

Research Article

Superresolution Reconstruction of Video Based on Efficient Subpixel Convolutional Neural Network for Urban Computing

Jie Shen ¹, Mengxi Xu,² Xinyu Du,^{1,3} and Yunbo Xiong¹

¹College of Computer and Information, Hohai University, Nanjing 211100, China

²College of Computer Engineering, Nanjing Institute of Technology, Nanjing 211167, China

³Bank of Shanghai, Shanghai 200120, China

Correspondence should be addressed to Jie Shen; shenjie_2003045@hhu.edu.cn

Received 22 April 2020; Revised 24 May 2020; Accepted 9 June 2020; Published 11 July 2020

Academic Editor: Wei Wang

Copyright © 2020 Jie Shen et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Video surveillance is an important data source of urban computing and intelligence. The low resolution of many existing video surveillance devices affects the efficiency of urban computing and intelligence. Therefore, improving the resolution of video surveillance is one of the important tasks of urban computing and intelligence. In this paper, the resolution of video is improved by superresolution reconstruction based on a learning method. Different from the superresolution reconstruction of static images, the superresolution reconstruction of video is characterized by the application of motion information. However, there are few studies in this area so far. Aimed at fully exploring motion information to improve the superresolution of video, this paper proposes a superresolution reconstruction method based on an efficient subpixel convolutional neural network, where the optical flow is introduced in the deep learning network. Fusing the optical flow features between successive frames can compensate for information in frames and generate high-quality superresolution results. In addition, in order to improve the superresolution, a superpixel convolution layer is added after the deep convolution network. Finally, experimental evaluations demonstrate the satisfying performance of our method compared with previous methods and other deep learning networks; our method is more efficient.

1. Introduction

Superresolution reconstruction is generating high-resolution results from the low-resolution images using construction models. In contrast to hardware, algorithm-based image construction can efficiently generate high-resolution images at a low cost. Due to a large number of images and video samples, more and more researchers focus on image reconstruction.

Recently, superresolution reconstruction methods have been proposed. For example, Farsiu et al. [1] proposed a regularized reconstruction method for solving inverse problems of ill-posed problems. A fast superresolution reconstruction algorithm is proposed for the pure translational motion and space invariant ambiguity [2]. The principle of neighborhood embedding [3] is to make the assumption that the local spatial structure between the low-resolution image block and the high-resolution image block is similar. The low-resolution image block represents the low-dimensional

data, and the high-resolution image block represents the high-dimensional space. Therefore, the low-resolution image can be linearly represented and mapped to the high-dimensional image block to reconstruct the high-resolution image. Yu and Zhang proposed an improved glowworm swarm optimization algorithm for superresolution reconstruction of video images [4]. Combined with the characteristics of superresolution reconstruction, the algorithm's swarm input, firefly's luciferase, and location update equation are redefined; the optimization objective function criterion is set.

Although the above methods can solve some problems of superresolution reconstruction, the disadvantage of the neighborhood embedding method is that the number k of image blocks selected for low-resolution image blocks is artificially specified, which will affect the reconstruction effect by the supervisor and may cause the phenomenon of underfitting and overfitting. Therefore, we proposed a

new superresolution reconstruction method based on optical flow and efficient subpixel convolutional neural network (ESPCN), which can solve the traditional reconstruction methods' defects.

The rest of the paper is organized as follows. The related work is discussed in Section 2. The methodology is described in Section 3. The details of our method are proposed in Section 4. Section 5 introduced the experiments and result analysis. Finally, the conclusions related to the paper are outlined in Section 6.

2. Related Works

There are three kinds of superresolution methods: interpolation method, reconstruction method, and learning-based method.

The interpolation method is to estimate the missing position of pixels by using the prior in original low-resolution images. Herein, the key of interpolation operation is to establish the mapping relationship between low-resolution and high-resolution images. In general, the interpolation methods include three categories: nearest-neighbor interpolation [5], bilinear interpolation [6], and bicubic interpolation [7]. However, the disadvantage of interpolation-based methods lies in the low-quality reconstruction results that commonly suffer from textual noises.

The reconstruction-based method is to model the degeneration process of image samples and generate a high-resolution image by inverse transformation. The degeneration model establishment is the key, which can fuse multiple images to reconstruct high-resolution images. There are three typical methods in the reconstruction methods: iterative back projection [8], projections onto convex sets [9], and maximum posterior probability [10]. In 2011, Zhang et al. improved the method based on convex set projection for sequential frame image [11] and applied the model to the image sequence of panoramic transformation. This method realizes the geometric transformation of the sequence frame accurately and improves the speed of reconstruction. In 2012, Wallach et al. improved the map method [12] and improved the contrast and resolution of a reconstructed image by fusing global information.

