applied sciences

MDPI

Article

BFRVSR: A Bidirectional Frame Recurrent Method for Video Super-Resolution

Xiongxiong Xue [1,2], Zhenqi Han [2], Weiqin Tong [1], Mingqi Li [2] and Lizhuang Liu [2,*]

[1] School of Computer Engineering and Science, Shanghai University, Shanghai 200444, China; xuexiongxiong2018@sari.ac.cn (X.X.); wqtong@shu.edu.cn (W.T.)
[2] Shanghai Advanced Research Institute, Chinese Academy of Sciences, Shanghai 201210, China; hanzq@sari.ac.cn (Z.H.); limq@sari.ac.cn (M.L.)
* Correspondence: liulz@sari.ac.cn

Received: 30 October 2020; Accepted: 4 December 2020; Published: 7 December 2020

Abstract: Video super-resolution is a challenging task. One possible solution, called the sliding window method, tries to divide the generation of high-resolution video sequences into independent subtasks. Another popular method, named the recurrent algorithm, utilizes the generated high-resolution images of previous frames to generate the high-resolution image. However, both methods have some unavoidable disadvantages. The former method usually leads to bad temporal consistency and has higher computational cost, while the latter method cannot always make full use of information contained by optical flow or any other calculated features. Thus, more investigations need to be done to explore the balance between these two methods. In this work, a bidirectional frame recurrent video super-resolution method is proposed. To be specific, reverse training is proposed that also utilizes a generated high-resolution frame to help estimate the high-resolution version of the former frame. The bidirectional recurrent method guarantees temporal consistency and also makes full use of the adjacent information due to the bidirectional training operation, while the computational cost is acceptable. Experimental results demonstrate that the bidirectional super-resolution framework gives remarkable performance and it solves time-related problems.

Keywords: video super-resolution; bidirectional; recurrent method; sliding window method

1. Introduction

Video super-resolution, which solves the problem of reconstructing high-resolution images from low-resolution images, is a classic problem in image processing. It is widely used in security, entertainment, video transmission, and other fields [1–3]. As compared with single image super-resolution, video super-resolution can use more information to output better high-resolution images, such as the feature information of adjacent frames. However, the reconstruction of video super-resolution images is generally difficult because of various issues, such as occlusion, adjacent frame information utilization, and computational cost.

With the rise of deep learning, video super-resolution has received significant attention from the research community over the past few years. The sliding window method and recurrent method are two of the latest state-of-the-art methods based on deep learning. Specifically, the sliding window video super-resolution (SWVSR) method solves this problem by combining a batch of low-resolution images to reconstruct a single high-resolution frame and divides the video super-resolution task into multiple independent super-resolution subtasks [4]. Each input frame is processed several times, which wastes calculations. In addition, the generation process is an independent subtask, which may reduce time consistency, resulting in flickering and artifacts. Unlike the SWVSR method, the recurrent video super-resolution (RVSR) method generates the current high-resolution image from the previous

high-resolution image, the previous low-resolution image, and the current low-resolution image [5,6]. Each input frame is processed once. The RVSR method is able to process video sequences of any length and enables the details of the video to be implicitly transmitted in longer video sequences. Insufficient use of information caused by the RVSR method leads to a correlation between image quality and time (as shown in Figure 1).

In short, video super-resolution methods still have the following problems: (a) The super-resolution network that uses the sliding window method has a high computational cost. Each frame of the image needs to be calculated $2N + 1$ times (window size $2N + 1$). (b) Direct access to the output of the previous frame helps the network generate a temporally consistent estimate for the next frame or previous frame. In the recurrent network, insufficient use of information leads to the correlation between image quality and time.

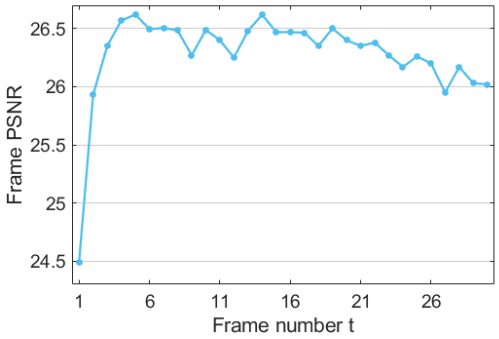

Figure 1. The recurrent video super-resolution (RVSR) method has a problem which is the correlation between time (frame number) and Peak Signal to Noise Ratio (PSNR) which is used to evaluate image quality.

In our work, we propose an end-to-end trainable bidirectional frame recurrent video super-resolution (BFRVSR) framework to address the above issues. We adopt forward training and reverse training to solve the problem of insufficient utilization of information and preserve temporal consistency, as shown in Figure 2. The BFRVSR has several benefits, which achieves a balance between RVSR and SWVSR. Each input frame needs to be processed no more than twice, while each output frame makes full use of the information contained by optical flow or any other calculated features. In addition, passing the previous high-resolution estimate directly to the other step helps the model to recreate fine details and produce temporally consistent videos. The work of the BFRVSR method is available at https://github.com/IlikethisID/BFRVSR.

