ATJO: Adaptive three-dimensional joint optimization for remote sensing video super-resolution

Tian Qin, Lijing Bu*, Zhengpeng Zhang, Mingjun Deng, Yin Yang, Jingxue Wang, Xinyu Lan, Wenjuan Peng, Yang Hu
* Xiangtan University, Xiangtan, China
E-mail: lijingbu@xtu.edu.cn

***Abstract*— Existing deep learning-based video super-resolution reconstruction methods are difficult in ensuring the reconstruction requirements of satellite remote sensing low-resolution video due to their dependence on many datasets for training, and produce erroneous detail reconstruction during the reconstruction process. In this paper, we propose a nonparametric local adaptive super-resolution reconstruction method for satellite remote sensing video, which can implicitly and adaptively adjust the inter-frame motion error to achieve better super-resolution video reconstruction. The method calculates the grayscale difference between the high-resolution image and the low-resolution image from three dimensions: global pixel points of the high-resolution image, the number of video frames, and the neighborhood pixel points of the low-resolution image. In order to effectively measure the effect of different motion estimation errors on image reconstruction, we combine the intra- and inter-pixel gray differences and geometric distances between pixel blocks of high- and low-resolution images to form an adaptive weight matrix with Gaussian decay function. On this basis, we construct this loss function term to achieve significant noise reduction. For the edge details of the image, we employ an edge sampling mechanism that updates the BTV regularization term with the deformable convolution idea. Smooth reconstruction of high-frequency features at the edges is achieved by off-setting the sampling points to better capture the details. Experiments show that the method proposed in this paper significantly improves the metrics such as fidelity and PSNR of the recon-structed images and obtains better high-resolution reconstructed videos compared with deep learning methods and traditional methods.**

I. INTRODUCTION

With the rapid advancement of remote sensing technology and the increasing demand for satellite imaging applications, the need for high-quality, high-resolution image and video content—an essential aspect of multimedia—is becoming more prominent in various practical scenarios. Fields such as disaster monitoring, military reconnaissance, urban planning, and agricultural assessment benefit significantly from high-resolution satellite video, as it provides richer, more detailed information. In response to the requirement for high-quality reconstruction of satellite video, using practical algorithms to improve the spatial resolution of satellite video, especially under limited hardware conditions, has become an important area of current research. Currently used video super-resolution reconstruction techniques can be broadly classified into two categories: traditional methods based on mathematical models and deep learning methods.

Traditional reconstruction-based methods usually rely on accurate motion estimation to model the relative motion between video frames. Originally, interpolation was generally used to improve the visualization of image reconstruction results[1,2,3], and the drawbacks of scaling up to multi-frames are apparent. To improve the algorithm's flexibility, the MAP regularization model has gradually become the mainstream of traditional methods [4,5]. However, the complex scene changes, lighting conditions, and motion uncertainties in satel-lite videos make it difficult for conventional motion estimation algorithms to cope with them, leading to the degradation of re-construction quality. Especially when motion estimation errors accumulate, the final super-resolution results tend to be blurred or "trailing", affecting the accuracy and reliability of the recon-struction.

On the other hand, deep learning methods have made significant progress in learning the mapping relationship between low and high resolution through large-scale data. Recent studies, such as deep neural networks based on end-to-end neural networks and adaptive feature consolidation networks (AFC)[6,7], have qualitative improvements in image super-resolution reconstruction. Feature extraction considering simi-larities between images RefSR [8] has a significant advantage in recovering image details. However, image alignment often produces bias in dynamic scenes due to motion estimation errors [9,10,11]. Most deep learning methods cannot effectively handle inter-frame dependencies when processing multi-frame images, resulting in the loss of dynamic information, and image detail is not guaranteed.