Learning-based methods intend to learn the correlation between the low-resolution and high-resolution images. Three phases are included in these methods: feature extraction, feature learning, and reconstruction. Typical methods include neighbor embedding [3] and sparse coding [13–15]. The representative method based on sparse representation is the superresolution image reconstruction method [16] proposed by Yang et al. in 2008. High- and low-resolution dictionaries are trained for LR image and HR image resolution, so that all low-resolution images to be reconstructed can extract sparse representation from dictionaries. But the style and edge of the reconstructed image need to be similar to the data set used in learning, and the quality and efficiency of reconstruction are not satisfactory.

In recent years, deep learning-based methods have demonstrated excellent performance on superresolution reconstruction [17, 18]. Based on the sparse coding method,

Dong et al. first proposed a superresolution method using a deep convolutional network, which is termed as the super-resolution convolutional neural network (SRCNN) [17]. A 3-layer convolutional neural network (CNN) was designed to learn the mapping relation from the low-resolution to high-resolution images. Aiming to accelerate the reconstruction process, Dong et al. further proposed a fast superresolution convolutional neural network (FSRCNN) that was based on an hourglass-shaped CNN comprising more layers than previous methods but had fewer parameters. The advantage of FSRCNN lies in that it eliminates the requirement of the image magnification and increases the efficiency of the reconstruction process.

Shi et al. proposed an efficient subpixel convolutional neural network method [18] that extracted features in the low-resolution domain and replaced the classic bicubic upsampling operation with an efficient subpixel convolution. Moreover, ESPCN was characterized by the subpixel convolution that generated the feature map in r^2 channel. Through periodic activation in subpixel convolution, the feature map that was of the size $H \times W \times r^2$ was reconstructed to generate the high-resolution images with the size $rH \times rW \times 1$. These image features were extracted by the hidden layers in the deep architecture. In practice, ESPCN was more efficient in contrast to other methods, thus had more opportunities to cater to the real-time tasks. Talab et al. used ESPCN and CNN for super-low-resolution face recognition [19].

In order to expand the receptive field for image reconstruction, Kim et al. proposed a very deep superresolution reconstruction method (VDSR) that increased the kernel size from 13×13 to 41×41 [20]. The sparsity property was used to accelerate the convergence process. Besides, VDSR could generate superresolution results with multiple scales. In contrast to a previous method, VDSR could capture more details of images to improve the performance of the resolution reconstruction.

Ledig et al. introduced the generative adversarial network (GAN) to the superresolution reconstruction and proposed the superresolution generative adversarial network (SRGAN) [21]. SRGAN comprised of a generative block and an adversarial block, where the first was used for reconstruction and the second was used to classify the quality of the reconstructed images. Although SRGAN could not generate competitive results in the PSNR scores, it could obtain results that were in line with the real images, providing a visually satisfying effect.

Sun et al. introduced and updated the residual network (ResNet) into the superresolution reconstruction [22]. Herein, the number of residual layers was increased from 16 to 32, expanding the model scale to generate better results.

3. Methodology

To improve the superresolution reconstruction method in speed and quality, this paper proposed a new method based on optical flow and ESPCN. The motion information between frames is considered in the reconstruction process to improve the reconstruction quality. The optical flow can calculate the motion of an object in a very small time [23].

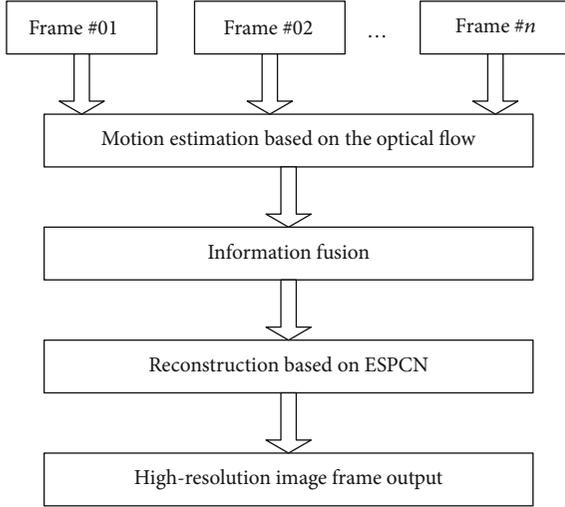


FIGURE 1: The framework of our superpixel reconstruction method.

It is suitable for motion estimation between video frames. Therefore, the optical flow performs well in accuracy and effectiveness compared with other motion estimation methods. Because of the subpixel convolution layer, ESPCN has very little additional computational cost compared to other deep learning methods. ESPCN can be used in real-time to improve the efficiency of urban computing and intelligence. In specific, the optical flow feature is extracted, which is jointly combined with the image to input into the network. Finally, the subpixel convolutional layer is applied to produce a high-resolution image from the feature map. The framework of our method is shown in Figure 1; there are three phases: motion information estimation, information fusion, and reconstruction.