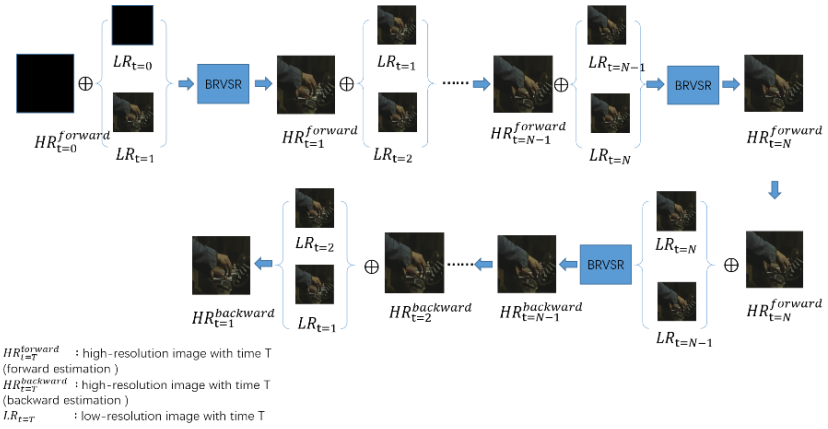

Figure 2. Overview of the proposed bidirectional frame recurrent video super-resolution (BFRVSR) framework. N frames are input as a group. In the group, each input is two frames of low-resolution and one frame of high-resolution, and the output is one frame of high-resolution. Forward estimation generates N frames of high resolution. Reverse estimation generates $N - 1$ frames of high resolution.

Our contributions are mainly reflected in the following: (a) Propose a bidirectional frame recurrent video super-resolution method, in which no pretraining step is required. (b) Address the correlation between image quality and time and preserve temporal consistency.

2. Related Work

With the rise of deep learning, computer vision, including image super-resolution and video security [7–9], have received significant attention from the research community over the past few years.

Image super-resolution (ISR) is a classic ill-posed problem. To be specific, in most cases, there are several possible output images corresponding to one given input image, thus, the problem can be seen as a task of selecting the most appropriate one from all the possible outputs. The methods are divided into interpolation methods such as nearest, bilinear, bicubic, and dictionary learning [10,11]; example-based methods [12–16]; and self-similarity approaches [17–20]. We refer the reader to three review documents [21–23] for extensive overviews of prior work up to recent years.

The recent progress in deep learning, especially in convolutional neural networks, has shaken up the field of ISR. Single image super-resolution (SISR) and video super-resolution are two categories based on ISR.

SISR uses a single low-resolution image to estimate a high-resolution image. Dong et al. [24] introduced deep learning into the field of super-resolution. They imitated the classic super-resolution solution method and proposed three steps, i.e., feature extraction, feature fusion, and feature reconstruction, to complete the SISR. Then, K. Zhang et al. [25] reached state-of-the-art results with deep CNN networks. A large number of excellent results have emerged [26–30]. In addition, the loss function also determines the result of image super-resolution, thus, some parallel efforts have studied the loss function [31–33].

Video super-resolution combines information from multiple low-resolution (LR) frames to reconstruct a single high-resolution frame. The sliding window method and recurrent method are two of the latest state-of-the-art methods.

The sliding window method divides the video super-resolution task into multiple independent subtasks, and each subtask generates a single high-resolution output frame from multiple low-resolution input frames [4,34–36]. The input is adjacent $2N + 1$ frames of low-resolution images like $\left\{I^{LR}_{t-N}, I^{LR}_{t-N+1} \ldots, I^{LR}_{t}, \ldots I^{LR}_{t+N-1}, I^{LR}_{t+N}\right\}$. Then, an alignment module is used to align $\left\{I^{LR}_{t-N}, I^{LR}_{t-N+1} \ldots I^{LR}_{t+N-1}, I^{LR}_{t+N}\right\}$ with the I^{LR}_{t}. Finally, I^{HR}_{t} is estimated through the aligned $2N + 1$ low-resolution frames. Drulea and Nedevschi et al. [29] used the optical flow method to align I^{LR}_{t-1} and I^{LR}_{t+1} with I^{LR}_{t} and used them to estimate I^{HR}_{t}.

