Application of Deep Learning in Super-resolution Processing of Face Images

Zhichao Zhang
University of Birmingham, Edgbaston B15 2TT, United Kingdom

Abstract. The resolution of an image is generally defined as the number of pixels contained in a unit size. The higher the resolution, the more details it contains and the clearer the image is. Because the results of reconstructing high-resolution images from low-resolution images are not unique, the super-resolution of images is a morbid inverse problem, and it is also a challenging and open research topic in the field of computer vision. With the development of machine learning and deep neural network, this paper studies the application of DL (Deep Learning) in face image super-resolution processing. The results show that the algorithm in this paper combines detail enhancement module and synthesizes fine-grained structure from high-resolution example image, which can generate low-frequency details of the image and transmit high-frequency details from the example image to the basic image for enhancement. It can be seen that the algorithm in this paper is better than other methods in evaluation index and visual effect. However, there are still some shortcomings in this algorithm and experiment, which need further research and improvement. The two image super-resolution algorithms proposed in this paper are both aimed at improving the perceptual quality of reconstructed images.

Keywords: Deep learning, Facial images, Super-resolution processing.

1. Introduction

In recent years, with the continuous development of society and the continuous progress of science and technology, people's work and life are increasingly pursuing high efficiency, and the demand for using scientific and technological means to assist in improving people's daily work efficiency is increasing. As an effective biological feature of personal identification, human face has extremely important applications. In our daily life, the face images obtained by monitoring cameras often have insufficient resolution and contain a lot of noise [1]. The resolution of the image is generally defined as anyone can easily and quickly obtain information from videos or images in daily life, but the amount of information obtained is closely related to the clarity of images and videos. Therefore, the clarity of images directly affects the quality of image information obtained. For low-resolution face recognition, the mainstream method is to directly extract features from low-resolution face images and directly recognize them, which is called a one-stage method; However, because the low-resolution image can contain less feature information and the features directly extracted are not sufficient, the accuracy and robustness of recognition are not high enough [2-3]. Because the results of reconstructing high-resolution images from low-resolution images are not unique, the super-resolution of images is a morbid inverse problem, and it is also a challenging and open research topic in the field of computer vision. With the development of machine learning and deep neural network, this paper studies the application of DL in super-resolution processing of face images. Under real conditions, visible light images are often affected by the limitations of image acquisition equipment or external interference during transmission, resulting in low resolution of images, so it is difficult to obtain high-resolution visible light images directly [4]. Therefore, the acquisition of high-resolution infrared images is also very important. A great breakthrough has been made generate a high-resolution image with more realistic texture details, and then face recognition is carried out to obtain more accurate and robust identity recognition [5].

2. Facial image super-resolution reconstruction

2.1 Deep learning based super-resolution facial image reconstruction method

DL belongs to one of the categories of machine learning. Compared with other machine learning technologies, DL methods are usually trained end-to-end, directly from the original image to the desired output image. This feature is in line with the real-time nature of super-resolution technology. CNN (Convolutional Neural Network) is a feedforward neural network, which simulates the neural structure in biology in structure and is a hierarchical network with a certain depth [6]. CNN has excellent feature extraction capabilities, which can extract high-frequency feature information required for matching image processing tasks. When the input feature map is multi-channel, the number of output channels is determined by setting the number of independent convolutional kernels, that is, a single convolutional kernel is convolutionally operated with each channel separately, and then added and summed to obtain the feature map results of the corresponding output channel of the convolutional kernel [7]. The advantage of CNN lies in completing the image upsampling task through traditional interpolation methods, utilizing convolutional layers to automatically learn features, and reducing the difficulty of manual learning. In addition, by adjusting the scaling factor of the interpolation method, reconstructed images of any size can be reconstructed, as shown in Figure 1.