Therefore, to target the reconstruction of remote sensing images, which is a high-precision and small-error demanded reconstruction goal under limited hardware resources, it is necessary to go for the design of a robust traditional reconstruction method more adapted to the needs of video reconstruction of satellite images in the popular trend of deep learning. Therefore, considering the edge-keeping constraints, we propose a video time-domain local neighborhood regularization model based on a nonparametric local adaptive algorithm for robust super-resolution reconstruction of motion scenes. The nonparametric local adaptive method we designed considers the similarities between local regions. When calculating the grayscale difference of sequence images frame by frame, we perform local image block segmentation and calculate the loss function of the grayscale difference between high-resolution images and low-resolution images block by block. Meanwhile, for different local regions, we construct weight matrices with Gaussian decay functions based on geometric distances and g-

979-8-3315-7206-8/25/$31.00

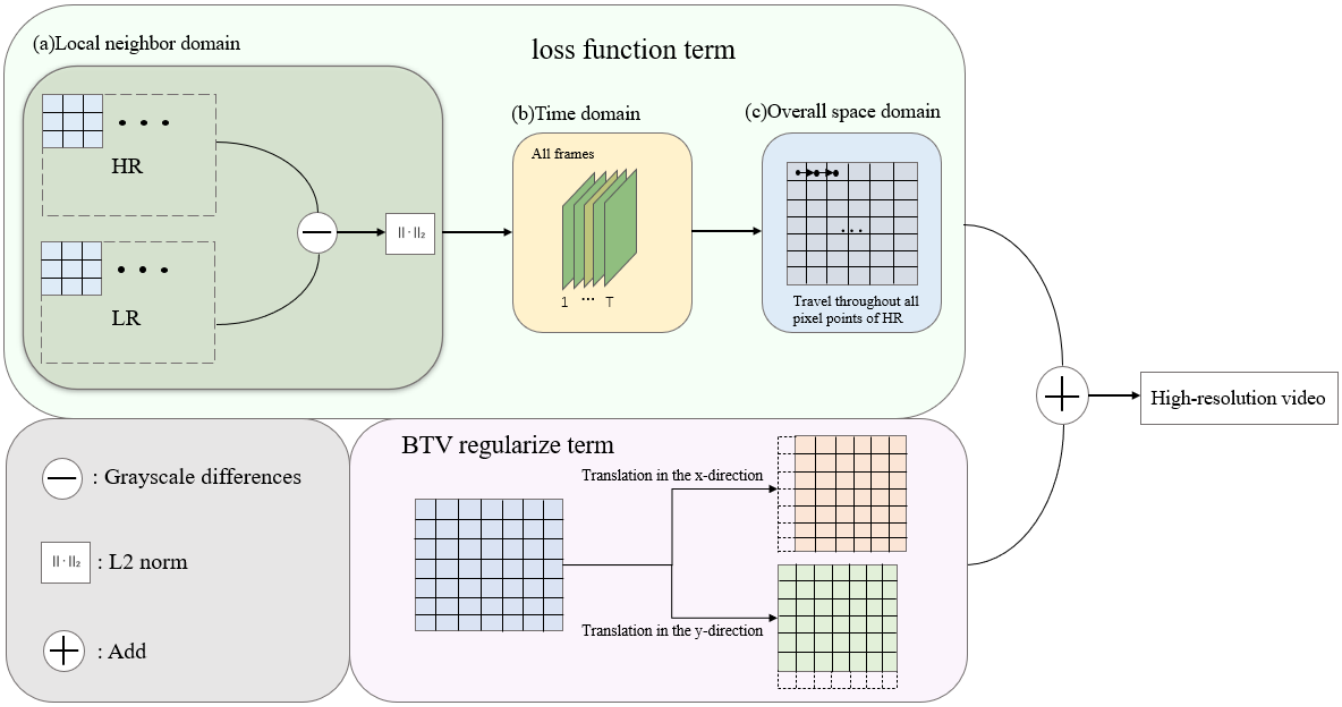

Figure 1: The overview of our method. (a) In the computation of the local neighborhood dimension, the gray-scale difference between the high-resolution image block and the low-resolution image block is computed; (b) In the time domain, the loss function term is computed frame-by-frame; (c) In the spatial domain, all pixel points of the high-resolution image are traversed. Calculate the BTV regularization term by translating it in the x and y directions.

ray differences and adaptively adjust the influence of local blocks of low-resolution images on the corresponding reconstruction effects for each frame. For edge regions, we draw on deformable convolution to sample the offset points and update the BTV regularization term to achieve local feature reconstruction. Our method transforms the super-resolution reconstruction problem into an optimization problem for MAP, and its objective function combines the loss function term and regularization term to achieve video super-resolution reconstruction of complex motion scenes while taking into account global consistency and local details.