Optical flow is defined as the instant velocity of the pixels including those in moving objects [24–26]. For the video sequence, the optical flow extraction is realized by estimating the changes between two frames. The correlation between successive frames is estimated to identify the changes in the time axis. The optical flow feature is represented as the two-dimensional vector that demonstrates the ratio of the intensity change.

Assume there are L layers in the ESPCN network, the front $L-1$ layers can be expressed as follows:

$$\begin{aligned} f^1(\mathbf{I}^{\text{LR}}; W_1, b_1) &= \phi(W_1 * \mathbf{I}^{\text{LR}} + b_1), \\ f^l(\mathbf{I}^{\text{LR}}; W_{1:l}, b_{1:l}) &= \phi\left(W_l * f^{l-1}(\mathbf{I}^{\text{LR}}) + b_l\right). \end{aligned} \quad (1)$$

In which $W_l b_l$ are the weight and bias of each layer, respectively; the value of l is between 1 and $L-1$; W_l is a two-dimensional convolution tensor with size of $n_{l-1} \times n_l \times k_l \times k_l$; n_l is the feature dimension of the l th layer; the dimension of n_0 is the channel number C , k_l is the convolution kernel size of the l th layer; ϕ is the activate function; and the last layer of the network f^L maps the low-resolution image to the high-resolution image \mathbf{I}^{SR} .

The subpixel convolutional method is first proposed in ESPCN where the convolutional kernel is used to activate different parts of images. When shifting the kernel, the subpixels are periodically activated according to their locations which can be mathematically defined as follows:

$$\mathbf{I}^{\text{SR}} = f^L(\mathbf{I}^{\text{LR}}) = PS\left(W_L * f^{L-1}(\mathbf{I}^{\text{LR}}) + b_L\right), \quad (2)$$

where $f^L(\cdot)$ and $f^{L-1}(\cdot)$ are the mapping functions in the L and $L-1$ layers, SR and LR denote the terms of the superresolution and low-resolution, W_L is the kernel in the L layer, b_L is the moderation factor, and $PS(\cdot)$ is devoted to the periodical transformation which can map the feature map to the high-resolution image:

$$PS(T)_{x,y,c} = T_{\lfloor x/r \rfloor, \lfloor y/r \rfloor, C \cdot r \cdot \text{mod}(y,r) + C \cdot r \cdot \text{mod}(x,r) + c} \quad (3)$$

where x and y locate the pixels in the high-resolution images; the size of the kernel is $n_{L-1} \times r^2 C \times k_L \times k_L$. Since the nonlinear mapping layer is eliminated in the bottom of the deep convolutional network, the subpixel convolution is operated on the low-resolution image when $k_L = (k_s/r)$ and $\text{mod}(k_s, r) = 0$.

We use the mean squared error (MSE) as the training strategy. The computing method is as follows:

$$\ell(W_{1:L}, b_{1:L}) = \frac{1}{r^2 H W} \sum_{x=1}^{rH} \sum_{y=1}^{rW} \left(\mathbf{I}_{x,y}^{\text{HR}} - f_{x,y}^L(\mathbf{I}^{\text{LR}}) \right)^2. \quad (4)$$

Meanwhile, we used the tanh function as the activation function, by introducing nonlinearity into the network with the tanh function; the output of the upper network is mapped to the input of the lower network; and the expression ability of the neural network is enhanced. The tanh function is defined as

$$\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}. \quad (5)$$

tanh function is a kind of hyperbolic function with fast convergence.

Moreover, the subpixel convolution can eliminate the initial interpolation process, which can generate a satisfying reconstruction result with low cost. This architecture nicely adapts to the superresolution reconstruction of the video with a large number of frames.

4. Algorithm

The phases of the superresolution reconstruction are shown in Figure 2. Successive 5 frames are selected as the input for our deep convolutional network. Here, we define the time scale of the third frame as n , while the time scales of the previous two frames are $n-2$ and $n-1$; the two time scales of the last two frames are $n+2$ and $n+1$. The motion feature estimation is based on the frame n ; the optical flow calculation is operated on the frames $n-2$, $n-1$, $n+1$, and $n+2$ to generate the feature maps for these frames. Then, these feature maps are combined with the original frames to obtain

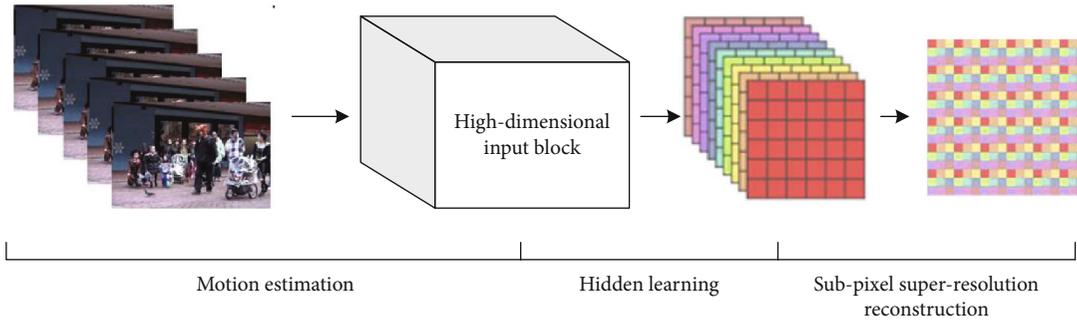


FIGURE 2: The phases of our algorithm.



