Research on Image Super-resolution Internet Transmission Technology Based on Virtual Optics

Gengyi Xiao
Department of Mathematics and Computer Technology, Guilin Normal College, Guilin, China
xiao_gy@glnc.edu.cn

Abstract. Image resolution is a set of performance parameters used to evaluate the richness of detailed information contained in an image, including temporal resolution, spatial resolution and color scale resolution, etc., which reflects the ability of the imaging system to reflect the detailed information of objects. In this paper, a super-resolution image transmission technology based on virtual optics is proposed. The original image is placed on the object plane of the virtual light path. Firstly, the light wave signal of the original image on the virtual diffraction plane is obtained by inverse diffraction calculation. Then, the light wave of the virtual diffraction plane is illuminated by spherical wave and then the forward diffraction calculation is performed. By changing the position of the observation plane, the original images with different magnification can be reconstructed. The simulation test results show that compared with the general interpolation amplification method, the amplified image has a good visual perception effect, especially in the significant area. The results obtained by the algorithm in this paper are better than the interpolation model and convolutional neural network algorithm as a whole; The above test results objectively show that the super-resolution effect of this algorithm is better.

Keywords: Virtual optics, Image super resolution, Internet transmission

1. Introduction
Image super-resolution technology studies how to recover a higher resolution image from a group of lower resolution image sequences or a lower resolution image [1]. High resolution emphasizes that the obtained image contains more pixel information and more high frequency information [2]. This technology can provide images with better visual effect and more image information, which can be divided into four types, namely, methods based on interpolation [3], reconstruction [4], enhancement [5] and learning [6]. Among them, the learning-based image super-resolution method has been developing rapidly in recent years. This method uses the image database or the image itself to obtain the association between the high-resolution image and the low-resolution image through learning, and takes it as a prior constraint to generate the high-resolution image. Learning-based methods can be divided into self-learning and external database-based methods.

Interpolation is the most direct and simple method in the study of super-resolution of image and video. However, this method will blur and smooth the image and increase the noise. The two key
problems of reconstruction-based method are registration and reconstruction. First of all, in the registration process, multi-frame low-resolution ordered image data can be effectively utilized, so as to obtain the required effective information. Secondly, the reconstruction process, according to the acquired knowledge, optimizes the image to be super-resolved. Learning-based methods think that low-resolution images have all useful information, and can directly predict the information corresponding to high-resolution images. Firstly, training is carried out through the data sets of low-resolution images and high-resolution images, and then a training optimization model is obtained. Through this model, the output super-resolution images obtained from the input low-resolution images can be directly calculated.

In this paper, an image super-resolution Internet transmission technology based on virtual optics is proposed. Its basic idea is to place digital images in virtual light path, and realize controllable image amplification by simulating the propagation of light waves in virtual light path. In addition, in the level set method, local salient targets are segmented by combining saliency map, and the contour of targets is extracted by canny operator after enlargement and reconstruction. In this paper, the contour smoothness of local salient targets after enlarged reconstruction is good, and the comparison results of objective evaluation indexes under different methods are given.

2. Principle of digital image controllable amplification based on virtual optics

The controllable amplification of digital images is based on Fresnel diffraction theory. Diffraction of light refers to the phenomenon that light deviates from straight line when it encounters obstacles or small holes during propagation. The diffraction of light waves in the near-field region is called Fresnel diffraction. Under the framework of scalar diffraction theory, Fresnel diffraction integral shown in formula (1) can be used to calculate the physical propagation process of light waves in the near-field region [7].

\[
U(x, y) = \frac{\exp(\jmath kd)}{\jmath \lambda d} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} f(x_0, y_0) \cdot \exp\left(\jmath k \left[\left(x-x_0\right)^2 + \left(y-y_0\right)^2\right]/(2d)\right) dx_0 dy_0
\]

(1)

In which, \(f(x_0, y_0)\) represents complex amplitude of object plane light wave, \(U(x, y)\) represents complex amplitude of observation plane light wave, \(d\) is diffraction distance, \(j = \sqrt{-1}\). \(\lambda\) is the wavelength of light, \(k = \frac{2\pi}{\lambda}\).