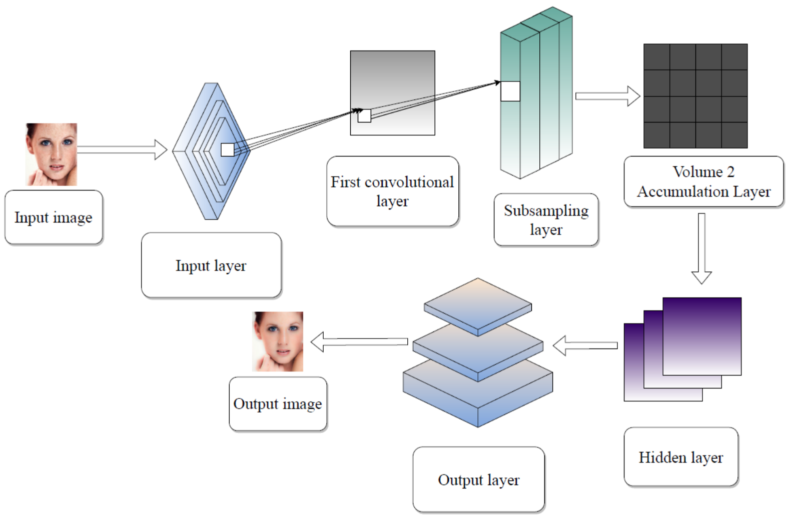

Figure 1. Flow chart of facial image super-resolution reconstruction using CNN

In qualitative analysis, in subjective vision, some details of the reconstructed image super-resolution algorithm based on attention mechanism are closer to the original high-resolution image, and the details and textures are clearer. The clarity of images directly affects the quality of image information obtained. For low-resolution face recognition, the mainstream method is to directly extract features from low-resolution face images and directly recognize them, which is called a one-stage method; However, because the low-resolution image can contain less feature information and the features directly extracted are not sufficient, the accuracy and robustness of recognition are not high enough. Although there are great differences between different image super-resolution models, their essence is to up-sample the image, and the difference is the up-sampling operation and its position in the model. Therefore, the key to the image super-resolution problem is how to up-sample, that is, how to generate high-resolution images from low-resolution images [8]. For an image I with N pixels, the brightness and contrast of the image can be obtained:

$$\mu_I = \frac{1}{N}\sum_{I=1}^{N} I \tag{1}$$

Among them, $I(i)$ is the intensity of the i pixel in image I, with brightness μ_I and contrast σ_I corresponding to the expected and variance of the image, respectively. Furthermore, the comparison coefficients $C_I\left(I,\hat{I}\right)$ and $C_c\left(I,\hat{I}\right)$ for brightness and contrast of two images can be obtained as follows:

$$C_I\left(I,\hat{I}\right)=\frac{2\mu_I\mu_i+C_1}{\mu_I^2+\mu_i^2+C_1} \tag{2}$$

$$C_c\left(I,\hat{I}\right)=\frac{2\sigma_I\sigma_i+C_2}{\sigma_I^2+\sigma_i^2+C_2} \tag{3}$$

Low resolution images are fed into deep convolutional networks for operation without increasing their resolution, and an end-to-end learnable upsampling layer is applied at the end of the network. The advantage of this method is that due to the high computational complexity of feature extraction in low dimensional space and only upsampling at the end of the network to improve resolution, the computational complexity and spatial cost are greatly reduced [9-10].

2.2 Example-based image super-resolution reconstruction method

The reference image needs to have a texture structure or content structure similar to that of the low-resolution image, and the reference image can be selected from adjacent frames in the video, images retrieved from the network, external database dictionaries or images based on self-examples. In the process of network information transmission, for the calibration of extracting feature weights, selectively paying attention to local features on some layers is helpful to the final reconstruction [11]. Because the results of reconstructing high-resolution images from low-resolution images are not unique, the super-resolution of images is a morbid inverse problem, and it is also a challenging and open research topic in the field of computer vision. Under real conditions, visible light images are often affected by the limitations of image acquisition equipment or external interference during transmission, resulting in low resolution of images, so it is difficult to obtain high-resolution visible light images directly. The attention-based model holds that not all features are necessary for super-resolution, but different feature information has different importance. Transferring high-resolution details from reference images to low-resolution images is the key to the success of RefSR, so two key problems need to be solved: first, how to establish the corresponding relationship between low-resolution images and reference images; The second is the high-resolution information synthesis of low-resolution images. The flow chart of this method is shown in Figure 2.

