IJIRSET
International Journal of Innovative Research in
SCIENCE | ENGINEERING | TECHNOLOGY

e-ISSN: 2319-8753 | p-ISSN: 2320-6710

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

Volume 10, Issue 6, June 2021

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

Impact Factor: 7.512

9940 572 462 | 6381 907 438 | ijirset@gmail.com | www.ijirset.com

International Journal of Innovative Research in Science, Engineering and Technology (IJIRSET)

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006256 |

Video Inpainting using CNN

Shreya Godse[1], Sankita Joshi[2], Aeshani Prabhu[3], Mugdha Shinde[4], Snehal Bhujbal[5], Mr. Pradeep P. Patil[6]

UG Students, Department of Computer Engineering, P.E.S.'s Modern College of Engineering, Pune, Maharashtra, India[1,2,3,4,5]

Assistant Professor, Department of Computer Engineering, P.E.S.'s Modern College of Engineering Pune, Maharashtra, India[6]

ABSTRACT: This project aims to propose a quick and automatic high-definition video painting technique that operates under several difficult conditions, such as a moving camera, a complex backdrop or a long occlusion. Our algorithm does not restrict the type of removal of objects, but extends to simultaneous context and foreground reconstruction, even if for a long time the moving objects are occluded. A new deep network architecture for quick painting with video. Our framework is designed to collect and refine information from neighbouring frames and synthesise still unknown regions, based on an image-based encoder-decoder model. At the same time, a recurring feedback and a temporal memory module force the output to be temporally consistent. Our system generates videos that are much more semantically accurate and temporally fluid compared to the state-of-the-art image painting algorithm. In comparison to the previous method of video completion, which relies on time-consuming optimization, our technique works almost in real time while delivering competitive video results.

KEYWORDS: Video inpainting, Video completion, Video object removal, Video caption removal, Video decaptioning, Video editing

I. INTRODUCTION

Specifically, the CNN encoder-decoder is built under the network structure, which allows us to estimate the inpinning potential for each pixel in the image. To guide CNN to automatically learn inpainting features, the matrix creates a label by assigning a class label to each pixel of the image for CNN training, and the designed weight cross-entropy loss works. The entire system having below feature which is defined in step by step: Multi-to-Single Frame Video or image Inpainting In image, the occluded or removed parts in a frame are often revealed in the past/future frames as the objects move and the viewpoint changes. If such hints exist in the temporal radius, those disclosed content can be borrowed to recover the current frame. Otherwise, the still-unknown regions should be synthesized. To achieve this, we construct our model as an encoder-decoder network that learns such temporal feature aggregation and single-frame inpainting simultaneously. The network is designed to be fully convolutional, which can handle arbitrary size input. Source encoders Encoder networks extract the features from the target and the aligned reference frames. The input to the encoder is a concatenation of an RGB image and the corresponding binary mask. The details on the architecture will be described in the supplementary materials. Feature Extraction Before directly combining the source and reference features, we propose to explicitly align the feature points. This strategy helps our model easily borrow traceable features from the neighbor frames. To achieve this, we insert flow sub-networks to estimate the flows between the source and reference feature maps in four different spatial scales. Decoder The decoder network completes the target frame given target features, aggregated reference features, and mask. The inputs are concatenated before being fed into the decoder. Decoder is basically our paste network that learns to fill the missing region by using the aggregated reference features and the visibility of those features. The pixels marked on mask are pixels that are never visible in all reference frames because those pixels always fall into holes. Therefore, the decoder has to be able to synthesize contents for those pixels as well. We add dilated convolution blocks to grow the receptive field and design the decoder network deeper than the other networks, in order to enhance the completion results for the unseen area by looking at other pixels within the image itself.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006256 |

II. MOTIVATION

- Video and Image painting for customers has always been a daunting and continuing area of discovery.
- A lot of work has been invested in developing better instruments for users to take athsetically beautiful pictures with the growing demand for taking nice pictures.
- But occasionally, when the unrelated object of interest is obstructed, users can take unnecessary photographs.
- For example, in our project, inspired by these could-be-improved video and images, we will try to tackle the problem of getting rid of these undesirable artefacts by using generative and adversarial networks in these video and images.

III. LITERATURE SURVEY

Kim, Dahun, et al[1] Centered on this structure, the architectural designs. Our first model is a Blind Video Decaptioning Network (BVDNet) that is programmed without any mask information to automatically erase and paint text overlays in images. In the ECCV Chalearn 2018 LAP Inpainting Competition Track 2: Video Decaptioning, our BVDNet took first place. Second, to deal with more random and larger gaps, we propose a network for more general video painting (VINet). Compared to state-of-the-art approaches, video findings display the value of our framework both qualitatively and quantitatively.

Chu, Mengyu, et al.[2]A self-supervised algorithm on a temporary basis. For both tasks, we show that the secret to achieving temporally coherent solutions without compromising spatial information is temporal adversarial learning. To boost long-term temporal continuity, we also suggest a novel Ping-Pong loss. It effectively keeps recurrent networks from momentarily accumulating objects without depressing detailed characteristics. In addition, we suggest a first set of measures to test the accuracy and perceptual consistency of temporal evolution quantitatively.

