International Journal of Speech Technology
https://doi.org/10.1007/s10772-021-09892-2

Accurate recognition of heterogeneous features in super resolution image visualization based on voice remote control system

Xuyang Wang[1] · Shijian Liu[1]

Received: 15 January 2021 / Accepted: 21 August 2021

Abstract
In the voice remote control system, the output of recognition accuracy is unstable due to the insufficient reconstruction of image resolution in the traditional image feature recognition. In order to solve the above problems, this paper proposes an image heterogeneous feature recognition method based on principal component analysis and convolution neural network. According to the shock response in the process of image super-resolution processing, the feasible region of super-resolution image reconstruction is established, and the image is reconstructed in the feasible region to achieve high-resolution image display. Principal component analysis (PCA) is used to reduce the dimension of super-resolution image, and the recognition of heterogeneous features of image is realized in convolution neural network. The performance of the proposed method is verified by comparative experiments. The results show that the accuracy of the heterogeneous feature recognition method is 94.26%, and the average recognition time is 7.8 s.

Keywords Super resolution image · Heterogeneous features · Feature recognition · Convolution neural network · PCA · Dimension reduction

1 Introduction

Super resolution technology refers to the technology of reconstructing corresponding high-resolution images from one or more low resolution images. Super resolution images are widely used, such as medical imaging, high-definition target detection, surveillance video acquisition, image compression and reconstruction, etc. It can provide more useful information for people to analyze and process images, and has high practical significance (Weimin et al., 2020). In the voice remote control system, the multimedia communication can get more vivid and intuitive visual experience through the processing of super-resolution image technology. At present, there are many methods to generate super resolution image, but the generated super resolution image has many visual heterogeneous information due to the influence of many factors in the processing process. Some of these information can help people or computers to process images more accurately. On the other hand, the useless information will affect the accuracy and efficiency of subsequent image information analysis (Ya, 2019), which has great limitations in use. Therefore, it is necessary to accurately identify the visualized heterogeneous features in super resolution images.

At present, some scholars study an image processing method, which uses dense depth convolution neural network and residual network, and extracts the high-frequency information of the original image separately. After up sampling, it is fused with super-resolution image to form high-frequency feature compensation. In short, it is a super-resolution image recognition method using feature compensation (Zilu & Xiang, 2019), but this recognition method is only good at processing and recognizing relatively concrete and single features. It has problems of low accuracy and high time cost for more feature recognition. Therefore, in order to improve the accuracy of super resolution image visualization heterogeneous feature recognition, this paper will study the accurate recognition method of heterogeneous features in super resolution image visualization and verify the feasibility of this method.

✉ Shijian Liu
1215034362@qq.com

[1] College of Computer and Communication, Lanzhou University of Technology, Lanzhou, China

Published online: 18 October 2021

2 Research on accurate recognition method of heterogeneous features in super resolution image visualization

2.1 Reconstruction of super resolution image

Image super-resolution reconstruction refers to the use of single or sequence of multiple low-resolution observation images to reconstruct a high-resolution image. Therefore, according to the different image sources that need to be processed, it can be divided into two categories: single image super-resolution algorithm and sequence image super-resolution algorithm. Single-image super-resolution reconstruction means that when there is only one low-resolution observation image, combined with some prior knowledge of the image, the high-frequency information lost when the image is acquired is restored, and a high-resolution image is reconstructed. Modeling-based methods use low-resolution images and observation models to provide constraints on the reconstruction of high-resolution images by modeling the imaging process of the image.

Image super-resolution reconstruction technology is widely used. The current research on super-resolution reconstruction mainly focuses on the following directions:

Seek a new image observation model. The modeling of the image imaging process is crucial to the image super-resolution reconstruction algorithm. However, since it is difficult to obtain an observation model that meets the actual imaging process, simple observation models are often used for approximation, which will inevitably affect the reconstruction effect. Therefore, it is necessary to construct a new image observation model to reflect the imaging process more accurately and comprehensively.

Due to the influence of many factors, the quality of super resolution image will be degraded. In order to avoid the loss of heterogeneous features in the degraded part of the image, it is necessary to reconstruct the image with high resolution. The following Fig. 1 shows the observation schematic diagram of super resolution image. According to the observation process of image, the constraint conditions of super resolution image reconstruction can be defined (Wei, 2019).