Fresnel diffraction integral can be expressed as Fourier transform and convolution, and there are one-time fast Fourier transform and fast convolution algorithm [8]. The fast convolution algorithm can be used to reconstruct the object light waves in the observation plane, which represents Fresnel diffraction transfer function, and \(F\) and \(F^{-1}\) represent Fourier forward transform and inverse transform respectively.

\[
U(x, y) = F^{-1}\{F[f(x_0, y_0)]H_F(f_x, f_y)\}
\]

(2)

\[
H_F(f_x, f_y) = \exp\left(\jmath kd \left[1 - \frac{\lambda^2}{2} (f_x^2 + f_y^2)\right]\right)
\]

(3)

From \(U(x, y)\), the object plane light wave signal \(f(x_0, y_0)\) can be obtained by inverse operation.

\[
f(x_0, y_0) = F^{-1}\{F[U(x, y)]H_F^*(f_x, f_y)\}
\]

(4)
Formula (2) and formula (4) constitute Fresnel diffraction transformation pair, which can be used to realize the transformation between the optical signals of the initial plane and the observation plane in the optical system.

3. Image super-resolution internet transmission technology in wireless communication environment

3.1. Interpolation model

We consider the case of double image interpolation. As shown in fig. 1, black dots represent known low-resolution dots, and the remaining blue and blank dots represent unknown high-resolution dots. The key of image interpolation is how to infer from the information in the neighborhood of unknown points.

![Figure 1](image)

**Figure 1.** Schematic diagram of the relationship between low-resolution images and high-resolution images. Black points represent low-resolution image pixels, while blue and blank points represent high-resolution points to be estimated.

In the proposed method, we first divide the image domain $\Omega$ into smaller overlapping sub-regions $\Omega = \{\Omega_1, \ldots, \Omega_k\}$. For each unknown pixel, we estimate the interpolation model according to its local neighborhood. Assume that $x_i \in R^2$ is the current point to be interpolated, which is located at the center of $\Omega_i$. As we know, natural images have strong local structure, which means that pixels in natural images have strong local dependence. According to this characteristic, we estimate the value of the current unknown point $\phi(x_i)$ by linearly combining the known points around us. In particular, we define the linear affine transformation model $f_i(\cdot; a_i, b_i)$ as shown in the following formula [9-10]:

$$f_i(x_i) = \langle a_i, \phi(x_i) \rangle + b_i$$

(5)

Here, $\phi(x_i)$ is the luminance value vector of k-nearest neighbor of $x_i$ along all directions, and $a_i$ and $b_i$ are the deviation between weight vector and linear estimation; $f_i(x_i)$ is the estimated value of $x_i$, and $\langle \cdot, \cdot \rangle$ represents the inner product. To simplify the description, we treat the deviation terms as follows:

$$\Phi(x_i)^T \left[ \phi(x_i)^T, 1 \right]^T w_i^T \left[ a_i^T, b_i \right]$$

(6)

The linear affine transformation model defined above can become:
\[ f_i(x_i) = w_i \Phi(x_i) \]  

(7)

Considering that the linear affine transformation function \( f_i \) is defined at every pixel, rather than shared by all points in the local neighborhood \( \Omega \), the proposed method is actually globally nonlinear and locally linear. This feature is more in line with the statistical law of natural images. We call the proposed interpolation model a local linear regression model.

3.2. Super-resolution algorithm based on convolutional neural network

Convolutional neural network is regarded as an image enhancement method to optimize the output results of thermal diffusion model and Rayleigh scattering model. We hope to learn a mapping \( F \) through a convolutional neural network, which is conceptually composed of four operations. Next, we will explain the definition of each operation in detail:

Pretreatment. The picture is enlarged to the required times by thermal diffusion model and Rayleigh scattering model. In the process of improving picture resolution, the parameters involved in these two algorithms based on point diffusion are generated adaptively, without extra input, so as to minimize the introduction of errors. The calculation results are also due to traditional methods such as bilinear interpolation.

Image block extraction and representation. This operation superimposes image blocks from the preprocessed image \( Y \). We use sparse representation for reference to represent each image block as a high-dimensional vector [11]. These vectors include a set of feature maps, the number of which is equal to the dimensions of the vectors. The densely extracted image blocks are further represented by a set of pre-trained bases.