FIGURE 3: The video frames in the training database: (a) news broadcast, (b) cars, (c) a patrol boat, (d) plants, and (e) a rugby game.

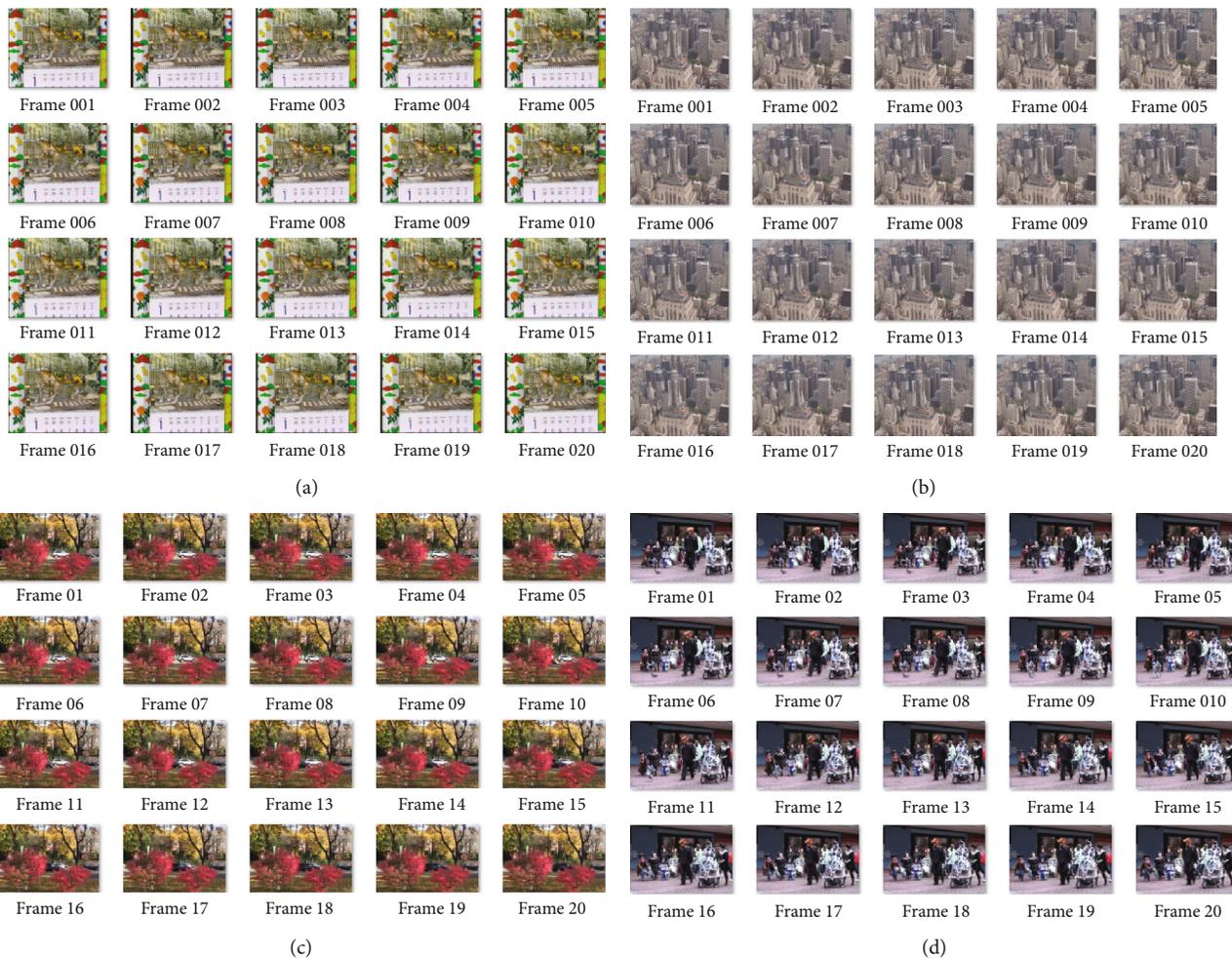


FIGURE 4: The testing database: (a) calendar, (b) building, (c) foliage, and (d) walk.



FIGURE 5: Superresolution reconstruction for the walk video: (a) original frame, (b) bicubic interpolation, (c) POCS, (d) sparse coding, (e) SRCNN, and (f) ours.

the input block of the deep convolutional network where the subpixel convolutional layer is applied to reconstruct the superresolution images [27].