According to the above figure, the observation expression of the image k is as follows,

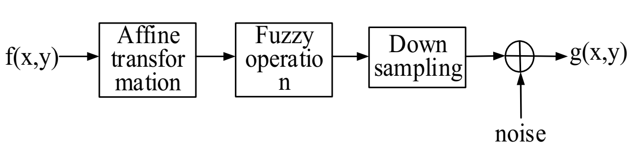

Fig. 1 Schematic diagram of super resolution image observation

$$g_k = \sigma_k\left(b_k\left(T_k(f(x,y))\right) + \eta_k(x,y)\right) \tag{1}$$

In formula (1), g_k is the super resolution image. $\eta_k(x,y)$ is the noise in the image which is usually subject to normal distribution after super resolution technology processing. $f(x,y)$ is the original image without super resolution technology processing. T_k is the affine transformation in the image observation process.b_k is the fuzzy operation in the image observation process. σ_k is the down sampling operation in the image observation process. According to the impact response in the process of image super resolution technology processing (Pengyu et al., 2019; Zhizheng et al., 2018), the feasible region of super resolution image reconstruction is established, and the image is reconstructed within the feasible region.

For the current super resolution image g_k, it can be obtained by the translation, flipping and rotation of image g_1Jiali et al., 2018; Yan et al., 2018). Therefore, the relationship between super resolution image g_k and image g_1 is as follows,

$$g_k(x,y) = g_1\left(x\cos\theta_k - y\sin\theta_k + a_k, y\cos\theta_k + x\sin\theta_k + b_k\right) \tag{2}$$

In the above formula, θ_k is the rotation angle when the image g_1 is rotated to the image g_k. a_k is the horizontal translation amount when the image g_1 is transformed into the image g_k through translation. b_k is the vertical translation amount when the image g_1 is transformed into the image g_k through translation. If the rotation angle of the image θ_k can be ignored, the following objective function can be obtained by processing formula (2) according to Taylor's mathematical principle (Huang et al., 2018).

$$\hat{\theta}_k, \hat{a}_k, \hat{b}_k = \underset{\theta_k, a_k, b_k}{\arg} \min E\left(\theta_k, a_k, b_k\right) \tag{3}$$

In the above formula, the third-order partial derivative can be obtained and the third-order partial derivative is set as 0. The translation motion parameter equation in the process of image transformation is obtained. After obtaining the translation and rotation of image super resolution processing through the above steps, the image is reconstructed according to the following steps.

(1) According to the frame number of the image, the affine transformation parameters between the original image and the super resolution image are calculated (Jingxuan, 2019).
(2) The diffusion function of image pixels is calculated, and the super resolution image is processed by linear interpolation.
(3) The above calculation process is repeated and the pixel value is updated until the set number of image itera-

tions is reached. Then the reconstructed super resolution image is stopped and output.

After the reconstruction of the super resolution image, the PCA dimension reduction algorithm optimized by particle swarm optimization is used to reduce the image dimension.

2.2 Dimension reduction of super resolution image

In both image enhancement and restoration, dimensionality reduction is required. Image enhancement methods have an ideal role in improving image contrast. The enhanced image changes the visual effect of the image, and the applicable method can be made more intuitive, mainly to highlight the details of the image and facilitate the subsequent processing of the image. In the obtained image, the contrast drop is very large. Using appropriate image enhancement methods, the obtained blurred image can be made clear. Traditional image enhancement techniques include histogram equalization and Laplacian sharpening. The histogram equalization method defines the distance between adjacent gray levels in the histogram, introduces a linear transformation based on histogram mean and standard deviation, and discusses the selection of histogram equalization parameters when the image is too bright. The Laplace sharpening method realizes the extraction and preservation of image details through Laplace sharpening, But the traditional image enhancement technology usually ignores the low frequency part of the image, resulting in the loss of part of the image features. The purpose of image restoration technology is to restore the degraded image to the original image before degradation. Therefore, this technology is also called image restoration, and it plays a key role in many image processing applications. The essence of image restoration is to judge and estimate different degradation mechanisms, and find a suitable restoration method for the degradation model. Therefore, the first problem to be solved in the image restoration technology is to find the degradation function. However, it is generally difficult to fully estimate the degradation function, so this process is also called a blind restoration method.

Before image dimension reduction, wavelet transform is used to extract the low frequency part to reduce the feature loss caused by image dimension reduction. The wavelet decomposition algorithm as shown in the Fig. 2 below is used to decompose the super resolution image (Gang, 2019).

