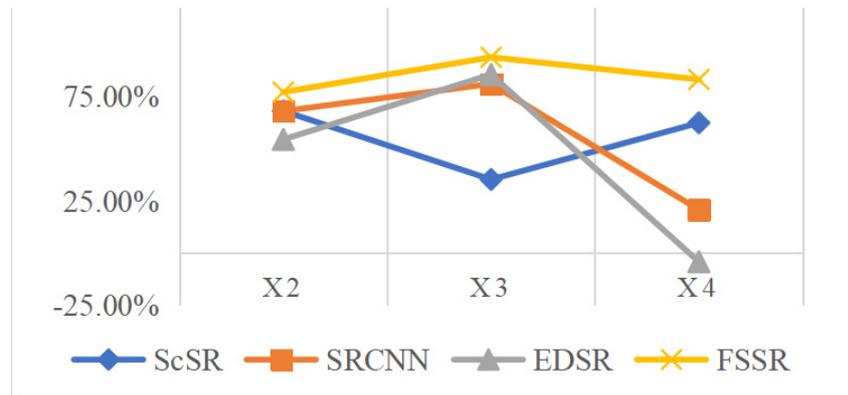
Figure 15. Comparison of the algorithm performance under the different reconstruction multiples (based on the BICUBIC algorithm)

4 Conclusion

In this study, a super-resolution reconstruction model that is based on deep learning was constructed to solve the low-resolution problem of the deep undercooled melt images that are acquired by high-speed cameras. In this study, we investigated the use of image frequency domain features for model learning from the perspective of the frequency on the image structure composition. The method expands the shallow features using the time–frequency domain transformation while arranging and reorganizing these shallow features to refine these features. A feature selection model was used to characterize the interrelationship between the sibling features and their contribution to the image composition. The experimental results show that the proposed method can effectively reconstruct the deep-undercooled melt images. In comparison to the typical super-resolution reconstruction algorithms that are used in the various development stages in this study, it achieved the best results for subjective and objective evaluation and the task-based evaluation. The experimental results on five public datasets also show that this algorithm achieves better results, which further verifies the effectiveness and applicability of this algorithm. In future work, we intend to further optimize the network structure and streamline the parameters; thus, reducing the requirement of excessive hardware equipment and a long training time for model training. It is also necessary to develop more optimized evaluation criteria from the perspective of image composition and practical application scenarios.

References

[1] J. Lee, S. Katamreddy, Y. C. Cho, S. Lee, G. W. Lee, Containerless materials processing for materials science on earth and in space, in: J. Lee, S. Wagstaff, A. Anderson, F. Tesfaye, G. Lambotte, A. Allanore (Eds.), *Materials processing fundamentals*, Springer International Publishing, 2021, pp. 187-199.

[2] L. Cao, R. F. Cochrane, A. M. Mullis, Solidification Morphology and Phase Selection in Drop-Tube Processed Ni–Fe–Si Intermetallics, *Intermetallics*, Vol. 60, pp. 33-44, May, 2015.

[3] K. Y. Kamal, R. Herranz, J. J. W. A. van Loon, P. C. M. Christianen, F. J. Medina, Evaluation of Simulated Microgravity Environments Induced by Diamagnetic Levitation of Plant Cell Suspension Cultures, *Microgravity Science and Technology*, Vol. 28, No. 3, pp. 309-317, June, 2016.

[4] Z. Zou, X. Luo, Q. Yu, Droplet Image Super Resolution Based on Sparse Representation and Kernel Regression, *Microgravity Science and Technology*, Vol. 30, No. 3, pp. 321-329, May, 2018.

[5] J. Li, J. Wang, X. Chen, Z. Luo, Z. Song, Multiple Task-driven Face Detection Based on Super-resolution Pyramid Network, *Journal of Internet Technology*, Vol. 20, No. 4, pp. 1263-1272, July, 2019.

[6] C. Dong, C. C. Loy, K. He, X. Tang, Learning a Deep Convolutional Network for Image Super-Resolution, *European Conference on Computer Vision*, Zurich, Switzerland, 2014, pp. 184-199.

[7] W. Shi, J. Caballero, F. Huszár, J. Totz, A. P. Aitken, R. Bishop, D. Rueckert, Z. Wang, 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*, Las Vegas, NV, USA, 2016, pp. 1874-1883.

[8] 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*, Las Vegas, NV, USA, 2016, pp. 1646-1654.

[9] C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, W. Shi, Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 105-114.