The architecture of the deep convolutional network is constructed with a 4-layer structure that comprises 3 convolutional layers and 1 subpixel convolutional layer. We transfer the color space of the original frame from the RGB space to the Ycb space. The first layer comprises 64 kernels that are $5 \times 5 \times 15$, which generate the 64-channel feature maps. The

second layer comprises 32 kernels that are 3×3 , which generate the 32-channel feature maps. $3 \times r^2$ kernels that are of 3×3 are included in the third layer that generates the $3 \times r^2$ feature map. The superresolution reconstruction is realized by the last subpixel convolutional layer.

In specific, we assume that the size of frames is $H \times W$, where H is the height of the frame and W is the width. Accordingly, the dimension of the input RGB frames is $H \times W \times 3$. In our model, successive 5 frames are jointly

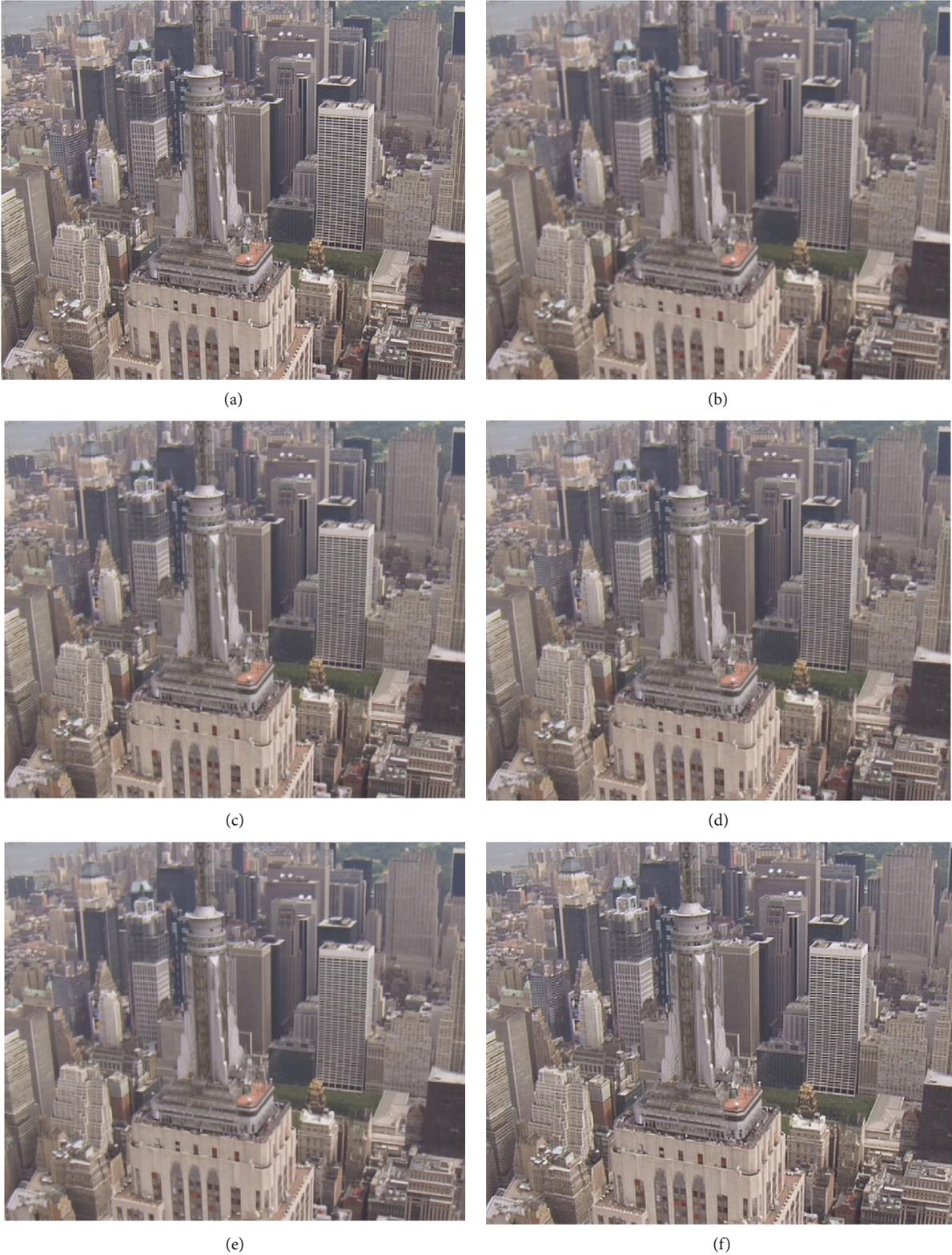


FIGURE 6: Superresolution reconstruction for the building video: (a) original frame, (b) bicubic interpolation, (c) POCS, (d) sparse coding, (e) SRCNN, and (f) ours.

considered to estimate the optical flow feature. As a result, the size of the input block is $H \times W \times (3 \times 5)$, and the size of the filter in the first layer is $f_1 \times f_1 \times 15$ and $f_1 = 5$. After

the mapping obtained by 4 hidden layers, $H \times W \times r^2$ feature maps are generated to further reconstruct the high-resolution images for the frame n .

