International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

Video Inpainting with Local and Global Refinement

Mrs. D. Srivalli, M. Tech[1]

[2]V. Divya Sri, [3]MD. Mymunnisa Begum, [4]N. Chandra Prakash, [5]R. Sai Kumar, [6]P. Pallavi

[1]Assistant professor, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru

[2]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru [3]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru [4]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru [5]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru [6]Student, Dept. of CSE, Dhanekula Institute of Engineering and Technology, Ganguru

Abstract— Video inpainting is a crucial task in computer vision and video editing, involving in removing unnecessary objects and restoration of missing or corrupted regions within a video sequence. One approach that has gained prominence in recent years is the combination of local and global refinement techniques. This innovative strategy leverages the strengths of both local and global information to produce high-quality inpainted videos. In the context of video inpainting, local refinement focuses on accurately restoring missing or damaged regions by considering nearby pixels or frames. Encoder-decoder network can be employed to fill in gaps with content that seamlessly blends with the surrounding context. On the other hand, global refinement seeks to ensure temporal consistency and smooth transitions between frames, preventing noticeable artifacts or jittering in the inpainted video.

In conclusion, by seamlessly blending local and global inpainting strategies, these methods can effectively remove unwanted elements from videos while preserving both spatial and temporal coherence. This technology finds applications in video editing, restoration of damaged archival footage, and even in the entertainment industry for special effects and scene corrections, ultimately contributing to the improvement of video quality and aesthetics.

Keywords: Video Inpainting, Local and Global Refinement, Encoder-Decoder network, Recurrent Flow Completion, mask-guided sparse video Transformer, dual-domain propagation.

I.INTRODUCTION

Inpainting refers to the process of reconstructing missing or damaged parts of images or videos. Object removal, a common task in video editing, involves seamlessly filling in the regions left behind after an object has been removed. Patch-based methods play a central role in video inpainting. This method typically involves breaking downthe video into smaller patches and then searching for similar patches in the surrounding areas to fill in the missing regions. By leveraging the similarity between patches, these methods can effectively reconstruct missing information. Local refinement techniques focus on refining the inpainted regions at a fine level of detail.By considering the local structure and texture of the image, local refinement helps to enhance the visual quality of the inpainted result. On the other hand, global refinement approaches consider the inpainting problem from a broader perspective. These techniques aim to enforce global consistency and coherence in the inpainted regions across frames in a video sequence. By analyzing the temporal context and motion information, global refinement methods can produce more temporally coherent results, reducing artifacts such as flickering or jittering in the inpainted video. The combination of both local and global refinement strategies enables video inpainting algorithms to achieve high-quality and visually appealing results. Local refinement ensures that the details and textures in the inpainted regions match well with the surrounding content, while global refinement ensures temporal consistency and smooth transitions between frames.

II.RELATED WORK

IEEE TRANSACTIONS ON IMAGE PROCESSING, VOL.XX, 20 22

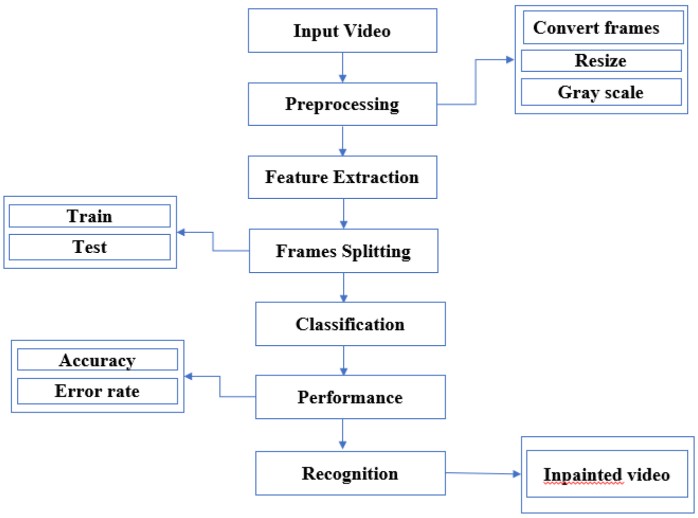
In this paper, from the perspective of the receptive field, they proposed a three-stage generative network for image inpainting. A coarse inpainting network with a large receptive field is applied to complete the whole structure and partial texture details. A local refinement network with a small receptive field is designed to eliminate the visual artifacts strongly related to its local region and prevent the negative effect from far and failed filling contents. An attention-based global refinement network with a large receptive field is proposed to further improve the visual quality of inpainted results using the global information and the more stable attention computation. The result shows 90% accuracy in means of LGNet. In this work, the framework concatenates two sub-networks with small and large receptive fields to handle different types of missing regions and artifacts. This study implies that it is beneficial to introduce different receptive fields for image inpainting tasks.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