The recurrent method generates a high-resolution image from the previous high-resolution image, the previous low-resolution image, and the low-resolution image. Huang et al. [37] used a bidirectional recurrent architecture but did not use any explicit motion compensation in their model. Recurrent structures are also used for other tasks, such as blurring [38] and stylization [39,40] of videos. Kim et al. [38] and Chen et al. [39] passed the feature representation to the next step, and Gupta et al. [40] passed the previous output frame to the next step, generating time-consistent stylizations in parallel work video. Sajjadi et al. [6] proposed a recursive algorithm for video super-resolution. The FRVSR [6] network estimates the optical flow $F^{LR}_{t\to t-1}$ of I^{LR}_{t-1} and I^{LR}_{t}, and uses I^{HR}_{t-1}. And $F^{LR}_{t\to t-1}$ to generate $\widetilde{I}^{HR}_{t}$, and finally, sends $\widetilde{I}^{HR}_{t}$ and I^{LR}_{t} to the network for reconstruction to obtain I^{HR}_{t}. However, insufficient use of information caused by FRVSR leads to the correlation between image quality and time.

3. Methods

The framework of BFRVSR is shown in Figure 2. All network modules can be replaced. For example, the optical flow module can use existing methods that have been pretrained instead of training and

building the network from scratch. You can also consider using a deformable convolution module [41] to replace the optical flow module.

After presenting an overview of the BFRVSR framework in Section 3.1, we define the loss functions used for training in Section 3.2.

3.1. Bidirectional Frame Recurrent Video Super-Resolution (BFRVSR)

The proposed model is shown in Figure 3. Trainable modules include the optical flow estimation network, i.e., FlowNet and the super-resolution network, i.e., SuperNet. The input of our model is the low-resolution image of the current frame I_t^{LR}, the low-resolution image of the previous frame I_{t-1}^{LR}, and the high-resolution image estimation of the previous frame I_{t-1}^{HR}. The output of our model is the high-resolution image estimation of the previous frame I_t^{HR}.

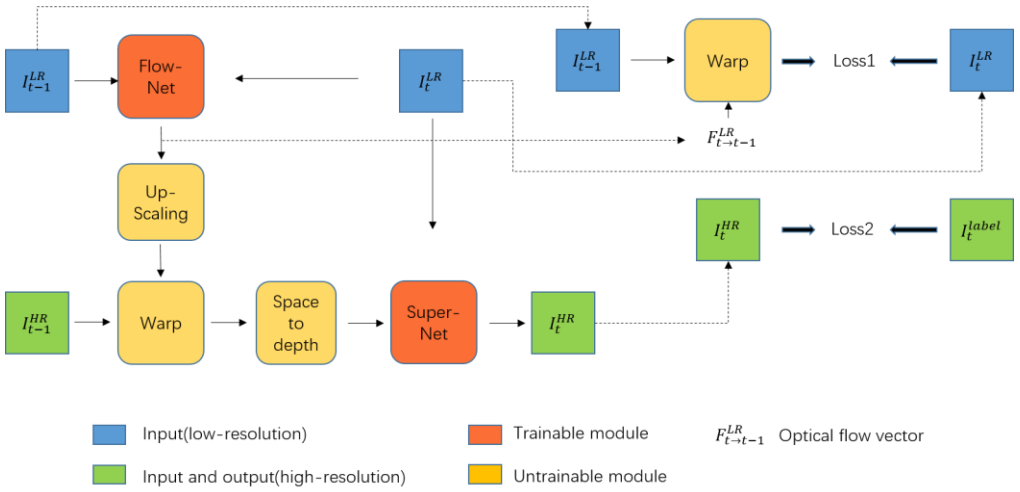

Figure 3. Overview of training network framework (**left**). Trainable modules include FlowNet and SuperNet. Upsampling uses bilinear interpolation. Loss function used during training (**right**).

3.1.1. Flow Estimation

The network structure of FlowNet is shown in the Figure 4. First, the network uses the optical flow estimation module to estimate the low-resolution image of the previous frame I_{t-1}^{LR} and the low-resolution image of the current frame I_t^{LR} to obtain a low-resolution motion vector diagram $F_{t\to t-1}^{LR}$. Our method of FlowNet is similar to the method in FRVSR [6].

$$F_{t\to t-1}^{LR} = FlowNet\left(I_{t-1}^{LR} \oplus I_t^{LR}\right) \in [-1,1]^{H\times W\times 2} \tag{1}$$

$F_{t\to t-1}^{LR}$ shows the position information from the current image to the previous frame.

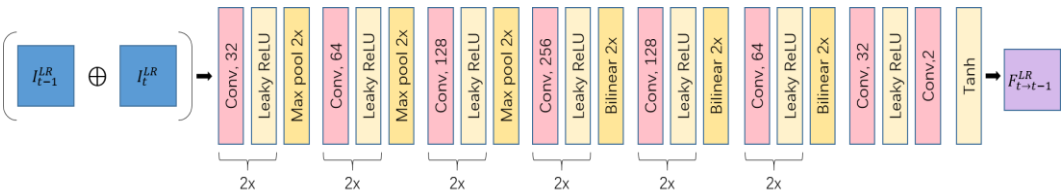

Figure 4. Overview of FlowNet. 2× represents the linear superposition of two identical modules.