5. Experimental Evaluations

5.1. Database. Training data is collected from the Xiph database which includes 10 video sequences with the length of 2695 frames [28]. The resolution of the frame is 144×176 . The training data shown in Figure 3 contains different contents in urban video surveillance, such as news broadcast, cars, a patrol boat, plants, and a rugby game. For training video sequences, the video segmentations are extracted and degenerated by adding the Gaussian noise and the downsampling processes, which can generate the low-quality and low-resolution samples. These samples are then inputted into the model. It is noticed that the input training data of the convolutional neural network is an image block composed of five consecutive frames after motion estimation. The output of the convolutional neural network is a high-resolution image. Meanwhile, the network generates the errors with the reference frame to update the network. Finally, the superresolution reconstruction model is going to be achieved after the network convergence.

The testing database shown in Figure 4 is given by the VideoSet4 database with 4 videos: calendar, buildings, foliage, and walk [29]. The building video includes 34 frames with the size of 704×576 ; the calendar video includes 41 calendar images with the size of 720×576 . In this training sample, the framework contains a large number of characters and regular textures, which can be evaluated by the details of edge reconstruction. Foliage video shows traffic scenes including vehicles and streets. The resolution of the original frame is 720×480 . Moreover, the contents of the foliage are quite complicated which can evaluate the robustness of our method against the dynamic noises. The walk video includes the walking pedestrians; the dataset of this video comprises 47 frames with the size of 720×480 . This dataset is characterized by slowly moving objects, thus can evaluate the performance of our method for motion estimation.

5.2. Parameter Setting. The input block is the combination of the motion estimation and original images, which is 15-dimensional. With respect to the network architecture, the first layer is constructed by 64 kernels with the size of $5 \times 5 \times 15$ and generates 64-dimensional feature maps. The second layer is constructed by 32 kernels which are applied to handle the feature map extracted by the first layer and generate 32-dimensional feature maps. The third layer comprises $3 \times r^2$ kernels with the size of 3×3 , which generates $3 \times r^2$ -dimensional feature maps. At last, the subpixel convolutional layer is used to reconstruct the superresolution images. The learning ratio is set as 0.001, the maximum epoch number is 100, and the batch size of the network is 32.

5.3. Experimental Comparisons. The comparison is divided into two parts. The first is comparison of traditional methods and the proposed method. The second is comparison between superresolution with motion estimation and without motion estimation. The data for experimental evaluation is selected from VideoSet4. The PSNR and time cost are selected as the metrics. The PSNR is calculated as follows:

TABLE 1: PSNR scores given by different methods (dB).

Methods	City	Calendar	Foliage	Walk
Bicubic interpolation	25.84	20.30	24.13	26.47
POCS	28.04	21.87	26.94	29.98
Sparse coding	28.53	22.09	27.33	30.58
SRCNN	30.23	23.39	28.78	32.43
Our method	30.44	23.59	29.00	32.61

TABLE 2: Improvement of PSNR (dB) value of video frame by motion estimation.

Methods	City	Calendar	Foliage	Walk
Bicubic interpolation	25.84	20.30	24.13	26.47
Our method	30.44	23.59	29.00	32.61
ESPCN	30.29	23.56	28.83	32.49

$$\text{PSNR} = 10 \log_{10} \left(\frac{\text{MAX}_I^2}{\text{MSE}} \right), \quad (6)$$

in which MAX is the image threshold and MSE is the mean square error.

5.3.1. The First Comparison. Traditional methods include the bicubic interpolation, POCS, and the sparse coding-based superresolution reconstruction methods, while the compared deep learning-based method is the SRCNN.

Figures 5 and 6 qualitatively present the results for frame #13 in the walk video and frame #8 in the building video. From the results, we can find that the reconstruction results obtained by traditional methods, i.e., sparse coding and POCS-based methods, are relatively more blurred in contrast to the results obtained by our methods. For example, our reconstructed images can clearly distinguish the details of the contour of the pedestrians and the texture of the baby carriage. Moreover, a significant advantage of our method can be demonstrated in Figure 5. In Figure 5, the windows of the building generate many complicated textures which cause periodic noise in the results obtained by compared methods. However, this effect is largely removed by our method due to the use of the motion features.

