Video Super-Resolution

Subjects: Computer Science, Artificial Intelligence Contributor: Yonggui Zhu , Guofang Li

Super-resolution (SR) refers to yielding high-resolution (HR) images from corresponding low-resolution (LR) images. As a branch of this field, video super-resolution (VSR) mainly utilizes the spatial information of the frame and the temporal information between neighboring frames to reconstruct the HR frame.

video super-resolution Super-resolution (SR) high-resolution (HR)

1. Introduction

At present, VSR encompasses non-blind VSR ^[1], blind VSR ^[2], online VSR ^[3], and other branches ^[4], and is widely used in remote sensing ^{[5][6]}, video surveillance ^{[7][8]}, face recognition ^{[9][10]}, and other fields ^{[11][12]}. At present, with the development of technology, the resolutions of videos are gradually increasing. Although this can enrich our lives and facilitate tasks such as surveillance and identification, it can put more pressure on areas such as video storage and transmission. In addressing these issues, VSR technology plays an important role. However, VSR is an ill-posed problem, and it is difficult to find the most appropriate reconstruction model. Thus, it remains a worthwhile endeavor to continue to explore VSR technology.

To obtain high-guality images, previous studies have proposed numerous effective methods. Initially, researchers utilize interpolation methods to obtain HR videos [13][14]. These methods possess higher computing speeds, but the results are poor. With the development of deep learning, constructing models [15][16][17] in different domains with deep learning has become a mainstream research method. Researchers have constructed different VSR models based on deep learning that can reconstruct high-quality videos. For example, researchers [18][19][20][21] have utilized explicit or implicit alignments to explore temporal flow between frames. This type of methodology can effectively align adjacent frames to the reference frame to extract high-quality temporal information. However, the alignment feature increases the computational effort of the model, thus exacerbating the burden during model training and testing. Meanwhile, inaccurate optical flow often leads to errors in alignment, which affects the performances of models. Moreover, scholars ^{[22][23][24]} have used 3D convolution or deformable 3D convolution to directly aggregate spatio-temporal information between different frames. Although this approach can quickly aggregate information from different times, it also incorporates a lot of temporal redundancy in features, which reduces the reconstruction ability of the model. In addition, in recent years, with the rise of Transformer, the application of Transformer to construct VSR models has also become a very popular research topic. Researchers [25][26][27] have applied Transformer to analyze and acquire the motion trajectories of videos to sufficiently aggregate the spatio-temporal information between consecutive frames. However, due to the relatively high level of computation required by Transformer, the further development of Transformer in the field of VSR is limited.

2. Video Super-Resolution

2.1. Single-Image Super-Resolution

Single-image super-resolution (SISR) is the basis of super-resolution. In recent years, with the development of deep learning, SR has ushered in a new revolution. Dong et al. ^[28] were the first to apply deep learning to SISR. They presented a three-layer convolution neural network and achieved a better effect. For example, when the review metric is peak signal-to-noise ratio (PSNR) and 4×SISR is performed, it outperforms the then state-of-theart A+ algorithm [33] with 0.21 dB and 0.18 dB on Set5 and Set14 datasets, respectively. It was thus proven that deep learning possesses great potential in the field of SR. After this paper, Kim et al. [34] presented a very deep neural network and applied the residual network to the SR model, achieving a better effect than SRCNN. For example, when performing 4× SISR and using PSNR as a metric for evaluation, it outperforms SRCNN with 0.87 dB and 0.52 dB on Set5 and Set14 datasets, respectively. Song et al. [35] came up with the idea of making use of the additive neural network for SISR, which replaced the traditional convolution kernel multiplication operation in the calculation of output layer. Experiments demonstrate that this additive neural network achieves performance and visual quality comparable to convolutional neural networks, while reducing energy loss by approximately 2.5 times when reconstructing a 1280×720 image. Liang et al. ^[29] introduced Swin Transformer into SISR and obtained high-quality recovered images. Tian et al. [30] proposed heterogeneous grouping blocks to enhance the internal and external interactions of different channels to obtain rich low-frequency structural information. In practice, Lee et al. ^[31] applied the SR technique to the satellite synthetic aperture radar, and could effectively recover the information of scatterers. Moreover, many scholars have also constructed SISR models using methods such as GAN or VAE, etc. [31][32][33][34][35]. Although the SISR model can also be used to reconstruct HR videos, the SISR model is only capable of capturing the spatial information of frames, and can not aggregate the temporal information between neighboring frames. As a result, the quality of the video recovered by the SISR is poor, while often suffering from artifacts and other problems. To reconstruct high-quality HR videos, researchers have shifted their focus to VSR models.