The vector space constructed by these bases represents the multi-dimensional space of low-resolution pictures. Commonly used basic training methods include PCA analysis, DCT, Haar and so on. They are essentially equivalent to convolving images through a set of filters. Each filter represents a vector direction in a multi-dimensional space. The above contents are abstractly summarized into the following mathematical formula, and the second layer is expressed as operation \( F_1 \):

\[ F_1(Y) = \max(0, W_1 * Y + B_1) \]  

(8)

In which \( W_1 \) and \( B_1 \) represent convolution layer operation and error, respectively. The size of \( W_1 \) here is \( c \times d_1 \times d_1 \times n_1 \). In which \( c \) is the number of channels of the input image, \( d_1 \) is the spatial size of convolution layer operation units, and \( n_1 \) is the number of convolution layer operation units. In short, \( W_1 \) represents \( n_1 \) convolution layer operation units, whose size is \( d_1 \times d_1 \).

Projection in high resolution space. The whole process is designed as the nonlinear projection of the extracted low-resolution image blocks in the high-resolution representation space. The above operation converts each low-resolution image block into an \( n_1 \)-dimensional description. In the following processing, the mapping in these \( n_1 \)-dimensional descriptors is a high-dimensional mapping of \( n_2 \)-dimensional high-resolution image blocks. This is equivalent to \( n_2 \) filters with a spatial size of \( 1 \times 1 \) using convolution layer operation unit. The operation of the third layer is:

\[ F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \]  

(9)

Here, the size of convolution layer operation unit \( W_2 \) is \( n_1 \times 1 \times 1 \times n_2 \), and error \( B_2 \) is \( n_1 \)-dimensional descriptor. Each output \( n_2 \) dimension description is a representation of a high-resolution image block. In this step, multiple layers of similar operations can be set up to increase accuracy.
Rebuild. In this process, the predicted high-resolution small blocks are overlapped to generate the final high-resolution image.

\[ F_3(Y) = W_3 \ast F_2(Y) + B_3 \]  

(10)

The size of convolution layer operation unit \( W_3 \) is \( n_2 \times f_3 \times f_3 \times c \), and \( B_3 \) is a \( c \)-dimensional description of characterization error. If the representation of high-resolution image blocks is in the high-resolution image spatial domain, each representation can be simply reconstructed to form image blocks. We expect this step to achieve an average effect; If the representation of high-resolution image blocks is in the high-dimensional vector space constructed on some bases, we expect that the coefficients should be projected onto the image domain first and then averaged.

### 3.3. Generate countermeasure network structure

The model of generating countermeasure network consists of generator and discriminator. The generator is used to generate pictures. The goal of the generator is to input the generated false pictures into the discriminator, so that the discriminator can output 1. The generator is reversely updated according to the discriminant results of the discriminator to optimize the generation model. The function of the discriminator is to judge whether the input picture is a picture generated by the generator or a real picture, and the discriminator outputs a real number between 0 and 1. The closer the output of the discriminator is to 1, the more likely the input picture is to be a real picture. The closer the output of the discriminator is to 0, the more likely the input picture is to be a false picture.

In the training process, this paper updates the network model by the value output by the discriminator. The generator and the discriminator are constantly trained in the process of mutual game, which finally makes the discriminator unable to distinguish whether the input picture is a real picture or a picture generated by the generator, and the final output value is around 0.5. The generated confrontation network finally reaches Nash equilibrium by mutual game between the generator and the discriminator [12], and the network converges. The objective function of generating the countermeasure network is as follows [13]:

\[
\min_{D} \max_{G} V(D,G) = E_{x \sim \text{pdata}(x)}[\log D(x)] + E_{z \sim \text{pc}(z)}[\log(1 - D(G(z)))]
\]  

(11)

When updating parameters in the program, the method adopted is to update the discriminator for \( n \) times, and then update the generator for one time [14].