In the above figure, I_j is the low-frequency component of the corresponding super resolution image. D_j^1 is the detail information component of the image in the horizontal direction. D_j^2 is the detail information component of the image in the vertical direction. D_j^3 is the detail information component of the image in the diagonal direction. The principle formula of wavelet decomposition is as follows (Yuefei, 2019):

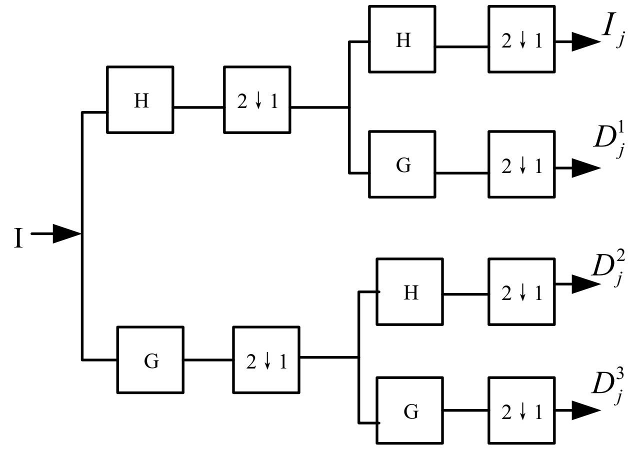

Fig. 2 Structure of wavelet decomposition algorithm

$$WT_s(a,b) = \frac{1}{\sqrt{a}} \int s(t)\psi^*\left(\frac{t-b}{a}\right)dt \tag{4}$$

In formula (4), a is the scale factor of wavelet decomposition, and its numerical value is positive. b is the time shift parameter of wavelet decomposition. $\psi(t)$ is the mother wave of wavelet change. $s(t)$ is the super resolution image sequence without wavelet transform decomposition. The scale factor and time-lapse parameters are discretized by binary sampling point lattice sampling. The discrete orthogonal dyadic wavelet transform after parameter discretization is shown in the following formula (Junli, 2019),

$$W_{j,n} = 2^{-\frac{j}{2}} \int_{-\infty}^{+\infty} s(t)\psi_{j,n}(t)dt \tag{5}$$

According to the above formula, the super resolution image can be decomposed into two dimensions. The two-dimensional scale function of wavelet decomposition can be expressed as the product of two one-dimensional scale functions. The principle of principal component dimension reduction is used.

The principal component is selected according to the pixel value of the image, and the principal component decision matrix is formed. The data correlation coefficient matrix R_i is obtained by moving the window down one window width. Then the matrix R_i is projected into the unitized space of the previous window coefficient matrix H_{i-1}, that is (Yu et al., 2019),

$$\begin{cases} H_{i-1} = \frac{1}{m} X_{i-1} V_{i-1} A_{i-1}^{-1} \\ \overline{R}_i = H_{i-1}^T R_i H_{i-1} \end{cases} \tag{6}$$

In the above formula,X_{i-1} is the window data set. V_{i-1} is the selected principal components with the number of k.

A_{i-1}^{-1} is the selected eigenvalues with the number of k in the previous window. $\overline{R}_i$ is the projection of R_i in space. The particle swarm optimization algorithm is used to optimize the parameters of principal components. The matrix parameters in the dimension reduction of principal component are taken as the initial population of particle swarm optimization algorithm, and the initial position and moving speed of particles are set. The fitness of each particle is calculated and compared, and the particle corresponding to the optimal fitness is selected to update the global extremum. After updating the global optimal parameters, the velocity and position of particles in the particle swarm are replaced by the corresponding positions and velocities of the particles with the current optimal parameters. The above steps are repeated and it is determined that whether the corresponding parameter solution of the optimal particle obtained by the current particle swarm optimization algorithm meets the stopping requirements of the algorithm. If it meets the requirements, the optimal parameters are output. If not, the initial iteration weight of the particle swarm optimization is adjusted and recalculated. The parameters optimized by particle swarm optimization are used to reduce the dimension of super-resolution image, and convolution neural network is used to identify the heterogeneous features in the image.

2.3 Realization of heterogeneous feature recognition

Visual recognition methods mainly include scatter plots, contour plots (parallel coordinate method), and multi-dimensional radial coordinate methods. The contour map maps the n-dimensional data space to a two-dimensional plane with n equidistant parallel axes. Each axis corresponds to an attribute dimension, that is, the abscissa represents each characteristic parameter, and the ordinate represents each characteristic value. Each data sample can be converted into a graphical form, represented by a broken line on n parallel axes. Use the contour map method to visualize all the fault samples in one graph. The feature vectors with the same attributes will be close together and be separable from other categories. Otherwise, it means that the feature parameters have no influence on the fault category. The essence of pattern recognition is classification. The more obvious the clustering characteristics, on this basis, convolution neural network method is introduced, the higher the recognition rate. Using this feature, feature selection is performed through visualized graphical observation. Intuitively remove features with inconspicuous clustering characteristics, and screen out high-quality features with obvious clustering.