II. METHODS

A. *Local image block difference - loss function term*

We introduce an additional reconstruction dimension on the local neighborhood, dividing the low-resolution input frames into blocks. We calculate the grayscale differences within the neighborhood blocks for edge pixels to adapt to the anisotropy of the ground object edges in satellite imagery. To eliminate the inter-frame offset caused by satellite hardware resource limitations, we compute the grayscale differences between the high-resolution and low-resolution images on a per-pixel block basis, considering three dimensions: global pixel points of the high-resolution image, video frame number, and neighborhood pixel points of the low-resolution image. Specifically, as shown in Figure , the image is first divided into multiple small pixel blocks. Then, the grayscale difference between the high-resolution and low-resolution images is calculated for each pixel block. This difference is computed progressively across three dimensions: local neighborhood comparison, video sequence frames, and global pixel points, forming a multi-dimensional grayscale difference analysis. This localized processing approach effectively captures image details while reducing the blurring effects that may arise from global processing.

$$F(X)=\sum_{(k,l)\in\Omega}\sum_{t\in[1,\cdots,T]}\sum_{(i,j)\in N^L(k,l)} w[k,l,i,j,t] \times\left\|D_p R_{k,l}^H BX-R_{i,j}^L Y_t\right\|_2^2 \tag{1}$$

Where matrices D and B represent the downsampling matrix and the blurring matrix, respectively. $R_{k,l}^H$ denotes the operation of extracting a high-resolution patch centered at pixel (k,l), while $R_{i,j}^L$ similarly represents the operation of extracting a low-resolution patch centered at pixel (i,j). Y_t refers to the input low-resolution image.

B. *Construction of adaptive weight matrix*

To effectively measure the effect of different motion estimation errors on image reconstruction, we constructed an adaptive weighting matrix by combining the gray level difference between the p image blocks of the high-resolution image and the low-resolution image and the geometric distance. The gray level difference of each facet and the geometric distance between facets are calculated, and the weights are adjusted using a Gaussian decay function. The resulting weight matrix can adaptively adjust the influence of the local aspects of each frame of the low-resolution image on the corresponding reconstruction effect, thus effectively reducing the negative impact of motion estimation errors on the reconstruction quality.

$$w[k,l,i,j,t]=\exp\left\{-\frac{\left\|D_p R_{k,l}^H BX-R_{l,j}^L Y_t\right\|}{r}\right\} \cdot f\left(\sqrt{(k-i)^2+(l-j)^2}\right) \tag{2}$$

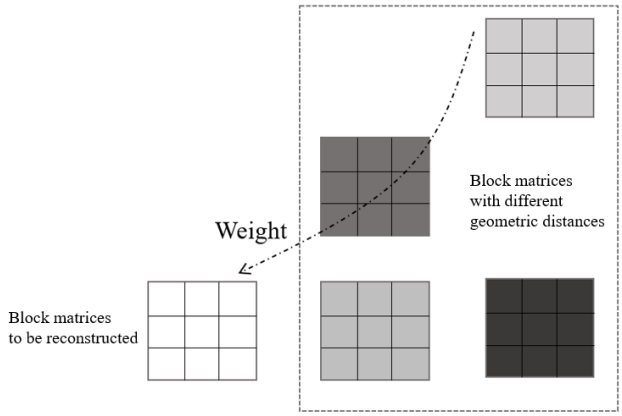

Figure 2: Schematic of the adaptive weight matrix. The shades of the matrix color indicate the size of the grayscale.