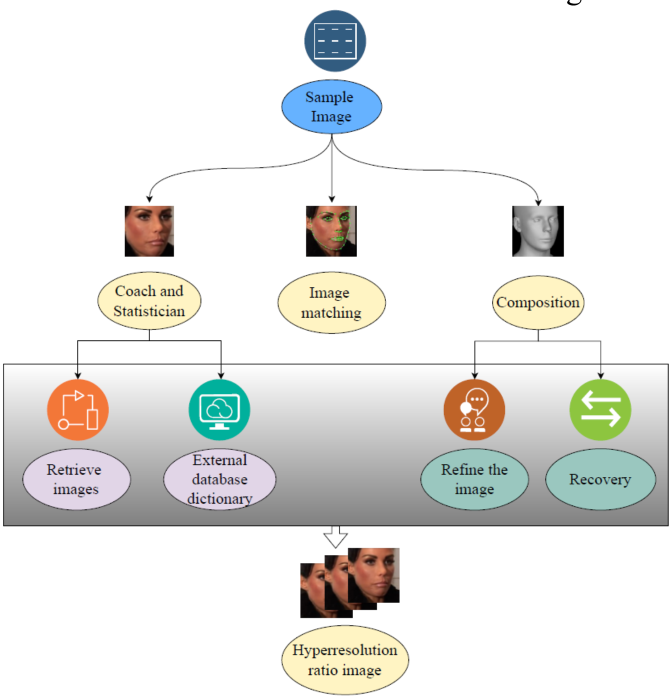

Figure 2. Flow chart of an example based super-resolution method

In super-resolution, the higher the peak signal-to-noise ratio of the reconstructed image, the better the quality of the reconstructed image and the performance of the reconstruction algorithm. As an effective biological feature of personal identification, human face has extremely important applications. In our daily life, the face images obtained by monitoring cameras often have insufficient resolution and contain a lot of noise. The calculation process of the peak signal-to-noise ratio of the example image is very simple and not easily affected by noise. It is one of the widely used objective evaluation indicators in evaluating image quality [12].

3. Experimental results and analysis

3.1 Ablation study

Compared with the original low-resolution image, the resolution of the reconstructed image is not significantly improved. Therefore, it can be used in some scenes where image quality is not strict. Because there are many parameters in K-NN search, such as patch size and number of candidate patches, this section analyzes how these parameters affect the final implementation effect of the algorithm. In the detail enhancement step, the patch size is set from 10 x 10 pixels to 30 x 30 pixels in increments of 5. Table 1 shows that this method works best when the patch size is 20 x 20.

Table 1. Influence of different patch sizes on image reconstruction effect

Patch size	PSNR	SSIM
10 x 10	35.45	0.88
20 x 20	38.17	0.91
30 x 30	34.28	0.85

Adding the component diagram of the face to the network input can further improve the quality of the generated image and enable it to have a clearer contour structure. By registering, the relative motion between the low resolution image and the referenced low resolution image can be obtained, and prior knowledge can be used to process the low resolution image. After years of development, existing reconstruction based methods have become very mature, but their effectiveness is poor when the scale factor of resolution amplification is large. And if there is noise in low resolution images, the algorithm results will be greatly affected. The detail enhancement module generates realistic and realistic textures in facial features and other areas. To address the issue of insufficient restoration of image details, generate realistic and realistic textures in facial features and other areas.

3.2 Result analysis

The relative motion between the low-resolution image and the referenced low-resolution image can be obtained by registration, and then the low-resolution image is processed by prior knowledge. After years of development, the existing methods based on reconstruction are very mature, but the effect is not good when the scale factor of resolution amplification is large. And if there is noise in the low-resolution image, the algorithm results will be greatly disturbed. For image super-resolution reconstruction, the amount of high-frequency information of image features will directly affect the details of the reconstructed image with large gradient jump. Only by extracting enough detailed information can we fit a more appropriate mapping relationship between face images and make the network have stronger generalization ability. In order to further verify the algorithm in this paper, this method is compared with two super-resolution methods, including bicubic interpolation, SRCNN and VDSR. Experiments are carried out on FEI dataset and Multi-PIE dataset, and PSNR and SSIM are used as evaluation indexes to evaluate the experimental results. The experimental results on FEI data set are shown in Table 2.