[10] J. Caballero, C. Ledig, A. Aitken, A. Acosta, J. Totz, Z. Wang, W. Shi, Real-Time Video Super-Resolution With Spatio-Temporal Networks and Motion Compensation, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 2017, pp. 2848-2857.

[11] J. Yu, Y. Fan, J. Yang, N. Xu, Z. Wang, X. Wang, T. Huang, *Wide activation for efficient and accurate image super-resolution*, arXiv:1808.08718, December, 2018. [Online]. Available: http://arxiv.org/abs/1808.08718.

[12] B. Lim, S. Son, H. Kim, S. Nah, K. M. Lee, Enhanced Deep Residual Networks for Single Image Super-Resolution, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, 2017, pp. 1132-1140.

[13] K. Xu, M. Qin, F. Sun, Y. Wang, Y. Chen, F. Ren, Learning in the Frequency Domain, *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Seattle, WA, USA, 2020, pp. 1737-1746.

[14] J. Li, S. You, A. Robles-Kelly, A Frequency Domain Neural Network for Fast Image Super-Resolution, *2018 International Joint Conference on Neural Networks (IJCNN)*, Rio de Janeiro, Brazil, 2018, pp. 1-8.

[15] E. Agustsson, R. Timofte, NTIRE 2017 Challenge on Single Image Super-Resolution: Dataset and Study, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops*, Honolulu, HI, USA, 2017, pp. 1122-1131.

[16] J. Yang, J. Wright, T. S. Huang, Y. Ma, Image Super-Resolution as Sparse Representation of Raw Image Patches, *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, Anchorage, Alaska, USA, 2008, pp. 1-8.

[17] K. He, X. Zhang, S. Ren, J. Sun, Deep Residual Learning for Image Recognition, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Las Vegas, NV, USA, 2016, pp. 770-778.

[18] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional Networks for Biomedical Image Segmentation, *International Conference on Medical image computing and computer-assisted intervention*, Munich, Germany, 2015, pp. 234-241.

[19] M. Bevilacqua, A. Roumy, C. Guillemot, M. A. Morel, Low-Complexity Single-Image Super-Resolution Based on Nonnegative Neighbor Embedding, *Proceedings of the British Machine Vision Conference*, Surrey, UK, 2012, pp. 135.1-135.10.

[20] R. Zeyde, M. Elad, M. Protter, On Single Image Scale-Up Using Sparse-Representations, *International conference on curves and surfaces*, Avignon, France, 2010, pp. 711-730.

[21] D. Martin, C. Fowlkes, D. Tal, J. Malik, A Database of Human Segmented Natural Images and Its Application to Evaluating Segmentation Algorithms and Measuring Ecological Statistics, *Proceedings of the Eighth IEEE International Conference on Computer Vision*, Vancouver, BC, Canada, 2001, pp. 416-423.

[22] J. B. Huang, A. Singh, N. Ahuja, Single Image Super-Resolution From Transformed Self-Exemplars, *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Boston, MA, USA, 2015, pp. 5197-5206.

[23] K. I. Matsuo, K. Ueda, M. Umeda, Extraction of character string from scene image by binarizing local target area, *IEEJ Transactions on Electronics, Information and Systems*, Vol. 122, No. 2, pp. 232-241, February, 2002.

Biographies

Keqing Ning received the M.S. degree from Beijing Jiao tong University, Beijing, China, in 2008. He is currently an Assistant Professor with the School of Information, North China University of Technology. He is currently pursuing the Ph.D. degree with Beijing Jiao tong University, Beijing, China. His research interests include image super-resolution, deep Learning and Application of embedded.

Ze Su was born in Beijing, China, in 1995. He recieved the bachelor's degree in Electronic information engineering from North China University of Technology, in 2018. He is currently pursuing the master's degree in Information and communications engineering from North China University of Technology. His research interests include image super resolution reconstruction and pattern recognition.

Zhihao Zhang was born in Shandong, China in 1996. He received his Bachelor's degree in Computer Science and Technology from Qingdao University of Technology in 2020. He is currently studying for his master's degree in computer technology at North China University of Technology. His research interests are deep learning and image super-resolution reconstruction.

Gwang-jun Kim received the B.E., M.E and Ph.D. degrees in computer engineering from Chosun University in 1993,1995 and 2000, respectively. He joined the department of computer engineering, Chonnam National University, in 2003 and became an Associate Professor in 2009. Since 2015, he has been an Professor in computer engineering at Chonnam National University. During 2000-2001, he was a researcher in the department of electrical and computer engineering at university of California, Irvine. His current research interest lie in the area sensor network, IoT, real-time communication and various kinds of communication systems.