2.2. Video Super-Resolution

VSR is an extension of SISR. In VSR, the temporal information between adjacent frames play a vital role. To reconstruct high-quality HR frames, studies have built a variety of models. For instance, Caballero et al. ^[19] applied the optical flow field, which included coarse flow and fine flow to align adjacent frames, and constructed an end-toend spatio-temporal module. Based on ^[19], Wang et al. ^[36] combined an optical flow field and long short-term memory to make more efficient use of inter-frame information and obtain more realistic details. Moreover, Tian et al. ^[37] presented the first model to substitute the deformable convolution into VSR, which amplified the feature extraction ability of the model. Based on ^[37], Wang et al. ^[20] proposed a pyramid, cascading, and deformable (PCD) module that further enhances the alignment capability of the model. Then, Xu et al. ^[38] designed a temporal modulation block to modulate the PCD module. Meanwhile, they conducted short-term and long-term feature fusion to better extract motion clues. These optical flow-based methods have also been applied to practical work such as video surveillance, etc. Guo et al. ^[8] utilized optical flow and other methods to construct the back-projection network, which can effectively reconstruct high-quality surveillance videos. Moreover, Isobe et al. ^[22] proposed the structure of intra-group fusion and inter-group fusion, and used 3D convolution to capture and supplement the spatio-temporal information between different groups. Ying et al. ^[23] proposed deformable 3D convolution with efficient spatio-temporal exploration and adaptive motion compensation capabilities. Fuoli et al. ^[39] devised a hidden space propagation scheme that effectively aggregates temporal information over long distances. Based on ^[39], Isobe et al. ^[40] explored the temporal differences between LR and HR space, effectively complementing the missing details in LR frames. Then, Jin et al. ^[5] used the temporal difference between long and short frames to achieve information compensation for satellite VSR. Liu et al. ^[26] designed a trajectory transformer that analyzes and utilizes motion trajectories between consecutive frames to obtain high-quality HR videos. Then, on the basis of ^[26], Qiu et al. ^[27] introduced the frequency domain into the VSR domain, which provided a new basis upon which to study VSR.

References

- 1. Zhang, W.; Zhou, M.; Ji, C.; Sui, X.; Bai, J. Cross-Frame Transformer-Based Spatio-Temporal Video Super-Resolution. IEEE Trans. Broadcast. 2022, 68, 359–369.
- Pan, J.; Bai, H.; Dong, J.; Zhang, J.; Tang, J. Deep Blind Video Super-resolution. In Proceedings of the IEEE/CVF International Conference on Computer Vision, Montreal, BC, Canada, 10–17 October 2021; pp. 4791–4800.
- 3. Xiao, J.; Jiang, X.; Zheng, N.; Yang, H.; Yang, Y.; Yang, Y.; Li, D.; Lam, K. Online Video Super-Resolution with Convolutional Kernel Bypass Graft. IEEE Trans. Multimed. 2022, 1–16.
- Wang, Y.; Isobe, T.; Jia, X.; Tao, X.; Lu, H.; Tai, Y. Compression-Aware Video Super-Resolution. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, Seattle, WA, USA, 14–19 June 2023; pp. 2012–2021.
- 5. Jin, X.; He, J.; Xiao, Y.; Yuan, Q. Learning a Local-Global Alignment Network for Satellite Video Super-Resolution. IEEE Geosci. Remote Sens. Lett. 2023, 20, 1–5.
- 6. Xiao, Y.; Yuan, Q.; Jiang, K.; Jin, X.; He, J.; Zhang, L.; Lin, C. Local-Global Temporal Difference Learning for Satellite Video Super-Resolution. arXiv 2023, arXiv:2304.04421.
- 7. Guarnieri, G.; Fontani, M.; Guzzi, F.; Carrato, S.; Jerian, M. Perspective registration and multiframe super-resolution of license plates in surveillance videos. Digit. Investig. 2021, 36, 301087.
- 8. Guo, K.; Guo, H.; Ren, S.; Zhang, J.; Li, X. Towards efficient motion-blurred public security video super-resolution based on back-projection networks. J. Netw. Comput. Appl. 2020, 166, 102691.
- Yu, F.; Li, H.; Bian, S.; Tang, Y. An Efficient Network Design for Face Video Super-resolution. In Proceedings of the Conference on Computer Vision Workshops, virtual event, 10–17 October 2021; pp. 1513–1520.