III.METHODOLOGY USED

3.1 Model Diagram:

This structure offers a broad summary of significant system elements, primary participants in the process, and crucial interconnections.



3.2 Algorithms Used:

3.2.1 Recurrent Flow Completion:

RFC is a technique used in video inpainting especially for object removal tasks. Video inpainting involves filling up missing or removed regions in a video sequence with visuals that are visually convincing. Object removal entails getting rid of an object or part of a video and replacing it with content that blends seamlessly into the background. However, temporal inconsistencies pose issues for traditional methods as there are some artifacts such as flickering or jittering which arise from this. In addition, RFC leverages on Recurrent Neural Networks (RNNs) to capture long-range dependencies and ensure temporal coherence. It involves several key steps:

1. Flow Estimation: Dense optical flow fields are calculated between consecutive frames to describe apparent motion of objects.

2. Recurrent Neural Networks (RNNs): These could be LSTM's or GRU's which attempt to model the temporal dynamics involved within the video sequence thus capturing long-term dependencies and relationship between times t_1 and t_2.

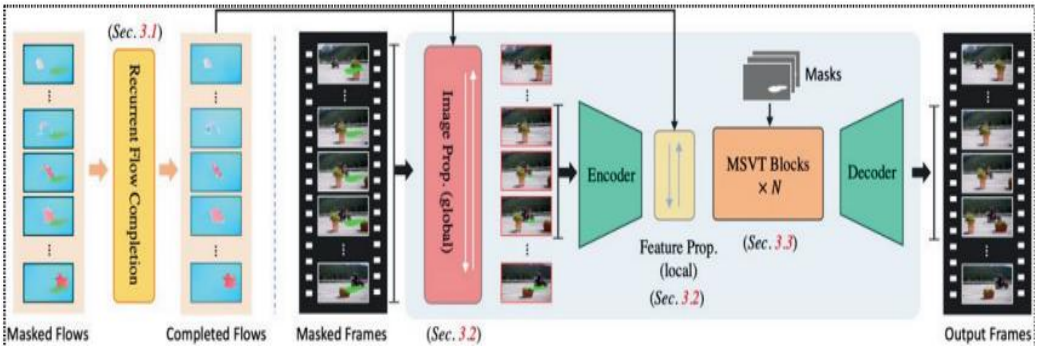
3. Flow Completion: Missing regions in each frame are filled iteratively using estimated flow fields and information from neighboring frames.

4. emporal Coherence: Consistency with surrounding frames can be achieved due to recurrent connections in the network thereby eliminating any temporal artifacts.

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

3.2.2 Dual-domain Propagation:

Dual-domain propagation is an advanced video inpainting technique that specializes in removing objects from videos. It cleverly combines spatial (image frames) and temporal (video sequence) information to ensure seamless and realistic inpainting. Using flow estimation, it understands how objects move between frames and transfers information accordingly. This synchronized approach guarantees both spatial (consistent within each frame) and temporal (smooth transitions between frames) coherence, eliminating unwanted effects like flickering or jittery movements. Additionally, iterative optimization and deep learning enhancements further improve the inpainting accuracy. Dual-domain processing addresses both spatial and temporal domains, allowing for precise object removal even in intricate scenes. Researchers are also actively working on improving computational efficiency for real-time applications.