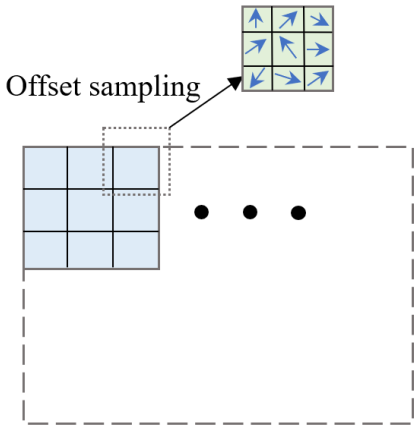

Figure 3: Offset point sampling schematic.

Where the parameters in $\left\|D_p R_{k,l}^H BX - R_{i,j}^L Y_t\right\|$ are related to the local patch grayscale differences described in the objective function. The δ_r parameter controls the impact of grayscale level differences between two pixels, while the function f is expressed as a Gaussian decay function based on geometric distance.

C. Edge sampling mechanism

The edge sampling mechanism enhances the detail recovery of ground object edges by intelligently selecting sampling points and applying a dynamic offset strategy, thereby improving the geometric measurement accuracy of the target. Traditional sampling methods uniformly cover the entire image, applying equal sampling to both edge and flat regions without specificity. Therefore, drawing inspiration from the offset point sampling concept in deformable convolution in deep learning, we have designed a deformable convolution-like sampling method. As shown in Figure 3, this method sets the primary sampling offset along the gradient direction, capturing the grayscale transition features on both sides of the edges. First, based on the edge and texture features of the image, we estimate the offset function using the image gradient. The gradient is calculated as follows:

$$G_x = \nabla_x I_{HR} = \frac{\partial I_{HR}}{\partial x} \tag{3}$$

$$G_y = \nabla_y I_{HR} = \frac{\partial I_{HR}}{\partial y} \tag{4}$$

Where G_x and G_y represent the gradients of the high-resolution image in the horizontal and vertical directions, respectively. Based on the gradients, the gradient magnitude and direction for each pixel are calculated as follows:

$$M = \sqrt{G_x^2 + G_y^2} \tag{5}$$

$$\theta = a\tan 2\left(G_y, G_x\right) \tag{6}$$

Where M is the gradient magnitude, and θ is the gradient direction. In edge regions, the gradient magnitude is more prominent, indicating significant image variations. Therefore, we can adjust the sampling point positions using the offset $\Delta p_{k,l}$. The specific estimation method is as follows:

$$\Delta p_{k,l,x} = \lambda_x \cdot \frac{G_x(k,l)}{M(k,l)}$$
$$\Delta p_{k,l,y} = \lambda_y \cdot \frac{G_y(k,l)}{M(k,l)} \tag{7}$$

Where λ_x and λ_y are the adjustment coefficients that control the size of the offset. $G_x(k,l)$ and $G_y(k,l)$ are the gradient values in the horizontal and vertical directions at the image block's position, respectively. $M(k,l)$ represents the gradient magnitude at this position, indicating the strength of the variation. The offset sampling operation is performed more accurately in the high-resolution image. For each pixel point (k,l), the new sampling position will be:

$$P_x' = k + \Delta p_{k,l,x}$$
$$P_y' = l + \Delta p_{k,l,y} \tag{8}$$

Where $\Delta p_{k,l,x}$ and $\Delta p_{k,l,y}$ represent the offset values in the horizontal and vertical directions, respectively.

D. BTV regularization term

For the edge details of the image, we introduce the term BTV (Binary Total Variation) regularization. The BTV regularization term uses the L1 norm to constrain the difference between the image and its translated version. It applies strong constraints in flat regions where the gradient is close to zero to suppress noise while preserving high-frequency information in edge regions with more significant gradients. The traditional expression for the BTV regularization term is:

$$Z_{BTV}(X) = \sum_{n=-p}^{p} \sum_{m=-p}^{p} \alpha^{|n|+|m|} \left\| X - s_x^n s_y^m X \right\|_1 \tag{9}$$

Table 1

Comparison of metrics with SOTA methods on the SATSOT dataset (4x). Our results are shown in bold. Advanced deep learning methods have stronger application capabilities. Our method still achieves stable high-resolution reconstruction in the 4x task, which outperforms all other methods in terms of metrics.