- 10. López-López, E.; Pardo, X.M.; Regueiro, C.V. Incremental Learning from Low-labelled Stream Data in Open-Set Video Face Recognition. Pattern Recognit. 2022, 131, 108885.
- Lee, Y.; Yun, J.; Hong, Y.; Lee, J.; Jeon, M. Accurate license plate recognition and superresolution using a generative adversarial networks on traffic surveillance video. In Proceedings of the IEEE International Conference on Consumer Electronics-Asia (ICCE-Asia), Jeju, Republic of Korea, 24–26 June 2018; pp. 1–4.
- 12. Seibel, H.; Goldenstein, S.; Rocha, A. Eyes on the Target: Super-Resolution and License-Plate Recognition in Low-Quality Surveillance Videos. IEEE Access 2017, 5, 20020–20035.
- 13. Zhang, L.; Wu, X. An edge-guided image interpolation algorithm via directional filtering and data fusion. IEEE Trans. Image Process. 2006, 15, 2226–2238.
- 14. Liu, X.; Zhao, D.; Zhou, J.; Gao, W.; Sun, H. Image Interpolation via Graph-Based Bayesian Label Propagation. IEEE Trans. Image Process. 2014, 23, 1084–1096.
- 15. Tian, C.; Yuan, Y.; Zhang, S.; Lin, C.; Zuo, W.; Zhang, D. Image super-resolution with an enhanced group convolutional neural network. Neural Netw. 2022, 153, 373–385.
- 16. Tian, C.; Zheng, M.; Zuo, W.; Zhang, B.; Zhang, Y.; Zhang, D. Multi-stage image denoising with the wavelet transform. Pattern Recognit. 2023, 134, 109050.
- 17. Zhu, Z.; He, X.; Li, C.; Liu, S.; Jiang, K.; Li, K.; Wang, J. Adaptive Resolution Enhancement for Visual Attention Regions Based on Spatial Interpolation. Sensors 2023, 23, 6354.
- 18. Wen, W.; Ren, W.; Shi, Y.; Nie, Y.; Zhang, J.; Cao, X. Video Super-Resolution via a Spatio-Temporal Alignment Network. IEEE Trans. Image Process. 2022, 31, 1761–1773.
- Caballero, J.; Ledig, C.; Aitken, A.P.; Acosta, A.; Totz, J.; Wang, Z.; Shi, W. Real-Time Video Super-Resolution with Spatio-Temporal Networks and Motion Compensation. In Proceedings of the Conference on Computer Vision Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 2848–2857.
- 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 Conference on Computer Vision Pattern Recognition Workshops, Long Beach, CA, USA, 15–20 June 2019; pp. 1954–1963.
- 21. Wang, W.; Liu, Z.; Lu, H.; Lan, R.; Zhang, Z. Real-Time Video Super-Resolution with Spatio-Temporal Modeling and Redundancy-Aware Inference. Sensors 2023, 23, 7880.
- 22. Isobe, T.; Li, S.; Jia, X.; Yuan, S.; Slabaugh, G.G.; Xu, C.; Li, Y.; Wang, S.; Tian, Q. Video Super-Resolution with Temporal Group Attention. In Proceedings of the Conference on Computer Vision Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 8005–8014.
- 23. Ying, X.; Wang, L.; Wang, Y.; Sheng, W.; An, W.; Guo, Y. Deformable 3D Convolution for Video Super-Resolution. IEEE Signal Process. Lett. 2020, 27, 1500–1504.