In the convolution neural network, the features in super resolution image are sensed by the low layer. The image features are classified and recognized in the subsequent convolution layer and hidden layer. Finally, the corresponding recognition results are output by the output layer of the convolution neural network.

In this paper, the restricted Boltzmann machine is added to the basic convolution neural network to reduce the neurons connected between layers, so as to improve the processing efficiency of the input image data. The energy expression of a single layer confined Boltzmann machine is as follows (Shibao et al., 2018),

$$E(v,h) = -\sum_{i=1}^{n}\sum_{j=1}^{m}\omega_{i,j}h_i v_j - \sum_{j=1}^{m} b_i v_j - \sum_{j=1}^{m} c_i h_i \tag{7}$$

In the above formula, $\omega_{i,j}$ is the connection weight between the visible neuron i in the upper layer and the hidden neuron j in the next layer. h_i is the hidden neuron node in the network. v_j is the visible neuron node in the network. c_i is the element in the hidden layer offset vector. b_i is the element in the visible layer offset vector.

According to the above energy expression, the joint probability distribution function of the restricted Boltzmann machine is obtained, and the conditional probability between the visible operation layer and the hidden layer of the convolution neural network is deduced. The single-layer restricted Boltzmann machine is trained by using the training sample set, so that the output of each layer of restricted Boltzmann machine is approximately equivalent to the original input. After training single-layer restricted Boltzmann machines, several restricted Boltzmann machines are connected in series to form the hidden layer of convolution neural network.

The convolution kernel is used to extract all the features of the image in the convolution layer. Different convolution kernels in the convolution layer can extract different kinds of features of the image. The image in the input convolution layer is convoluted by the convolution kernel to obtain the local induction domain. It is difficult to define the boundary of local sensing domain. By mapping the local sensing domain under the calculation of sigmoid function, the nonlinear feature extraction of super-resolution image is realized.

The image processed by convolution layer is recognized in pooling layer, and the recognition principle is clustering. Cluster centers are randomly selected to calculate the distance between other image feature vectors and cluster centers. The feature vector is divided into the cluster with the smallest clustering distance. After all the feature vectors are divided, the cluster center is updated. The clustering process is repeated until no new cluster center is generated.

After clustering, the remaining feature vector set is output to the output layer of convolution neural network, and the image feature vector output by the output layer is the recognized heterogeneous feature.

Through the above steps, the research on the accurate recognition method of heterogeneous features in super resolution image visualization is completed. The performance of this method will be tested through experiments.

3 Simulation experiment

The performance of heterogeneous feature recognition in super-resolution image visualization is verified by comparative experiments, and MATLAB is used for simulation experiments.

3.1 Experimental content

In this experiment, from the accuracy and time cost of heterogeneous feature recognition, the effectiveness and feasibility of the proposed method are tested intuitively and effectively by comparing the above method with the traditional method based on feature compensation.

Table 1 Experimental image data parameters

Image number	Size/K	Format	Image number	Size	Format
P1	300	png	P11	580	png
P2	320	jpg	P12	600	jpg
P3	351	png	P13	620	jpg
P4	400	jpg	P14	570	png
P5	350	png	P15	480	jpg
P6	360	png	P16	490	jpg
P7	380	jpg	P17	380	png
P8	420	jpg	P18	400	jpg
P9	550	png	P19	640	png
P10	560	png	P20	720	jpg

3.2 Experimental preparation and process

In this experiment, the images in the image database are selected as the experimental objects, and the images are numbered randomly and input to the computer simulation platform. The image is reconstructed into super-resolution image by using the proposed method and the traditional method. After image reconstruction, the heterogeneous feature information of the image is obtained. The images are numbered randomly and input to the computer simulation platform. The specific parameters of the data set parameters used in the experiment are shown in the Table 1 below.

In the computer simulation platform, the experimental group and the control group methods are run to recognize the experimental images of different types. The time consumption of the two methods to identify the heterogeneous features is recorded, and the recognition time cost is represented by the recognition time-consuming. According to the known heterogeneous feature information of the image, the recognition accuracy of the two feature recognition methods in the recognition process is calculated. The experimental data is processed and analyzed to get the corresponding experimental conclusion and complete the verification process.