When updating the parameters of the discriminant model, it is necessary to get the maximum value of the above formula, so it is hoped that the output of \( D(x) \) of sample \( x \) from real sample data \( \text{pdata} \) should be as close as possible to 1, that is, \( \log D(x) \) should be as large as possible; It is also expected that the output \( D(G(z)) \) of data \( G(z) \) generated by random noise should be as close to 0 as possible.

When updating the parameters of the generated model, the data generated by the generator should be as close as possible to the real data, that is, \( p_g = \text{pdata} \). In this way, \( D(G(z)) \) is as large as possible (the upper limit is 1), and \( \log(1 - D(G(z))) \) is as small as possible, so it is necessary to minimize equation (8).

### 4. Experimental analysis

#### 4.1. Algorithm comparison

Figure 2 below shows the comparison result between the new algorithm and bilinear interpolation. After the original image is reduced, it is used as the input of the subsequent reconstruction process. With the increase of magnification, the linearly enlarged image appears sawtooth phenomenon, while
the algorithm in this paper keeps the edge of smooth transition. When expanded by 16 times, the linearly enlarged image is blurred; However, this paper still achieved good results.

Figure 2. Comparative result

4.2. Discussion on controllable image magnification based on virtual optics
Compared with classical image interpolation methods, such as nearest neighbor, bilinear and bicubic methods, this method involves Fresnel diffraction integral calculation, and its execution time is relatively long, but its time complexity has obvious advantages compared with image interpolation methods, bilinear and bicubic methods.

Fresnel diffraction light wave coding of digital image is realized by means of discrete Fresnel diffraction transformation formula. Because of the non-band limitation of Fresnel diffraction transfer function and the approximate satisfaction of sampling theorem, the high frequency components in reconstructed image are lost, which shows that the sharpness at sharp corners in reconstructed image is lost to a certain extent, but the reconstruction effect on smooth edge objects is good.

The formation of magnified reconstructed image is based on the interpolation of discrete diffracted light wave signal sinc function. When the magnification is large, the interpolation points of reconstructed image surface are larger, and the diffraction fringes of reconstructed image become obvious, which affects the visual perception effect.

4.3. Objective effect
Objective evaluation indexes are mainly PSNR and image SSIM [15]. The former quantitatively calculates the error between the processed result and the original image, and the higher the PSNR, the smaller the distortion; The closer SSIM is to 1, it shows that the processed structure is very similar to the original structure, that is, the resulting graph is better.

The results of interpolation model, convolutional neural network and the algorithm in this paper are evaluated objectively by using the above two indexes. Table 1 gives the objective test results.
Table 1. Comparison of objective indicators

| Image  | Index | Interpolation model | Convolutional Neural Network | This algorithm |
|--------|-------|---------------------|------------------------------|----------------|
| Butterfly | PSNR  | 23.01               | 25.66                        | 29.69          |
|         | SSIM  | 0.71                | 0.81                         | 0.91           |
| Bird    | PSNR  | 22.33               | 28.90                        | 31.20          |
|         | SSIM  | 0.68                | 0.83                         | 0.99           |
| Figure  | PSNR  | 27.88               | 27.16                        | 30.21          |
|         | SSIM  | 0.79                | 0.79                         | 0.87           |
| Flower  | PSNR  | 28.06               | 29.36                        | 31.36          |
|         | SSIM  | 0.82                | 0.87                         | 0.97           |

Except that the objective index of the image tested by convolution neural network is slightly higher than that of this algorithm, other test results show that the PSNR and SSIM obtained by this algorithm are higher than those obtained by interpolation model and convolution neural network algorithm, which fully shows that the results produced by this algorithm are very close to the original image and the processing effect is better.

In order to highlight the effective improvement of the algorithm in this paper, a series of training and learning processes are carried out, and the results of various network learning models are obtained. According to the two different network learning models obtained by the algorithm in this paper under the same iteration times, the above test images are processed with super resolution, and the corresponding PSNR and SSIM values are obtained. Taking the butterfly diagram of the test image as an example, the network model obtained by different iterations is used to test and record it, and finally the change trend of PSNR and SSIM is obtained, as shown in Figures 3 and 4 respectively. For convenience of comparison, interpolation model and test results of convolution neural network algorithm are added in the Figure.