Table 2. Comparison Results of Different Methods

Algorithm	PSNR	SSIM
SRCNN	36.15	0.915
VDSR	32.58	0.924
Algorithm in this article	40.19	0.957

From Table 2, it can be seen that the algorithm in this article integrates the detail enhancement module and synthesizes fine-grained structures from high-resolution example images. While generating low-frequency details in the image, it can transmit high-frequency details from the example image to the base image for enhancement. It can be seen that the algorithm in this article has shown significant improvement in evaluation indicators and visual effects compared to other methods.

4. Conclusions

With the rapid development of computer artificial intelligence and the progress and development of digital image processing technology. Super-resolution of face image refers to generating high-resolution or super-resolution face image according to low-resolution image input. Target super-resolution image are extracted, and then the features matching the reference image are searched in the feature space, and then the effective texture is migrated to the feature map of low-resolution image through multi-scale feature migration. When the input feature map is multi-channel, the number of output channels is determined by setting the number of independent convolutional kernels, that is, a single convolutional kernel is convolutionally operated with each channel separately, and then added and summed to obtain the feature map results of the corresponding output channel of the convolutional kernel. Finally, the simulation integrates the detail enhancement module and synthesizes the fine-grained structure from the high-resolution example image. While generating the low-frequency details of the image, it can transmit in evaluation index and visual effect. However, there are still some shortcomings in this algorithm and experiment, which need further research and improvement. The two image super-resolution algorithms proposed in this paper are both aimed at improving the perceptual quality of reconstructed images. It is equivalent to solving the same problem from different angles with two different methods, and then considering how to combine these two methods to further improve the perceptual quality of images is the focus of subsequent research.

References

[1] Kwon O. Face recognition Based on Super-resolution Method Using Sparse Representation and Deep Learning[J]. Journal of Korea Multimedia Society, 2022, 21(5):18-30.

[2] Jiang J, Wang C, Liu X, et al. Deep Learning-based Face Super-resolution: A Survey[J]. International journal of computer assisted radiology and surgery, 2021, 67(14):30-57.

[3] Li X, Wu Y, Zhang W, et al. Deep learning methods in real-time image super-resolution: a survey[J]. Journal of Real-Time Image Processing, 2020, 17(10):22-30.

[4] Hashimoto R, Okura Y, Yamamoto M, et al. An Approach towards 4 times Super-Resolution Processing for SPECT Images by Deep Learning Technique[J]. International journal of computer assisted radiology and surgery. 2020, 68(62):15-45.

[5] Jha B, Jz A, Yuan S B, et al. Facial Image Synthesis and Super-Resolution With Stacked Generative Adversarial Network[J]. Neurocomputing, 2020, 402(10):359-365.

[6] Zhang Y, Zhang Z, Li T. Research on Image Super-resolution Reconstruction Based on Deep Learning[J]. Journal of Physics: Conference Series, 2021, 1802(4):042034-042051.

[7] Li D, Wang Y, Liu D, et al. A Novel Image Super-Resolution Method Based on Attention Mechanism[J]. Journal of Physics Conference Series, 2020, 1518(9):012036-012057.

[8] Zhu P, Zhu H, Jian Z. A deep auto-encoder based technique for face image super resolution[J]. Computer Era, 2019, 69(15):40-62.

[9] Menezes A G. Analysis and evaluation of Deep Learning based Super-Resolution algorithms to improve performance in Low-Resolution Face Recognition[J]. IEEE Signal Processing Letters, 2021, 23(11):51-59.

[10] Niu C, Nan F, Wang X. A super resolution frontal face generation model based on 3DDFA and CBAM[J]. Displays, 2021, 69(51):102043-102069.

[11] Wang H, Hu Q, Wu C, et al. DCLNet: Dual Closed-loop Networks for face super-resolution[J]. Knowledge-Based Systems, 2021, 222(33):106987-106987.

[12] Chen J, Chen J, Wang Z, et al. Identity-Aware Face Super-Resolution for Low-Resolution Face Recognition[J]. IEEE Signal Processing Letters, 2020, 112(99):21-51.