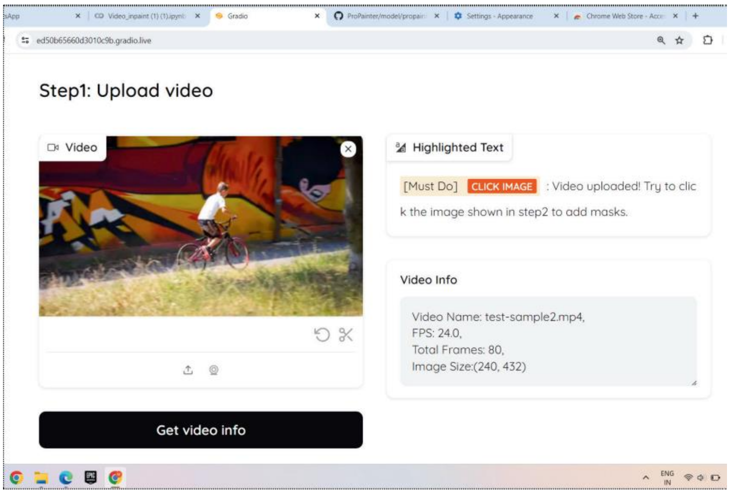
3.2.3 Mask-Guided Sparse Video Transformer:

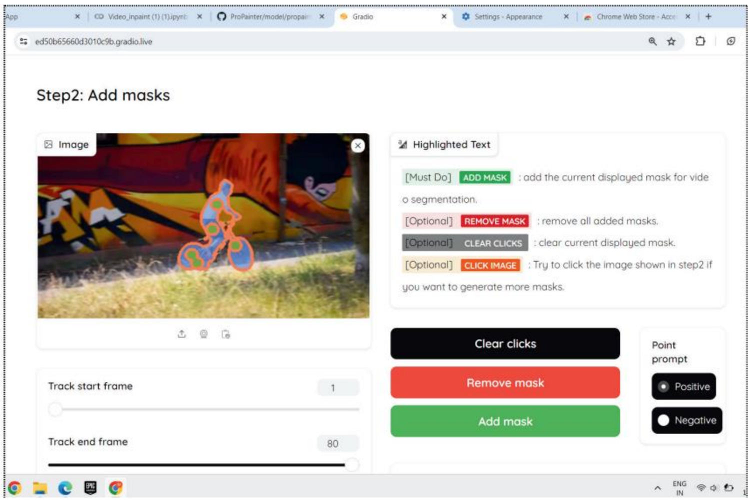
The Mask-Guided Sparse Video Transformer is a cutting-edge method for video inpainting, focusing on object removal. It employs a specialized Transformer architecture, tailored for video, with sparse attention mechanisms to handle large frames efficiently. By utilizing masks to guide the process, it ensures accurate reconstruction, even in complex scenes. The Transformer captures long-range dependencies, ensuring temporal coherence. Through iterative refinement and neighboring frame data, it produces visually coherent results with minimal artifacts. This technique supports parallel processing, enabling faster speeds and user interaction through masks for guidance. Evaluation metrics confirm its superiority over traditional methods in quality and efficiency. It finds applications in video restoration, scene completion, and special effects generation. Ongoing research aims to enhance quality, improve efficiency, and explore new applications, making it a promising advancement in video inpainting for object removal.

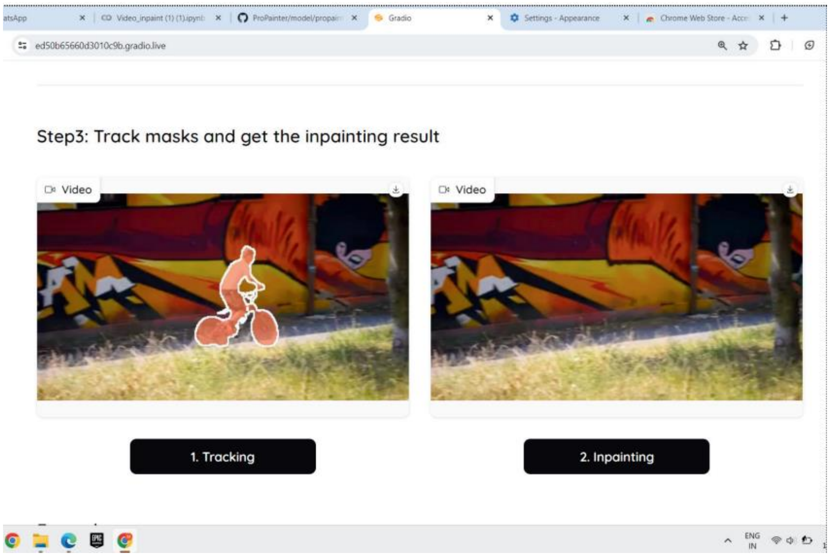
International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 08 Issue: 03 | March - 2024 SJIF Rating: 8.176 ISSN: 2582-3930