Video	Metrics	Bicubic	IBP	POCS	BasicVSR	TTVSR	HiRN	**Ours**
"car_52" from SATSOT	PSNR	29.56	33.97	23.75	27.70	32.49	33.58	**35.94**
	SSIM	0.8387	0.9117	0.7868	0.8185	0.9102	0.9190	**0.9471**
"plane_06" from SATSOT	PSNR	28.26	30.10	26.62	29.66	31.99	31.97	**34.75**
	SSIM	0.8158	0.8324	0.8278	0.8409	0.8709	0.8729	**0.9226**
"train_19" from SATSOT	PSNR	28.45	32.29	23.94	28.52	33.72	33.88	**34.84**
	SSIM	0.8649	0.8999	0.7598	0.8305	0.9268	0.9288	**0.9342**

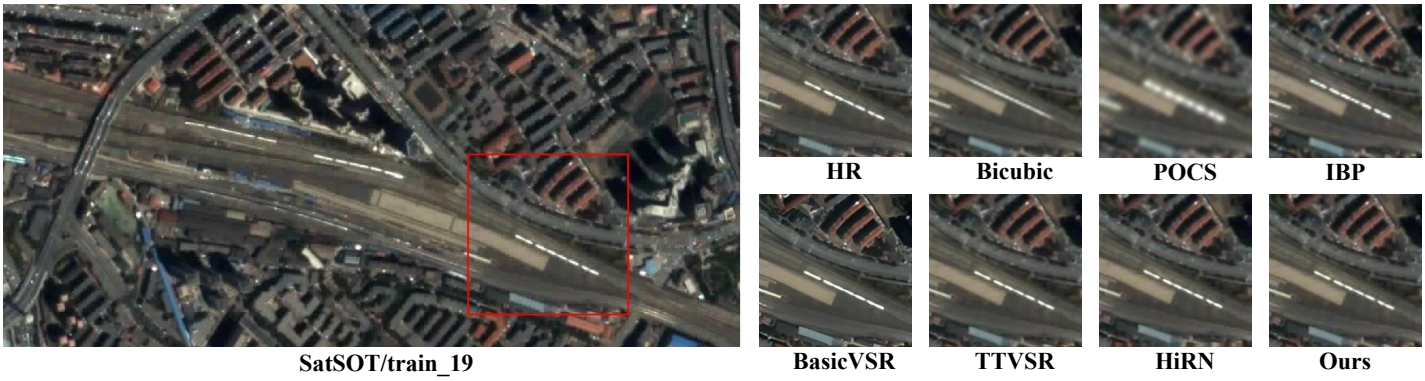

Figure 4: Comparison of our method with SOTA on the SatSOT dataset, the metrics are better than the other methods.

Where p represents the maximum translation amount, α is the scaling weight factor, with a range of [0,1], and s_x^n and s_y^m represent the horizontal and vertical pixel shifts, respectively. In our model, the BTV term introduces a decay factor to adjust the influence of pixels at different distances within the neighborhood, enabling more precise edge preservation. At the same time, the dynamic offset edge sampling mechanism allows the original s_x^n and s_y^m to be combined with the offset for more flexible offset processing. The updated regularization term is as follows:

$$Z_{BTV}{}'(X) = \lambda \sum_{n=-p}^{p} \sum_{m=-p}^{p} \alpha^{|n|+|m|} \| X - (s_x^n + \Delta p_{k,l,x}, s_y^m + \Delta p_{k,l,y}) X \|_1 \quad (10)$$

The offset affects the translation operation, allowing the regularization term to better account for the detail recovery in the edge regions. By incorporating the offset, the regularization term can dynamically adjust the influence of pixels near edges, leading to more accurate preservation of edge details and enhancing the overall reconstruction quality.