Table 1 quantitatively shows the PSNR scores obtained by different methods. From Table 1, we can find that our method generates better reconstruction results with respect to the PSNR scores. In contrast to the traditional reconstruction methods, the advantage of our method is that motion feature learning can improve the correctness of reconstruction. Although SRCNN and our methods are commonly based on convolutional operations, our method is the best. Different from the SRCNN, motion information is estimated and introduced in our method, which is the reason for our better performance.

5.3.2. The Second Comparison. We use the same video training set to train them. Table 2 shows the PSNR of four videos of VideoSet4 reconstructed by two different models. In the case of two times of magnification, it can be seen that the



FIGURE 7: Comparison of different reconstruction methods to reconstruct the 19th frame of foliage: (a) original frame, (b) LR images, (c) bicubic interpolation, (d) ESPCN, and (e) ours.

algorithm without motion estimation of adjacent frames is generally about 5 dB higher than bicubic interpolation. The average increase of our method is about 0.12 compared with SRCNN (without motion estimation).

Figure 7 shows the reconstruction quality of the 19th frame of the foliage sequence in VideoSet4 under different algorithms. From the results, it can be found that the biggest difference between the ESPCN model with motion estimation and the ESPCN model without motion estimation is that for moving objects, adding motion estimation

can produce better reconstruction effect, which is more obvious in high-speed moving objects. In Figure 7, the black car with the midview motion has obvious distortion and halo (Figure 7(d)). In Figure 7(e), in the reconstruction image using motion estimation, the edge of the black car is relatively straight and clear. This is because the new model takes into account the motion information before and after the frame, and it can better consider the temporal coherence of the video than the single frame image with single information.



FIGURE 8: Comparison of different reconstruction methods to reconstruct the 31st frame of calendar: (a) original frame, (b) LR images, (c) bicubic interpolation, (d) ESPCN, and (e) ours.

The same phenomenon is shown in Figure 8. On the 31st frame of the calendar video sequence, the characters in the bicubic interpolation image calendar in Figure 7(c) are blurred. In Figure 8(d), although the superresolution method of the single frame image improves the character edge to a certain extent, there is a certain degree of dislocation between smaller characters. In Figure 7(e), in the reconstructed image of the model combined with motion estimation, the edges of the smaller characters become more separated and the deformation degree is lower. The difference between foliage and calendar is that foliage mainly reflects the main motion of the object, while calendar reflects the relative motion of the object caused by the lens motion, which shows that the motion estimation has a certain effect on the two kinds of motion.

6. Conclusions

Aimed at improving the performance of the reconstruction of videos, this paper proposes deep convolutional network-based reconstruction methods where the motion information is extracted and introduced. Moreover, the proposed method introduces the subpixel convolution, which can significantly speed up the reconstruction process. Experimental results demonstrate that our method generates better reconstruction results in contrast to previous methods, the SRCNN, and the ESPCN. The proposed method can be applied to hardware architecture composed of integrated circuit chips such as digital signal processor (DSP) and programmable system on chip (SOC) in the form of embedded program. So, our method can be configured in the front-end device of video surveillance. In the future, it will be used as a form of edge intelligent and provide a feasible method for reducing the computing load of the centre system of urban computing, and we will evaluate the contribution of other types of motion information for superresolution reconstruction.

Data Availability

The data used to support the findings of this study are included within the article.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

This work is supported partly by the National Natural Science Foundation of China (Nos. 51979085 and 61903124), the Fundamental Research Funds for the Central Universities (No. B200202186), and the National Natural Science Foundation of China (No. 61563036).

References

[1] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar, "Fast and robust multiframe super resolution," *IEEE Transactions on Image Processing*, vol. 13, no. 10, pp. 1327–1344, 2004.

[2] N. K. Bose, M. K. Ng, and A. C. Yau, "A fast algorithm for image super-resolution from blurred observations," *EURASIP Journal on Advances in Signal Processing*, vol. 2006, no. 1, Article ID 035726, 2006.

[3] H. Chang, D. Y. Yeung, and Y. Xiong, "Super-resolution through neighbor embedding," in , Article ID I-IProceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004, vol. 1, Washington, DC, USA, June 2004.

[4] W. Yu and M. Zhang, "Super resolution reconstruction of video images based on improved glowworm swarm optimization algorithm," in *2018 IEEE 3rd International Conference on Image, Vision and Computing (ICIVC)*, pp. 331–335, Chongqing, China, June 2018.

[5] W. T. Freeman, T. R. Jones, and E. C. Pasztor, "Example-based super-resolution," *IEEE Computer Graphics and Applications*, vol. 22, no. 2, pp. 56–65, 2002.

[6] J. D. van Ouwerkerk, "Image super-resolution survey," *Image and Vision Computing*, vol. 24, no. 10, pp. 1039–1052, 2006.

[7] D. Glasner, S. Bagon, and M. Irani, "Super-resolution from a single image," in *2009 IEEE 12th International Conference on Computer Vision*, pp. 349–356, Kyoto, Japan, September 2009.