3.1.2. Upscaling Flow

In this step, we process the low-resolution optical flow map that has been obtained, and we use bilinear interpolation with scaling factor s for upsampling to obtain the high-resolution optical flow map.

$$F^{HR}_{t\to t-1} = Upsample\left(F^{LR}_{t\to t-1}\right) \in [-1,1]^{sH\times sW\times 2} \tag{2}$$

3.1.3. Warping HR Image

Use the obtained high-resolution optical flow diagram and the high-resolution image of the previous frame to estimate the high-resolution image of the current frame.

$$\widetilde{I}^{HR}_{t} = Warp\left(I^{HR}_{t-1}, F^{HR}_{t\to t-1}\right) \tag{3}$$

We implemented warping as a differentiable function using bilinear interpolation similar to Jaderberg et al. [42].

3.1.4. Mapping to Low Resolution (LR) Space

We map high-dimensional spatial information to low-dimensional depth information using the space-to-depth transformation.

$$H^{depth}_{t} = DM\left(\widetilde{I}^{HR}_{t}\right) \tag{4}$$

Our method of mapping to low-dimensional space is similar to the method in FRVSR [6]. The mapping to LR space operation process is shown in the Figure 5.

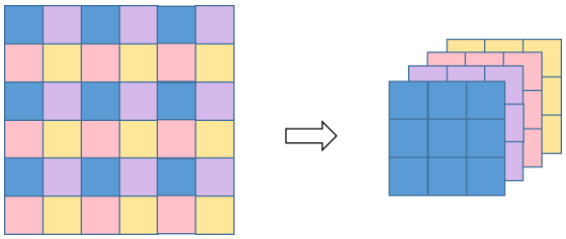

Figure 5. Space-to-depth module. Compress the spatial information of high-resolution images into low-resolution image depth information.

3.1.5. Super-Resolution

In this step, the low-dimensional depth map of the high-resolution image of the current frame H^{depth}_{t} and the low-resolution image of the current frame I^{LR}_{t} are sent to the SuperNet to obtain the final high-resolution frame. The network structure of SuperNet is shown in the Figure 6.

$$I^{HR}_{t} = SuperNet\left(H^{depth}_{t} \oplus I^{LR}_{t}\right) \tag{5}$$

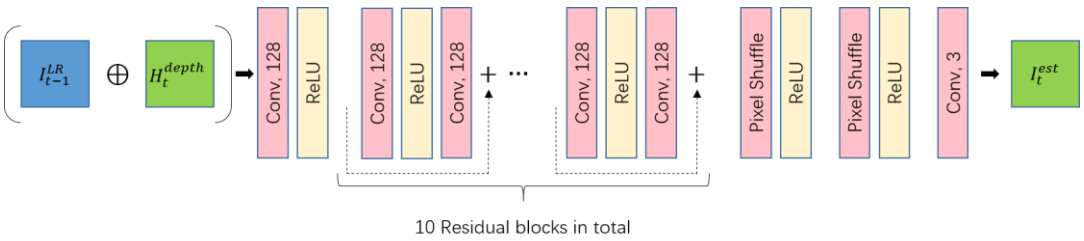

Figure 6. Overview of SuperNet. SuperNet uses the RESNET framework and Pixel Shuffle upsampling operation. SuperNet is an open framework and can be replaced by other networks.

In summary, the overall process of the network is as follows:

$$I_t^{HR} = SuperNet(DM(Warp(I_{t-1}^{HR}, Upsample\Big(FlowNet\Big(I_{t-1}^{LR} \oplus I_t^{LR}\Big)\Big)) \oplus I_t^{LR})) \quad (6)$$

3.2. Loss Functions

In our network architecture, the optical flow estimation module and the super-resolution module are trainable, therefore, in the training process, two loss functions are used to optimize the results.

The first loss function is the error between the high-resolution image generated by the super-resolution module and the real image label I_t^{lable} as follows:

$$L_1 = \left\| I_t^{HR} - I_t^{label} \right\|_2^2 \quad (7)$$

Because the dataset does not have the ground truth of optical flow, we use a method similar to the FRVSR [6] to calculate the spatial mean square error on the curved LR input frame to optimize the optical flow estimation module as the second loss function as follows:

$$L_2 = \left\| Warp\Big(I_{t-1}^{LR}, F_{t\to t-1}^{LR}\Big) - I_t^{LR} \right\|_2^2 \quad (8)$$

The loss function of training final backpropagation is $L_{total} = L_1 + L_2$.

4. Experiment

4.1. Training Datasets and Details

4.1.1. Training Datasets

Vimeo-90k [43] is our training and testing dataset. We abbreviate the Vimeo-90k test dataset as Vimeo-Test and the Vimeo-90k train dataset as Vimeo-Train. The Vimeo-90k dataset contains 91,701 7-frame continuous image sequences, and is divided into Vimeo-Train and Vimeo-Test. In Vimeo-Train, we randomly cropped the original 448 × 256 image to the 256 × 256 real label image. In order to generate LR images, we performed Gaussian blur and downsampling processing on the real label image and used a Gaussian blur with standard deviation σ = 2.0.