E. Optimization model and solution

We transform the super-resolution reconstruction problem into a MAP optimization problem. The objective function combines the fidelity and regularization terms, achieving video super-resolution reconstruction in complex motion scenes while considering global consistency and local details. Specifically, the objective function consists of two parts:

Table 2 : Comparison of metrics with adaptive weighting matrix removed.

Metrics	Weighting calculation method	
	Adaptive weighting matrix	Constant weighting of 1
PSNR	**34.75**	32.54
SSIM	**0.9226**	0.9007

Table 3: Metrics of gradient-based offsettable edge sampling mechanism vs. fixed sampling.

Sampling	PSNR	SSIM
Offset	**34.84**	**0.9342**
Fixed	31.98	0.9110

Loss Function Term: This includes an adaptive weight matrix used to measure the difference between the reconstructed and low-resolution images.

BTV Regularization Term: This term constrains the quality of the reconstructed image, preserving edge details and reducing noise.

By optimizing this objective function, we can effectively handle motion estimation errors while ensuring image quality, ultimately improving the reconstruction results.

$$\hat{X} = \arg\min_{X \in R} \left\{ \begin{array}{l} \sum_{(k,l)\in\Omega} \sum_{t\in[1,\ldots,T]} \sum_{(i,j)\in N^L(k,l)} w[k,l,i,j,t] \times \\ \| D_p R_{k,l}^H B X - R_{i,j}^L Y_t \|_2^2 \\ + \\ \lambda \sum_{n=-p}^{p} \sum_{m=-p}^{p} \alpha^{|n|+|m|} \| X - \\ (s_x^n + \Delta p_{k,l,x}, s_y^m + \Delta p_{k,l,y}) X \|_1 \end{array} \right\} \quad (11)$$

Where $\hat{X}$ is the reconstructed high-resolution image, λ is the regularization parameter, and $(s_x^n + \Delta p_{k,l,x}, s_y^m + \Delta p_{k,l,y})$ represents the offset sampling operation.

To solve the objective function, let:

$$\begin{aligned} C(X) = & \sum_{(k,l)\in\Omega} \sum_{t\in[1,\ldots,T]} \sum_{(i,j)\in N^L(k,l)} w[k,l,i,j,t] \times \\ & \| D_p R_{k,l}^H B X - R_{i,j}^L Y_t \|_2^2 \\ & + \\ & \lambda \sum_{n=-p}^{p} \sum_{m=-p}^{p} \alpha^{|n|+|m|} \| X - \\ & (s_x^n + \Delta p_{k,l,x}, s_y^m + \Delta p_{k,l,y}) X \|_1 \end{aligned} \quad (12)$$

To compute the gradient of Equation (12), we obtain Equation (13) as follows：

$$\begin{aligned} \nabla C(X) = & 2 \sum_{(k,l)\in\Omega} \sum_{i\in[1,\ldots,T]} \sum_{(i,j)\in N^L(k,l)} w(k,l,i,j,t) \times \\ & D_p^T R_{k,l}^T B^T \| D_p R_{k,l}^H B X - R_{i,j}^L Y_t \|_2^2 \\ & + \\ & \lambda \sum_{n=-p}^{p} \sum_{m=-p}^{p} \alpha^{|n|+|m|} (I - (s_y^{-m} + \Delta p_{k,l,y}, s_x^{-n} + \Delta p_{k,l,x})) \\ & sign(X - (s_x^n + \Delta p_{k,l,x}, s_y^m + \Delta p_{k,l,y}) X) \end{aligned} \quad (13)$$

To minimize the objective function, we use the gradient descent method for optimization. The update rule is defined as:

$$X^{(t+1)} = X^{(t)} - \eta \nabla X^{(t)} \quad (14)$$

Where $X^{(t)}$ is the model output at the t-th iteration, and η is the learning rate.