[8] S. Dai, M. Han, Y. Wu, and Y. Gong, "Bilateral back-projection for single image super resolution," in *Multimedia and Expo, 2007 IEEE International Conference on*, pp. 1039–1042, Beijing, China, July 2007.

[9] C. Fan, C. Wu, G. Li, and J. Ma, "Projections onto convex sets super-resolution reconstruction based on point spread function estimation of low-resolution remote sensing images," *Sensors*, vol. 17, no. 2, p. 362, 2017.

[10] G. K. Chantas, N. P. Galatsanos, and N. A. Woods, "Super-resolution based on fast registration and maximum a posteriori reconstruction," *IEEE Transactions on Image Processing*, vol. 16, no. 7, pp. 1821–1830, 2007.

[11] Z. Zhang, X. Wang, J. Ma, and G. Jia, "Super resolution reconstruction of three view remote sensing images based on global weighted POCS algorithm," in *2011 International Conference on Remote Sensing, Environment and Transportation Engineering*, Nanjing, China, June 2011.

[12] D. Wallach, F. Lamare, G. Kontaxakis, and D. Visvikis, "Super-resolution in respiratory synchronized positron emission tomography," *IEEE Transactions on Medical Imaging*, vol. 31, no. 2, pp. 438–448, 2012.

[13] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation," *IEEE Transactions on Image Processing*, vol. 19, no. 11, pp. 2861–2873, 2010.

[14] Z. Chen, Y. Zhang, C. Wu, and B. Ran, "Understanding individualization driving states via latent Dirichlet allocation model," *IEEE Intelligent Transportation Systems Magazine*, vol. 11, no. 2, pp. 41–53, 2019.

[15] C. Xu, "A novel recommendation method based on social network using matrix factorization technique," *Information Processing & Management*, vol. 54, no. 3, pp. 463–474, 2018.

[16] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8, Anchorage, AK, USA, June 2008.

[17] K. Umehara, J. Ota, N. Ishimaru et al., "Super-resolution convolutional neural network for the improvement of the image quality of magnified images in chest radiographs," in

Proceedings Volume 10133, Medical Imaging 2017: Image Processing, Orlando, Florida, USA, February 2017.

- [18] W. Shi, J. Caballero, F. Huszár et al., “Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1874–1883, Las Vegas, NV, USA, June 2016.
- [19] M. A. Talab, S. Awang, and S. A. M. Najim, “Super-low resolution face recognition using integrated efficient sub-pixel convolutional neural network (ESPCN) and convolutional neural network (CNN),” in *2019 IEEE International Conference on Automatic Control and Intelligent Systems (I2CACIS)*, pp. 331–335, Selangor, Malaysia, June 2019.
- [20] J. Kim, J. K. Lee, and K. M. Lee, “Accurate image super-resolution using very deep convolutional networks,” in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 1646–1654, Las Vegas, NV, USA, June 2016.
- [21] C. Ledig, L. Theis, F. Huszár et al., “Photo-realistic single image super-resolution using a generative adversarial network,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 4681–4690, Honolulu, HI, USA, July 2017.
- [22] B. Sun, J. Lu, and X. Wei, “Image super-resolution reconstruction based on deep residual network,” in *Proceedings of the 2018 2nd International Conference on Artificial Intelligence: Technologies and Applications (ICAITA 2018)*, Osaka, Japan, March 2018.
- [23] W. L. Zhong, *Research on Application of Convolution Neural Network in Video Super Resolution*, vol. 5, University of Electronic Science and Technology, 2018.
- [24] E. Ilg, N. Mayer, T. Saikia, M. Keuper, A. Dosovitskiy, and T. Brox, “Flownet 2.0: evolution of optical flow estimation with deep networks,” in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 2462–2470, Honolulu, HI, USA, July 2017.
- [25] X. Zhu, S. Zhang, R. Hu, Y. Zhu, and J. Song, “Local and global structure preservation for robust unsupervised spectral feature selection,” *IEEE Transactions on Knowledge and Data Engineering*, vol. 30, no. 3, pp. 517–529, 2018.
- [26] W. Zheng, X. Zhu, Y. Zhu, R. Hu, and C. Lei, “Dynamic graph learning for spectral feature selection,” *Multimedia Tools and Applications*, vol. 77, no. 22, pp. 29739–29755, 2018.
- [27] X. Y. Du, *Research on Super-Resolution Reconstruction Method Based on Deep Learning*, vol. 5, Hohai University, 2019.
- [28] P. V. Vu, ‘*New Approaches and a Subjective Database for Video Quality Assessment*’, *Doctoral Dissertation*, Oklahoma State University, 2015.
- [29] D. Li and Z. Wang, “Video superresolution via motion compensation and deep residual learning,” *IEEE Transactions on Computational Imaging*, vol. 3, no. 4, pp. 749–762, 2017.