![Figure 3. PSNR obtained by different algorithms](image-url)
It can be seen from fig. 3 and fig. 4 that the results of this algorithm are better than those of interpolation model and convolutional neural network. The above test results objectively show that the super-resolution effect of this algorithm is better.

5. Conclusion
Deep learning has shown great potential in the field of image super-resolution reconstruction, which has greatly promoted the vigorous development of this field. However, there is still a long way to go for deep learning super-resolution image reconstruction technology, which can not only retain all kinds of details of the original image, but also meet people's subjective evaluation. The virtual optical method is applied to image super-resolution internet transmission technology, and the test results of degraded and non-degraded images show that the proposed method shows good visual perception effect compared with general methods, especially in significant areas. The enlarged outline of local salient regions is smoother, and the method proposed in this paper can also be applied to stereo images. In the next step, we will explore improvement schemes for these two problems and apply them to more scenes, such as amplification and reconstruction of significant targets in 3D scenes.

Acknowledgments
This work was financially supported by Guangxi university scientific research fund: Research on the key technologies of vehicle-mounted network based on CPS (2013YB286).

References
[1] Sun Chao, Lu Junwei, Li Jianwei, et al. Fast image super-resolution method based on deconvolution [J]. Acta Optica Sinica, 2017, 037(012): 142-152.
[2] Cerio D G, Witmer L M. Modeling visual fields using Virtual Ophthalmoscopy: Incorporating geometrical optics, morphometrics, and 3D visualization to validate an interdisciplinary technique [J]. Vision Research, 2020, 167: 70-86.
[3] Gao Han, Li Kun-Wei, Jiang Ai-ping. Improvement of super-resolution reconstruction algorithm for optical region image based on learning [J]. Journal of Natural Science, Heilongjiang University, 2019, 036(005): 611-617.
[4] Radel J C, Belanger-Garnier V, Hegedus M P. Virtual image determination for mirrored surfaces [J]. Optics Express, 2018, 26(3): 2599.
[5] Zhao Xiaodong, Zhang Yuying, Cheng Xuemei. Research progress of super-resolution reconstruction of optical remote sensing images [J]. Unmanned System Technology, 2019(3): 20-27.

[6] Scully M, Sokolov A, Svidzinsky A. Virtual Photons: From the Lamb Shift to Black Holes [J]. Optics & Photonics News, 2018, 29(2): 34.

[7] Chen Wenjing, Tang Yi. Overview of image super-resolution reconstruction methods based on deep learning [J]. Journal of Yunnan Minzu University (Natural Science Edition), 2019, v.28; No.118(06): 71-79.

[8] Luo Lizhi, Wu Fei, Cao Kun, et al. Application of image super-resolution in fuzzy license plate recognition system [J]. Software Guide, 2019, 018(005):177-180,186.

[9] Tian Songwang, Lin Suzhen, Lei Haiwei, Li Dawei, Wang Lifang. Multi-band image synchronization supersession and fusion method based on improved WGAN-GP [J]. Acta Optica Sinica, 2020, v.40; No.473(20): 77-84.

[10] Combined super-resolution and deblurring method of text and image based on generated countermeasure network [J]. Computer Applications, 2020, 040(003):859-864.

[11] Dai Qi, Fu Rao, Deng Lian GUI, et al. Research on super surface image display technology [J]. Applied optics, 2019, v.40; No.236 (06): 125-129

[12] Guo Youdong, Ling Furi, Yao Jianquan. Terahertz image super-resolution reconstruction based on gradient transformation [J]. Laser Technology, 2020, 044(003): 271-277.

[13] reny, Liu Chaohui, Essays. Parallel super-resolution reconstruction method of staring image of remote sensing area array [J]. Infrared and Laser Engineering, 2019, v.48; No.291 (01): 299-306.

[14] Sentis M, Onofri F, Fabrice L. Bubbles, drops, and solid particles recognition from real or virtual photonic jets reconstructed by digital in-line holography [J]. Optics Letters, 2018, 43(12): 2945.

[15] Gamo J. Assessing a Virtual Laboratory in Optics as a Complement to On-Site Teaching [J]. IEEE Transactions on Education, 2019, 62(2): 119-126.