4.1.2. Training Details

Our network is end-to-end trainable, and there are no modules that need to be pretrained. The Xavier method is used for initialization. We train 600 epochs, and the batch size is 4; the optimizer uses Adam optimizer; and the initial learning rate is 10^{-4}, which is reduced by 0.1 times every 100 epochs. In a batch, each sample is 7 consecutive images. We conduct video super-resolution experiments at 4× factor.

In order to obtain the first high-resolution image I_1^{HR}, two methods can be used. In the first method, we set I_0^{HR} to a completely black image. This can force the network to learn detailed information from low-resolution images. In the second method, we upsample I_1^{LR} to I_1^{HR} through the bicubic interpolation method and estimate I_2^{HR} from $\left\{I_2^{LR}, I_1^{LR}, I_1^{HR}\right\}$. In order to compare with the RVSR method, we used the first method for experimentation.

4.2. Baselines

For a fair evaluation of the proposed framework on equal ground, we compare our model with the following three baselines that use the same optical flow and super-resolution networks:

SISR Only a single low-resolution image is used to estimate a high-resolution image without relying on timing information. The input is I_t^{LR} and the output is I_t^{HR}.

VSR Through $\left\{I_{t-1}^{HR}, I_{t-1}^{LR}, I_t^{LR}\right\}$, without the optical flow network estimation, relying on the learning space deformation ability of the convolution operation itself to obtain I_t^{HR}.

RVSR Through $\left\{I_{t-1}^{HR}, I_{t-1}^{LR}, I_t^{LR}\right\}$, with the optical flow network estimation, and then sent to SuperNet to obtain I_t^{HR}. The operation process is the same as the forward propagation in the BFRVSR network.

We ensure that the network model is consistent during the evaluation. The key parameters of the training parameters are the same. The initialization uses Xavier initialization, and the accelerator uses Adam optimizer. The initial learning rate is 10×10^{-4}, which is reduced to 0.1 times every 100 rounds. All networks are trained with the same training set, and the coefficient of Gaussian blur is 2.0.

4.3. Analysis

We train baselines and BFRVSR to convergence under the same parameter conditions. We compare and test the pretrained model on the Vimeo-Test. Table 1 shows the comparison image PSNR results of baselines and BFRVSR. As compared with baselines, our proposed framework has the best effect in continuous 7-frame video sequences, and it is 0.39 dB higher than the RVSR method. PSNR of BICUBIC and SISR is only related to current low-resolution images, and no correlation between high-resolution images. PSNR of VSR and RVSR has correlation between image quality and time. Because of motion compensation by optical flow network, the RVSR performance is better than the VSR.

Table 1. The PSNR index of the image generated by the five methods of BFRVSR, RVSR, video super-resolution (VSR), single image super-resolution (SISR), and BICUBIC are compared. As can be seen in the table, BFRVSR is an upgrade of RVSR, which has the best effect, and also overcomes the shortcomings of RVSR's unidirectional gain.

	Frame1	**Frame2**	**Frame3**	**Frame4**	**Frame5**	**Frame6**	**Frame7**	**Average**
BICUBIC	29.3057	27.3187	29.3173	29.3120	27.3087	27.3051	27.2900	27.3082
SISR	28.5332	28.5633	28.5240	28.5468	28.5523	28.5447	28.5593	28.5462
VSR	28.7632	29.4320	29.8012	29.8122	29.8310	29.9001	29.9212	29.6373
RVSR	29.0803	29.8807	30.1547	30.2898	30.3553	30.3980	30.3991	30.0797
BFRVSR (ours)	30.4772	30.4836	30.4833	30.4739	30.4670	30.4547	30.4145	30.4649

BFRVSR performs a forward estimation and a reverse estimation. The BRVSR is equivalent to an RVSR network in forward estimation. It transmits global detail information by using I_{t-1}^{HR} and performs timing alignment operations. However, there are some problems, that is, the details of I_j^{LR} cannot be obtained for I_i^{LR} to optimize the image ($i > j$). Reverse estimation solves this problem. Reverse estimation makes each frame implicitly use all the information to estimate the high-resolution image of the frame. Use the $\left\{I_t^{HR}, I_{t-1}^{LR}, I_t^{LR}\right\}$ to generate I_{t-1}^{HR}.