IV.RESULT AND OBSERVATION

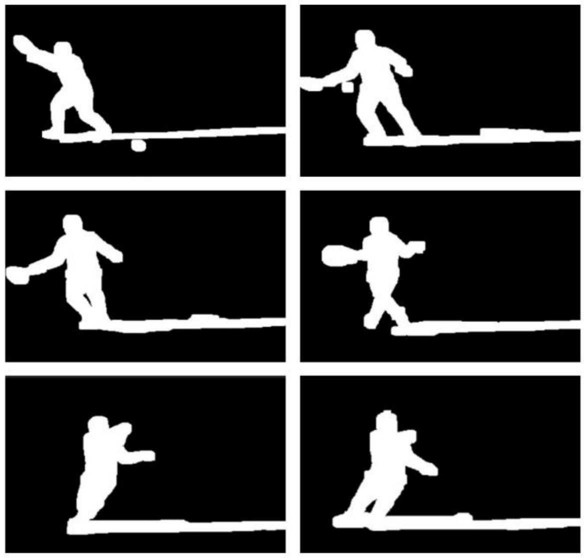
The Web Application Interface is given below:







For example, the input video is given to the system to perform inpainting and is depicted like the following:

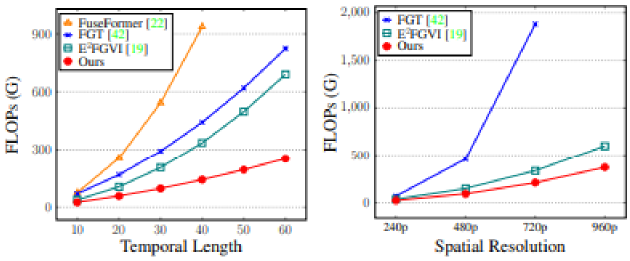


The Inpainting Model performs the Inpainting Process by initially splitting the video into several frames with specific dimensions, then generating a mask to inpaint the particular regions using machine learning and deep learning algorithms. The output video is obtained as following:

Here, the undesired object is removed from the video with the help of an inpainting mask, and the desired video is obtained as the final result.

V.GRAPHICAL REPRESENTATION

The graphical representation of the performance is depicted as follows:



VI.CONCLUSION

In conclusion, the combination of video inpainting with local and global refinement using encoder-decoder network for object removal, presents a powerful and effective approach in the field of video processing and restoration. By leveraging the strengths of both local and global inpainting techniques, the model can address various complexities within video frames, such as small-scale details and large-scale structures. The incorporation of the encoder-decoder network for object removal adds another layer of versatility to the approach. This technique efficiently identifies and removes unwanted objects or artifacts from the video frames, contributing to the overall enhancement of visual quality. The encoder-decode network ensures a swift and accurate removal process, crucial for real-time or time-sensitive applications.

VII.REFERENCES

1. D. Ding, S. Ram, and J. J. Rodr´ıguez, "Image inpainting using nonlocal texture matching and nonlinear filtering,"IEEE Trans. Image Process., vol. 28, no. 4, pp. 1705–1719, 2019.

2. Y. Zeng, J. Fu, H. Chao, and B. Guo, "Learning pyramid-context encoder network for high-quality imageinpainting," in IEEE Conf. Comput. Vis. Pattern Recog., 2019, pp. 1486–1494.

3. C. Xie, S. Liu, C. Li, M.-M. Cheng, W. Zuo, X. Liu, S. Wen, and E. Ding, "Image inpainting with learnablebidirectional attention maps," in IEEE Int. Conf. Comput. Vis., 2019, pp. 8858–8867.