3.3 Experimental results

The recognition accuracy and time consumption of the two methods for image texture feature recognition are shown in the following two tables, and the data in the Tables 2, 3 are analyzed.

By comparing and analyzing the data in the above two tables, it can be seen that the recognition accuracy of the experimental group method for different super-resolution images is higher than that of the existing methods. The recognition accuracy of the experimental group is more than 91%. On the premise of high recognition rate, the

Table 2 Recognition accuracy (%) and time consumption (s) of experimental methods

Image number	Recognition accuracy	Time consuming	Image number	Recognition accuracy	Time consuming
P1	91.47	6.74	P11	97.05	8.71
P2	92.85	8.05	P12	92.44	9.33
P3	95.21	6.46	P13	96.68	8.60
P4	91.57	8.14	P14	90.12	7.07
P5	96.07	7.52	P15	96.09	6.91
P6	91.26	8.62	P16	96.18	7.41
P7	95.05	6.33	P17	94.8	8.49
P8	93.24	9.41	P18	97.48	9.92
P9	92.59	7.86	P19	95.95	8.41
P10	93.72	8.95	P20	95.26	827

Table 3 Recognition accuracy (%) and time consumption (s) of traditional methods

Image number	Recognition accuracy	Time consuming	Image number	Recognition accuracy	Time consuming
P1	83.82	25.54	P11	82.21	23.66
P2	81.48	26.03	P12	79.44	19.79
P3	80.66	17.71	P13	80.57	25.78
P4	82.34	19.87	P14	80.27	24.45
P5	83.74	15.09	P15	81.22	27.69
P6	80.70	18.91	P16	81.51	26.89
P7	80.66	19.04	P17	82.26	30.38
P8	83.21	22.52	P18	82.99	32.58
P9	83.14	24.68	P19	83.69	29.91
P10	82.87	44.12	P20	80.94	26.38

recognition time of the experimental group is less than 10 s, which is far less than that of the control group. The lowest recognition time of the traditional method is 15.09 s. For further processing of table data, the average accuracy of the experimental group was 94.26%, while the average accuracy of the control group was 81.89%. The experimental results show that the recognition rate of this method is improved by 12.9%. The average time of the experimental group was 7.8 s, while the average time of the control group was 23.35 s. The experimental results show that the time cost of this method is reduced by 15.55 s. This is because the method in this paper avoids the loss of some heterogeneous features in image reconstruction by limiting the feasible region of super-resolution image reconstruction, and improves the accuracy of recognition of heterogeneous features in super-resolution image. Through the subsequent convolution neural network method for feature clustering analysis, the repetitive heterogeneous feature recognition process is reduced, and the recognition time is shortened. To sum up, the heterogeneous feature recognition method studied in this paper has good recognition accuracy in practical application, reaching 94.26%, with low recognition time, with an average of 7.8 s, which has a certain degree of practical value.

4 Conclusion

Super resolution image is widely used in military, science and technology, industrial production and daily life. After low resolution image is processed, it contains more details. Therefore, it is of great significance to further process the features in super resolution images.

Affected by factors such as imaging equipment, the resolution of the image actually obtained by the user is often limited, resulting in insufficient image details. This will not only affect the user's visual experience, but also constitute an obstacle to related computer vision applications. Therefore, how to improve the resolution of an image (so-called image super-resolution reconstruction) has always been a basic research topic in the field of image processing, and has received extensive attention and in-depth research from many scholars. Image super-resolution reconstruction refers to a digital image processing technology that reconstructs a high-resolution image from one or more low-resolution images. According to different image types, image super-resolution reconstruction mainly includes color image super-resolution reconstruction and depth image super-resolution reconstruction. With the deepening of research and application, the research of super-resolution reconstruction of mixed color and depth images has also appeared one after another, and has gradually become a research hotspot. In the context of the above research, Compared with other common features, the information contained in heterogeneous features is more important for the application of super resolution image. Aiming at the problems of traditional image feature recognition methods, this paper studies the accurate recognition method of heterogeneous features in super resolution image visualization. By calculating the feasible region of image reconstruction, the image is reconstructed within the constraints. Principal component analysis (PCA) is used to reduce the dimension of super-resolution image, and the recognition of heterogeneous features of image is realized in convolution neural network. In the simulation experiment, the feasibility of the research method is verified. The experimental results show that the accuracy of the heterogeneous feature recognition method is 94.26%, and the average recognition time is 7.8 s. In the future research, it is necessary to further study the processing efficiency of the method in order to improve the performance of the recognition method.