RVSR can be trained on video clips of any length. However, if the video clip is too long, RVSR has a problem which is the correlation between image quality and time. In fact, RVSR also has the problem on shorter video clips. BFRVSR solves this problem, as shown in Figure 7. BFRVSR has two processes, i.e., forward estimation and reverse estimation. I_{t-1}^{HR} is used to transmit global information and perform timing alignment operations. In the forward estimation, BFRVSR is equivalent to RVSR. The problem in forward estimation is obvious. When the generated video sequence is $\left\{I_1^{LR}, \ldots, I_i^{HR}, \ldots, I_j^{LR}, \ldots, I_N^{HR}\right\}$, the reference information generated by the forward estimation of I_t^{HR} is $\left\{I_1^{LR}, \ldots, I_{t-1}^{HR}\right\}$, not global information. Reverse estimation solves this problem. Reverse estimation makes each frame implicitly use all global information to estimate the high-resolution image of the frame by using $\left\{I_t^{HR}, I_{t-1}^{LR}, I_t^{LR}\right\}$ to generate I_{t-1}^{HR}. As shown in Figure 8, the result of reverse estimation is better than result of forward estimation.

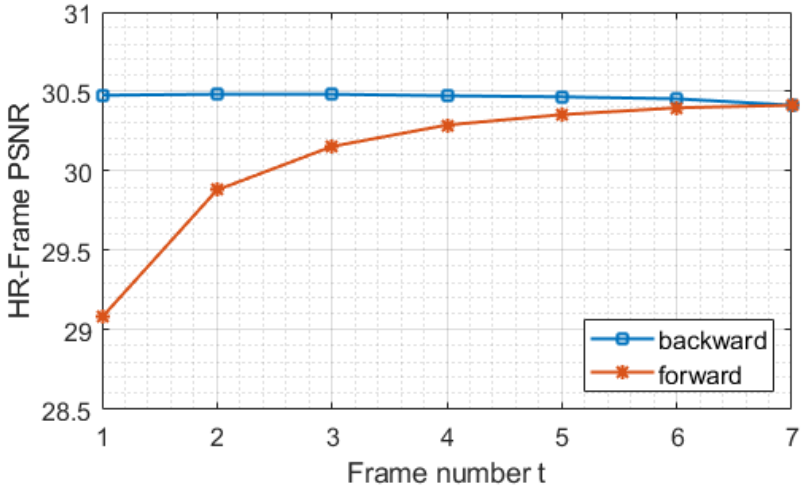

Figure 7. We show the quality of each frame in the forward propagation of BFRVSR and the quality of each frame in the reverse propagation. We found that global information is implicitly used in backpropagation to generate high-resolution images.

The video super-resolution, based on the sliding window method processes each frame $2N + 1$ times, the video super-resolution based on the recurrent method processes each frame once, and the BFRVSR processes each frame, at most, two times.

On the RTX-2080Ti, the time for a single image Full HD frame for 4× super-resolution is 291 ms.

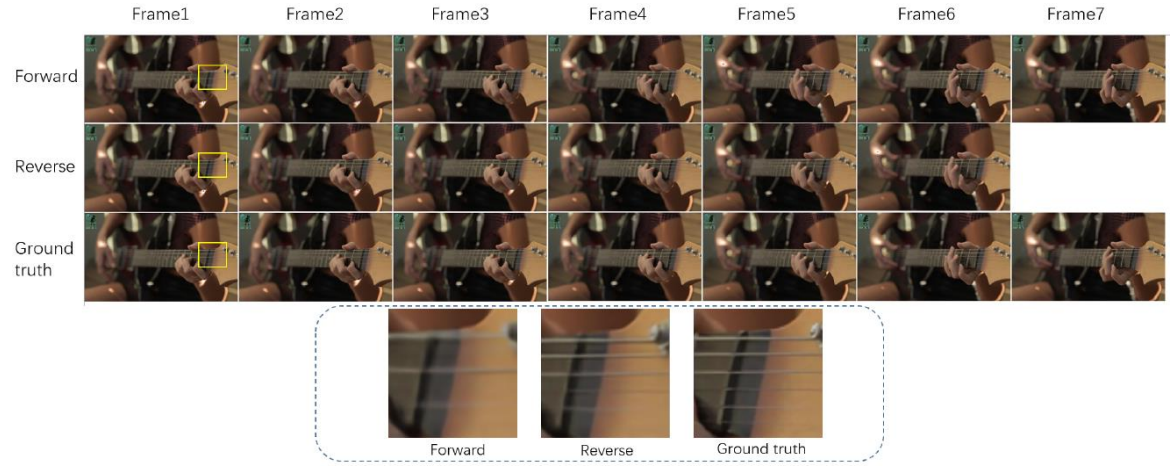

Figure 8. Visual comparison on Vimeo-Test. The image of reverse estimation is sharper and contains.

5. Conclusions

We propose an end-to-end trainable bidirectional frame recurrent video super-resolution method. Due to the operation of bidirectional training, with more information utilized to feed the model to deal with the correlation between image quality and time, BFRVSR successfully solves the problem shown in Figure 1. To be specific, it decouples the correlation between image quality and time. In addition, the proposed method achieves better image quality, while the computational cost is lower than the sliding window method.