III. EXPERMENTS

A. Comparison of methods

We choose representative traditional methods such as Bicubic [12], IBP [13], POCS [14] and advanced deep learning methods such as TTVSR [15], RealBasic VSR [16], HiRN [17] to compare with our method on the widely used satellite video dataset SatSOT to demonstrate the effectiveness of our method. During the experiments, we select multi-frame video frames containing moving objects, downsample the original video frames, and add a fuzzy kernel to simulate low-resolution videos. To verify the application potential of the method, the input low-resolution video frames are reconstructed with downsampling factors (4x) for the corresponding super-resolution. Also, to demonstrate the ability of our method to capture inter-frame dependencies, we perform experiments on equally spaced frames with significant offsets. The robustness of this capability will be further verified by consecutive frame experiments, the results of which are shown in the Supplementary Material. PSNR and SSIM are used as evaluation metrics for SR performance.

Our method divides the low-resolution image into pixel blocks of 3*3 pixel size, and the geometric distance and gray level difference between the high-resolution image block and the low-resolution image block are computed for weighting matrix and loss value summation. To better capture the shape and size of the object, the BTV regularization term is computed by combining the deformable-like convolution edge sampling mechanism. The gradient of the loss function term with the BTV regularization term is calculated and updated using the gradient descent method for 100 iterations to obtain high-resolution video.

As shown in Table 1, our method cannot only effectively restore the shape of moving objects without non-existent de-tailings in terms of reconstructed image fidelity but also vastly outperforms other methods in terms of distortion metrics PSNR and SSIM. In the 4x large-scale zoom task, our method achieved the best results, with an average improvement of 2.03 dB over other methods. As shown in Figure 4, despite the superiority of some traditional methods, there is apparent blurring or even invisibility of moving objects in visual perception. Accordingly, methods such as BasicVSR [16] have superior de-blurring effects but pursue excessive smoothing, which is not conducive to the high realism of satellite video and affects our judgment. Our method, which is specifically designed to address the requirements of satellite videos for realism and high-precision reconstruction, is equally effective in everyday videos, not only with excellent distortion metrics but also by reducing artifacts and false textures in the video reconstruction of motion scenes, thus ensuring consistency with the original image and pixel accuracy of the reconstruction.

B. Ablation Study

Adaptive weight matrix. Keeping the sampling mechanism, the BTV regularization term unchanged, we remove the adaptive weight matrix (so that it is constant to 1); that is, the calculation of the loss function term of the grayscale difference only considers the direct gray scale difference between the high-resolution face sheet and the low-resolution face sheet. The effects of different distances and different gray scales are the same. The results in Table 2 show that after removing the adaptive weight matrix, the PSNR and SSIM indexes decrease significantly, by 2.21 dB and 0.0219, indicating that the introduction of the adaptive weight matrix can effectively deal with the impact of pixels with different geometric distances and different gray differences on the reconstructed points, and reduce the part of the interference of the reconstruction effect with a significant motion estimation error and a slight correlation.

Edge Sampling Mechanism. We replaced the offset point sampling with fixed position sampling without offset processing, i.e., baseline BTV regularization was used.

According to Table 3, the PSNR and SSIM metrics are decreased in terms of metrics by 2.86 dB and 0.0232. Meanwhile, the offset point sampling method is visually better than normal sampling for processing edge features and restoring object shape. This is because by offsetting, the model can consider the specific shape of the object ignored by the image block delineation, which leads to a better feature reconstruction of the object edges.

IV. CONCLUSIONS

This paper proposes a non-parametric local adaptive video super-resolution reconstruction algorithm to reconstruct satellite remote sensing image videos fully. We design a loss function term based on temporal, global pixel, and local neighborhood gray level differences with an updated BTV regularization term that introduces offset point sampling to achieve high-frequency detail restoration with temporal consistency. It is worth noting that the adaptive weight matrix we develop can effectively handle the intra-block effects of different gray level differences at different geometric distances after image block segmentation. At the same time, the proposed edge sampling mechanism has advantages in edge reconstruction and object morphology consistency preservation. Many experiments have proved that our method can efficiently reconstruct videos with limited data resources, especially in meeting the authenticity requirements of satellite images. Future research can explore increasing the computational efficiency of the model for more tasks.