The text mainly conducts in-depth research on the super-resolution reconstruction of super-resolution images, and has achieved preliminary results. However, there are still some problems that need to be further studied and improved. In the problem of single image super-resolution reconstruction, building an image observation model that meets the characteristics of the actual imaging system is the key to the success of the super-resolution algorithm. However, in practical applications, it is more difficult to obtain the point spread function in the image observation model. We usually use a simple point spread function to approximate, and such an approximation has a large gap with the actual imaging system, which is bound to affect. The effect of the reconstruction algorithm. Therefore, seeking a new image observation model to make the observation model reflect the imaging process more accurately and comprehensively, and to achieve accurate estimation of the point spread function is an important issue to be further studied and improved. In the future, we can also study the following directions:

(1) In the problem of depth map super-resolution reconstruction, it is generally assumed that the color image and depth image of the same scene have been calibrated and registered. However, in practical applications, we need to control the color camera and the depth map. Calibration and registration are carried out between cameras. Although the current camera calibration and registration can achieve very high accuracy, the level of automation is still very low, and the calibration and registration process is relatively cumbersome. Therefore, it is very necessary to study highly automated camera calibration and registration technology in combination with the problem of depth image super-resolution reconstruction.
(2) In the problem of depth image super-resolution reconstruction, currently only a high-resolution color image of the same scene is used as a guide to explore the intrinsic relationship between the same scene color image and the depth image, and the high-resolution.
(3) The inherent information of the color image is fused into the super-resolution algorithm of the depth image. Considering that the depth image can also be obtained through the left and right image stereo matching algorithm, in the next step, we can combine the left and right high resolution color images of the same scene, and apply the stereo matching algorithm idea to the depth image super-resolution problem. Further improve the quality of the depth image.

Acknowledgements Gansu Province Key R&D Program: Research on UAV Key Technologies (No: 18YF1GA060).

References

Chen, W. (2019). Research on the method of super-resolution image recognition in machine vision system. *Microelectronics & Computer, 36*(06), 105–108.

Huang, X., Ling, Z., & Li, X. (2018). Discriminative deep feature learning method by fusing linear discriminant analysis for image recognition. *Journal of Image and Graphics, 23*(04), 510–518.

Lei, W., Wang, Y., & Li, J. (2020). Optimization research of image super resolution reconstruction algorithm based on FSRCNN. *Transducer and Microsystem Technologies, 39*(02), 54–57.

Li, Y. (2019). High-precision recognition method for face image texture features based on PCA model. *Bulletin of Science and Technology, 35*(07), 135–138.

Luo, J. (2019). A method of texture feature extraction based on rotation invariant co -occurrence pattern. *Journal of Xuchang University, 38*(02), 124–131.

Sun, S., Yin, L., Yan, X., et al. (2018). Defect recognition of welding image based on texture feature. *Computer Applications and Software, 35*(05), 242–246.

Tang, J., Zhu, G., & Du, Z. (2018). Super-resolution restoration algorithm based on SVM multi-figure classification learning. *Journal of Nanjing Normal University, 41*(03), 28–34.

Wang, G. (2019). Research on RELM oil painting recognition based on color characteristic and texture characteristic. *Microcomputer Applications, 35*(10), 57–60.

Wang, P., Zeng, L., & Wu, Y. (2019). Image feature intelligent recognition model based on large dataset domain adaptive fast algorithm. *Foreign Electronic Measurement Technology, 38*(04), 7–11.

Xu, Y., Mi, Q., Liu, B., et al. (2018). Feature recognition method of coal and gangue image based on dictionary learning. *Journal of Shandong University of Science and Technology, 37*, 66–72.

Yang, J. (2019). Image splicing detection based on composite texture features. *Software Guide, 18*(01), 178–182.

Yang, Z., & Wang, C. (2018). Research on image super-resolution method based on convolutional neural network. *Journal of Shandong Normal University, 33*(04), 427–433.

Ying, Z., & Long, X. (2019). Super-resolution image reconstruction based on feature compensation depth neural network. *Journal of Wuyi University, 33*(03), 66–71.

Yu, Y. (2019). Image feature recognition technology simulation in mobile terminal gesture interaction. *Computer Simulation, 36*(10), 462–466.

Yu, F., Wu, X., Chen, J., et al. (2019). Exploiting images for video recognition: heterogeneous feature augmentation via symmetric adversarial learning. *IEEE Transactions on Image Processing, 28*, 5308–5321.