6. Future Work

There is still room for improvement in the field of video super-resolution. If the problem of occlusion and blur is considered, much more computational cost would be required. We can deal with the problem by adding cross connections. In addition, a deformable convolution module, which has been frequently investigated recently, shows enormous potential in the field of image classification,

semantic segmentation, etc. Thus, it may achieve better results if we replace the optical flow module with a deformable convolution module. Furthermore, it is believed that video super-resolution and frame insertion have considerable similarities, thus, we may try to utilize BFRVSR to perform these two tasks simultaneously.

Author Contributions: Project administration, X.X.; Validation, X.X.; investigation, X.X., Z.H.; resources, W.T., M.L., L.L.; visualization, X.X., Z.H. All authors have read and agreed to the published version of the manuscript.

Funding: Supported by the National Natural Science Foundation of China (grant no. 61972007) and National Key R&D Program: Key Special Project for International Cooperation in Science and Technology Innovation between Governments (grant no. 2017YFE0192800).

Conflicts of Interest: The authors declare no conflict of interest. The founding sponsors had no role in the design of the study; in the collection, analysis, or interpretation of data; in the writing of the manuscript, and in the decision to publish the results.

References

1. Rajnoha, M.; Mezina, A.; Burget, R. Multi-frame labeled faces database: Towards face super-resolution from realistic video sequences. *Appl. Sci.* **2020**, *10*, 7213. [CrossRef]
2. Nam, J.H.; Velten, A. Super-resolution remote imaging using time encoded remote apertures. *Appl. Sci.* **2020**, *10*, 6458. [CrossRef]
3. Li, J.; Peng, Y.; Jiang, T.; Zhang, L.; Long, J. Hyperspectral image super-resolution based on spatial group sparsity regularization unmixing. *Appl. Sci.* **2020**, *10*, 5583. [CrossRef]
4. Wang, X.; Chan, K.C.K.; Yu, K.; Dong, C.; Loy, C.C. EDVR: Video restoration with enhanced deformable convolutional networks. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Los Angeles, CA, USA, 16–19 June 2019.
5. Tai, Y.; Yang, J.; Liu, X. Image super-resolution via deep recursive residual network. In Proceedings of the Internaltional Conference on Computer Vision and Pattern Recogintion, Hawaii, HI, USA, 21–26 July 2017.
6. Sajjadi, M.S.M.; Vemulapalli, R.; Brown, M. Frame-recurrent video super-resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–22 June 2018.
7. Wahab, A.W.A.; Bagiwa, M.A.; Idris, M.Y.I.; Khan, S.; Razak, Z.; Ariffin, M.R.K. Passive video forgery detection techniques: A survey. In Proceedings of the International Conference on Information Assurance & Security IEEE, Okinawa, Japan, 28–30 November 2014.
8. Bagiwa, M.A.; Wahab, A.W.A.; Idris, M.Y.I.; Khan, S.; Choo, K.-K.R. Chroma key background detection for digital video using statistical correlation of blurring artifact. *Digit. Investig.* **2016**, *19*, 29–43. [CrossRef]
9. Bagiwa, M.A.; Wahab, A.W.A.; Idris, M.Y.I.; Khan, S. Digital video inpainting detection using correlation of hessian matrix. *Malays. J. Comput. Sci.* **2016**, *29*, 179–195. [CrossRef]
10. Yang, J.; Wang, Z.; Lin, Z.; Cohen, S.; Huang, T. Coupled dictionary training for image super-resolution. *IEEE Trans. Image Process.* **2012**. [CrossRef]
11. Shi, W.; Caballero, J.; Huszar, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
12. Duchon, C.E. Lanczos filtering in one and two dimensions. *J. Appl. Meteorol.* **1979**, *18*, 1016–1022. [CrossRef]
13. Freedman, G.; Fattal, R. Image and video upscaling from local self-examples. *ACM Trans. Graph.* **2011**, *28*, 1–10. [CrossRef]
14. Freeman, W.T.; Jones, T.R.; Pasztor, E.C. Example-based super-resolution. *IEEE Comput. Graph. Appl.* **2002**, *22*, 56–65. [CrossRef]
15. Timofte, R.; Rothe, R.; Van Gool, L. Seven ways to improve example-based single image super resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
16. Yang, J.; Lin, Z.; Cohen, S. Fast image super-resolution based on in-place example regression. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Portland, Oregon, OR, USA, 23–28 June 2013.
17. Liu, C.; Sun, D. A bayesian approach to adaptive video super resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Colorado Springs, CO, USA, 21–25 June 2011.