4. H. Liu, B. Jiang, Y. Song, W. Huang, and C. Yang, "Rethinking image inpainting via a mutual encoder-decoderwith feature equalizations," in Eur. Conf. Comput. Vis., 2020.

5. R. Zhang, W. Quan, B. Wu, Z. Li, and D.-M. Yan, "Pixel-wise dense detector for image inpainting," Comput.Graph. Forum, vol. 39, no. 7, pp. 471–482, 2020

6. , "Free-form image inpainting with gated convolution," in IEEE Int. Conf. Comput. Vis., 2019, pp.

4471–4480.

7. K. Nazeri, E. Ng, T. Joseph, F. Qureshi, and M. Ebrahimi, "EdgeConnect: Structure guided image inpainting usingedge prediction," in Int. Conf. Comput. Vis. Worksh., 2019.

8. Y. Ren, X. Yu, R. Zhang, T. H. Li, S. Liu, and G. Li, "Structureflow: Image inpainting via structure-awareappearance flow," in IEEE Int. Conf. Comput. Vis., 2019, pp. 181–190.

9. Z. Yi, Q. Tang, S. Azizi, D. Jang, and Z. Xu, "Contextual residual aggregation for ultra high-resolution imageinpainting," in IEEE Conf. Comput. Vis. Pattern Recog., 2020, pp. 7508–7517.

10. Y. Wang, Y.-C. Chen, X. Tao, and J. Jia, "Vcnet: A robust approach to blind image inpainting," in Eur. Conf.Comput. Vis., 2020.

11. T. Wang, H. Ouyang, and Q. Chen, "Image inpainting with externalinternal learning and monochromicbottleneck," in IEEE Conf. Comput. Vis. Pattern Recog., 2021, pp. 5120–5129.

12. X. Li, B. Zhang, J. Liao, and P. V. Sander, "Let's see clearly: Contaminant artifact removal for moving cameras,"in IEEE Int. Conf. Comput. Vis., 2021, pp. 2011–2020.

13. Z. Guo, Z. Chen, T. Yu, J. Chen, and S. Liu, "Progressive image inpainting with full-resolution residual network,"in ACM Int. Conf. Multimedia, 2019, p. 2496–2504.

14. J. Li, F. He, L. Zhang, B. Du, and D. Tao, "Progressive reconstruction of visual structure for image inpainting," inIEEE Int. Conf. Comput. Vis., 2019, pp. 5961–5970.

15. J. Li, N. Wang, L. Zhang, B. Du, and D. Tao, "Recurrent feature reasoning for image inpainting," in IEEE Conf.Comput. Vis. Pattern Recog., 2020, pp. 7757–7765.

16. Y. Zeng, Z. Lin, J. Yang, J. Zhang, E. Shechtman, and H. Lu, "Highresolution image inpainting with iterativeconfidence feedback and guided upsampling," in Eur. Conf. Comput. Vis., 2020.

17. H. Liu, B. Jiang, Y. Xiao, and C. Yang, "Coherent semantic attention for image inpainting," in IEEE Int. Conf.Comput. Vis., 2019, pp. 4170– 4179.

18. L. Liao, J. Xiao, Z. Wang, C.-W. Lin, and S. Satoh, "Guidance and evaluation: Semantic-aware image inpaintingfor mixed scenes," in Eur. Conf. Comput. Vis., 2020.

19. Z. Hui, J. Li, X. Wang, and X. Gao, "Image fine-grained inpainting," arXiv preprint arXiv:2002.02609, 2020.

20. N. Wang, Y. Zhang, and L. Zhang, "Dynamic selection network for image inpainting," IEEE Trans. ImageProcess., vol. 30, pp. 1784–1798, 2021.

21. M. Zhu, D. He, X. Li, C. Li, F. Li, X. Liu, E. Ding, and Z. Zhang, "Image inpainting by end-to-end cascadedrefinement with mask awareness," IEEE Trans. Image Process., vol. 30, pp. 4855–4866, 2021.

22. H. Wu, J. Zhou, and Y. Li, "Deep generative model for image inpainting with local binary pattern learning andspatial attention," arXiv preprint arXiv:2009.01031, 2020