V. ACKNOWLEDGEMENT

This research was funded by the National Key R&D Program of China (grant number 2020YFA0713503), Hunan Provincial Department of Education Hunan Provincial Teaching Reform Research Project for Regular Higher Education Institutions (grant number HNJG-20230279) and Hunan Provincial Department of Education Scientific Research Project (grant number 23C0059).

REFERENCES

[1] Sanchez-Beato, A., and Pajares, G. 2008. Noniterative interpolation-based super-resolution minimizing aliasing in the reconstructed image. IEEE Transactions on Image Processing, 17(10): 1817–1826.

[2] Nasonov, A. V., and Krylov, A. S. 2010. Fast super-resolution using weighted median filtering. In Proceedings of the International Conference on Pattern Recognition, IEEE Computer Society, 2230–2233.

[3] Lin, S. C., and Chen, C. T. 2007. Reconstructing vehicle license plate image from low-resolution images using nonuniform interpolation method. International Journal of Image Processing, 1(2): 21–28.

[4] Marquina, S., and Osher, J. 2008. Image super-resolution by TV-regularization and Bregman iteration. Journal of Scientific Computing, 37: 367–382.

[5] Gilboa, G., and Osher, S. 2006. Nonlocal linear image regularization and supervised segmentation. ACM Transactions on Graphics, 25(3): 761–768.

[6] Arefin, M. R., et al. 2020. Multi-Image Super-Resolution for Remote Sensing using Deep Recurrent Networks. In Proceedings of the 2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 816–825. Seattle, WA, USA.

[7] Mehta, N., Dudhane, A., Murala, S., Zamir, S. W., Khan, S., and Khan, F. S. 2022. Adaptive Feature Consolidation Network for Burst Super-Resolution. In Proceedings of the 2022 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), 1278–1285. New Orleans, LA, USA.

[8] Zhang, Z., Wang, Z., Lin, Z., and Qi, H. 2019. Image super-resolution by neural texture transfer. In Proceedings of the 2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR), 7974–7983. Long Beach, CA, USA. https://doi.org/10.1109/CVPR.2019.00817.

[9] Lertrattanapanich, S., and Bose, N. K. 1999. Latest results on high resolution reconstruction from video sequences. [EB/OL]. [2022-02-08].

[10] Bare, B., Yan, B., Ma, C. X., and Li, K. 2019. Real-time video super-resolution via motion convolution kernel estimation. Neurocomputing, 367: 236–245.

[11] Bao, W. B., Lai, W. S., Zhang, X. Y., Gao, Z. Y., and Yang, M. H. 2019. MEMC-Net: Motion estimation and motion compensation driven neural network for video interpolation and enhancement. IEEE Transactions on Pattern Analysis and Machine Intelligence, 43(3): 933–948.

[12] Schiemenz, S., and Hentschel, C. 2010. Scalable high-quality nonlinear up-scalar with guaranteed real-time performance. In Proceedings of the 14th IEEE International Symposium on Consumer Electronics (ISCE), 1–6. Braunschweig, Germany, June 7–10.

[13] Nayak, R., and Patra, D. 2018. Enhanced iterative back-projection based super-resolution reconstruction of digital images. Arabian Journal for Science and Engineering, 43: 7521–7547. https://doi.org/10.1007/s13369-018-3150-1.

[14] Stark, H., and Oskoui, P. 1989. High-resolution image recovery from image-plane arrays using convex projections. Journal of the Optical Society of America A, 6(11): 1715.

[15] Liu, C., Yang, H., Fu, J., and Qian, X. 2022. Learning trajectory-aware transformer for video super-resolution. arXiv preprint, arXiv:2204.04216. https://doi.org/10.48550/arXiv.2204.04216.

[16] Chan, K. C. K., Wang, X., Yu, K., Dong, C., and Loy, C. C. 2021. BasicVSR: The search for essential components in video super-resolution and beyond. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 4945–4954. Nashville, TN, USA.

[17] Choi, Y. J., and Kim, B. G. 2023. HiRN: Hierarchical recurrent neural network for video super-resolution (VSR) using two-stage feature evolution. Applied Soft Computing, 143, C (Aug. 2023).