- Liu, H.; Zhao, P.; Ruan, Z.; Shang, F.; Liu, Y. Large Motion Video Super-Resolution with Dual Subnet and Multi-Stage Communicated Upsampling. In Proceedings of the AAAI Conference on Artificial Intelligence, in virtua, 2–9 February 2021; pp. 2127–2135.
- Geng, Z.; Liang, L.; Ding, T.; Zharkov, I. RSTT: Real-time Spatial Temporal Transformer for Space-Time Video Super-Resolution. In Proceedings of the Conference on Computer Vision Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17420–17430.
- 26. Liu, C.; Yang, H.; Fu, J.; Qian, X. Learning Trajectory-Aware Transformer for Video Super-Resolution. In Proceedings of the Conference on Computer Vision Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 5677–5686.
- 27. Qiu, Z.; Yang, H.; Fu, J.; Liu, D.; Xu, C.; Fu, D. Learning Spatiotemporal Frequency-Transformer for Low-Quality Video Super-Resolution. arXiv 2022, arXiv:2212.14046.
- Dong, C.; Loy, C.C.; He, K.; Tang, X. Learning a Deep Convolutional Network for Image Super-Resolution. In Proceedings of the Conference on Computer Vision, Zurich, Switzerland, 6–12 September 2014; Volume 8692, pp. 184–199.
- Timofte, R.; De Smet, V.; Gool, L.V. Anchored Neighborhood Regression for Fast Example-Based Super-Resolution. In Proceedings of the IEEE International Conference on Computer Vision, ICCV 2013, Sydney, Australia, 1–8 December 2013; pp. 1920–1927.
- Kim, J.; Lee, J.K.; Lee, K.M. Accurate Image Super-Resolution Using Very Deep Convolutional Networks. In Proceedings of the Conference on Computer Vision Pattern Recognition, Las Vegas, NV, USA, 26 June–1 July 2016; pp. 1646–1654.
- Song, D.; Wang, Y.; Chen, H.; Xu, C.; Xu, C.; Tao, D. AdderSR: Towards Energy Efficient Image Super-Resolution. In Proceedings of the Conference on Computer Vision Pattern Recognition, virtual event, 10–17 October 2021; pp. 15648–15657.
- Liang, J.; Cao, J.; Sun, G.; Zhang, K.; Gool, L.V.; Timofte, R. SwinIR: Image Restoration Using Swin Transformer. In Proceedings of the Conference on Computer Vision Workshops, virtual event, 10–17 October 2021; pp. 1833–1844.
- 33. Tian, C.; Zhang, Y.; Zuo, W.; Lin, C.; Zhang, D.; Yuan, Y. A heterogeneous group CNN for image super-resolution. arXiv 2022, arXiv:2209.12406.
- 34. Lee, S.J.; Lee, S.G. Efficient Super-Resolution Method for Targets Observed by Satellite SAR. Sensors 2023, 23, 5893.
- Shi, Y.; Han, L.; Han, L.; Chang, S.; Hu, T.; Dancey, D. A Latent Encoder Coupled Generative Adversarial Network (LE-GAN) for Efficient Hyperspectral Image Super-Resolution. IEEE Trans. Geosci. Remote Sens. 2022, 60, 1–19.

- Yang, F.; Yang, H.; Fu, J.; Lu, H.; Guo, B. Learning Texture Transformer Network for Image Super-Resolution. In Proceedings of the Conference on Computer Vision Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 5790–5799.
- 37. Malczewski, K. Diffusion Weighted Imaging Super-Resolution Algorithm for Highly Sparse Raw Data Sequences. Sensors 2023, 23, 5698.
- 38. Zhang, D.; Tang, N.; Zhang, D.; Qu, Y. Cascaded Degradation-Aware Blind Super-Resolution. Sensors 2023, 23, 5338.
- 39. Wang, Z.; Yi, P.; Jiang, K.; Jiang, J.; Han, Z.; Lu, T.; Ma, J. Multi-Memory Convolutional Neural Network for Video Super-Resolution. IEEE Trans. Image Process. 2019, 28, 2530–2544.
- Tian, Y.; Zhang, Y.; Fu, Y.; Xu, C. TDAN: Temporally-Deformable Alignment Network for Video Super-Resolution. In Proceedings of the Conference on Computer Vision Pattern Recognition, Seattle, WA, USA, 14–19 June 2020; pp. 3357–3366.
- Xu, G.; Xu, J.; Li, Z.; Wang, L.; Sun, X.; Cheng, M. Temporal Modulation Network for Controllable Space-Time Video Super-Resolution. In Proceedings of the Conference on Computer Vision Pattern Recognition, virtual event, 10–17 October 2021; pp. 6388–6397.
- Fuoli, D.; Gu, S.; Timofte, R. Efficient Video Super-Resolution through Recurrent Latent Space Propagation. In Proceedings of the Conference on Computer Vision Workshops, Long Beach, CA, USA, 15–20 June 2019; pp. 3476–3485.
- Isobe, T.; Jia, X.; Tao, X.; Li, C.; Li, R.; Shi, Y.; Mu, J.; Lu, H.; Tai, Y.W. Look Back and Forth: Video Super-Resolution with Explicit Temporal Difference Modeling. In Proceedings of the Conference on Computer Vision Pattern Recognition, New Orleans, LA, USA, 18–24 June 2022; pp. 17411–17420.

Retrieved from https://encyclopedia.pub/entry/history/show/114999