18. Huang, J.-B.; Singh, A.; Ahuja, N. Single image super-resolution from transformed self-exemplars. In Proceedings of the Internaltional Conference on Computer Vision and Pattern Recogintion, Boston, MA, USA, 7–12 June 2015.
19. Makansi, O.; Ilg, E.; Brox, T. End-to-end learning of video super-resolution with motion compensation. In Proceedings of the Global Conference on Psychology Researches, Lara-Antalya, Turkey, 16–18 March 2017.
20. Ranjan, A.; Black, M.J. Optical flow estimation using a spatial pyramid network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Hawaii, HI, USA, 21–26 July 2017.
21. Anwar, S.; Khan, S.; Barnes, N. A Deep Journey into Super-resolution: A survey. *ACM Comput. Surv.* **2020**, *53*. [CrossRef]
22. Wang, Z.; Chen, J.; Hoi, S.C.H. Deep Learning for Image Super-resolution: A Survey. *IEEE Trans. Pattern Anal. Mach. Intell.* **2020**, 1. [CrossRef] [PubMed]
23. Nasrollahi, K.; Moeslund, T.B. Super-resolution: A comprehensive survey. *Mach. Vis. Appl.* **2014**, *25*, 1423–1468. [CrossRef]
24. Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a deep convolutional network for image super-resolution. In Proceedings of the Europeon Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014.
25. Zhang, K.; Zuo, W.; Chen, Y.; Meng, D.; Zhang, L. Beyond a gaussian denoiser: Residual learning of deep cnn for image denoising. *IEEE Trans. Image Process.* **2017**, *26*, 3142–3155. [CrossRef] [PubMed]
26. Lai, W.-S.; Huang, J.-B.; Ahuja, N.; Yang, M.-H. Deep laplacian pyramid networks for fast and accurate superresolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
27. Dong, C.; Loy, C.C.; He, K.M.; Tang, X.O. Image Super-Resolution Using Deep Convolutional Networks. *IEEE Trans. Pattern Anal. Mach. Intell.* **2016**, *38*, 295–307. [CrossRef] [PubMed]
28. Perez-Pellitero, E.; Salvador, J.; Ruiz-Hidalgo, J.; Rosenhahn, B. PSyCo: Manifold span reduction for super resolution. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016.
29. Drulea, M.; Nedevschi, S. Total variation regularization of local-global optical flow. In Proceedings of the International IEEE Conference on Intelligent Transportation Systems, Washington, DC, USA, 5–7 October 2011.
30. Tao, X.; Gao, H.; Liao, R.; Wang, J.; Jia, J. Detail-revealing deep video super-resolution. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
31. Yang, C.-Y.; Huang, J.-B.; Yang, M.-H. Exploiting selfsimilarities for single frame super-resolution. In Proceedings of the Asian Conference on Computer Vision, Queenstown, New Zealand, 8–12 November 2010.
32. Johnson, J.; Alahi, A.; Fei-Fei, L. Perceptual losses for realtime style transfer and super-resolution. In Proceedings of the European Conference on Computer Vision, Amsterdam, The Netherlands, 8–16 October 2016.
33. Ledig, C.; Theis, L.; Huszar, F.; Caballero, J.; Cunningham, A.; Acosta, A.; Aitken, A.P.; Tejani, A.; Totz, J.; Wang, Z.; et al. Photorealistic single image super-resolution using a generative adversarial network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017.
34. Milanfar, P. *Super-Resolution Imaging*; CRC Press: Boca Raton, FL, USA, 2010.
35. Tian, Y.P.; Zhang, Y.L.; Fu, Y.; Xu, C.L. TDAN: Temporally Deformable Alignment Network for Video Super-Resolution. In Proceedings of the Internaltional Conference on Computer Vision and Pattern Recogintion, Seattle, WA, USA, 16–20 June 2020.
36. Xiang, X.; Tian, Y.; Zhang, Y.; Fu, Y.; Allebach, J.P.; Xu, C. Zooming slow-mo: Fast and accurate one-stage space-time video super-resolution. In Proceedings of the Internaltional Conference on Computer Vision and Pattern Recogintion, Seattle, WA, USA, 16–20 June 2020.
37. Huang, Y.; Wang, W.; Wang, L. Bidirectional recurrent convolutional networks for multi-frame super-resolution. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 11–12 December 2015.
38. Kim, T.H.; Lee, K.M.; Scholkopf, B.; Hirsch, M. Online video deblurring via dynamic temporal blending network. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
39. Chen, D.; Liao, J.; Yuan, L.; Yu, N.; Hua, G. Coherent online video style transfer. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.

40. Gupta, A.; Johnson, J.; Alahi, A.; Fei-Fei, L. Characterizing and improving stability in neural style transfer. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
41. Dai, J.; Qi, H.; Xiong, Y.; Li, Y.; Zhang, G.; Hu, H.; Wei, Y. Deformable convolutional networks. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017.
42. Jaderberg, M.; Simonyan, K.; Zisserman, A.; Kavukcuoglu, K. Spatial transformer networks. In Proceedings of the Advances in Neural Information Processing Systems, Montreal, QC, Canada, 11–12 December 2015.
43. Xue, T.; Chen, B.; Wu, J.; Wei, D.; Freeman, W.T. Video enhancement with task-oriented flow. *Int. J. Comput. Vis.* **2019**, *127*, 1066–1125. [CrossRef]