Wang, Haodi, et al. [3]The inclusion of the global discriminator helps the network to guarantee the complete image's global accuracy and background continuity while retaining local information. The loss function is adjusted accordingly on this basis. The role of loss is composed of two parts: loss of reconstruction and multi-scale versus loss of learning. The network architecture, which shows fair repair efficiency, is applied to the Paris dataset. It tends to boost the patching and blurring issues of the Background Encoders, and produces better results in painting. The experiment improves the network's generalisation potential in comparison with the painting of existing face pictures. Comparing this paper's network structure with Background Encoders and GAN-based face image restoration projects It is much more succinct and quicker.

Wang, Chuan, et al. [4]On video inpainting, with the added temporal dimension, an expanded problem from image inpainting. Such expansion introduces new technological problems that are challenging to overcome. First, retrieving missing video content requires not only the spatial context of each frame, but also the context of motion across frames to be understood. Second, in both global context-level and local image-feature-level, the output video must maintain high spatiotemporal accuracy.

Rosset Antoine et al. [5] design navigation feature a multi-dimensional image display and image interpretation and Multimodality multidimensional group. The system is very fast 3D graphics standard and optimized widely used for computer games optimized for tracking hardware graphics accelerator every available advantage. Methods of imaging radiology groups included 2D slices spiral volumetric 3D topographic dimension object is to get a fourth or fifth data temporarily and functional, which can handle ultrafast CT scanning and MR scanner with combined PET / CT.

A.Bardera et al. [6] represents a new approach based thresholding segmentation using information theoretical entropy measure excessive structural information of a 2D or 3D object and find the optimal threshold. The main motivation for this approach is to use an excess entropy as a measure of structural information of an image. The experimental results showed good behavior apply access.

Carlos et al. [7] presents an efficient package for image viewing ; this package to describe a graphical tool for collaborative analysis and visualization of model image created by the volume of data slices ; This package allows a

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006256 |

number of users of medical image analysis , at the same time and coordinated . In addition, clinical , direct measures of structures in 3D space seem mandatory.

DaminiDey et al[8]. Endoscopic operating area resulting in improved 3D viewing surgical preoperative scans method developed full video input. This method allows visualization of stereoscopic and panoramic navigation arbitrarily painted perspective after the procedure. Validation Experiments ghosts applies when image data pre operating conditions accurately reflect the operating system.

JR Swedlow [9] shows that the server open development environment Microscopy (home) and Home distant objects server to provide a flexible tool set for managing image data . This project has developed and released the data model that provides a description of the structure of image data acquisition and analysis results.

Sun K. Yoo et al. [10] explains the widely accepted 3D visualization of medical imaging facility to help diagnose the patient. Terms rated view streaming optimization and control medical 3D image object that has a high quality web. 3D object visualization system is applied in various fields such as bio-interactive guided surgery, diagnostic technologies and computer-aided telemedicine.

III. PROPOSED SYSTEM DESIGN

To Design and Develop a Video Inpainting System for Video De-captioning, Object Removal and Quality Enhancement are using Deep Learning.

ALGORITHM DESIGN

Training
Input: Training dataset TrainData[], Various activation functions[], Threshold Th
Output: Extracted Features Feature_set[] for completed trained module.
Step 1: Set input block of data d[], activation function, epoch size,
Step 2 : Features.pkl←ExtractFeatures(d[])
Step 3 : Feature_set[] ← optimized(Features.pkl)
Step 4 : Return Feature_set[]

Testing
Input: Test Dataset which contains various test instances TestDBLits [], Train dataset which is build by training phase TrainDBLits[] , Threshold Th.
Output: HashMap<class_label, SimilarityWeight> all instances which weight violates the threshold score.

Step 1: For each read each test instances using below equation

$$testFeature(m) = \sum_{m=1}^{n} (.featureSet[A[i] \dots\dots A[n] \leftarrow \text{TestDBLits})$$

Step 2 :extract each feature as a hot vector or input neuron from $testFeature(m)$ using below equation.
Extracted_FeatureSetx[t……n] = $\sum_{x=1}^{n}(t) \leftarrow testFeature$ (m)
Extracted_FeatureSetx[t] contains the feature vector of respective domain

Step 3: For each read each train instances using below equation

$$trainFeature(m) = \sum_{m=1}^{n} (.featureSet[A[i] \dots\dots A[n] \leftarrow \text{TrainDBList})$$

Step 4 :extract each feature as a hot vector or input neuron from $testFeature(m)$ using below equation.
Extracted_FeatureSetx[t……n] = $\sum_{x=1}^{n}(t) \leftarrow testFeature$ (m)
Extracted_FeatureSetx[t] contains the feature vector of respective domain.

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006256 |

Step 5 : Now map each test feature set to all respective training feature set

$$weight = calcSim\ (\text{FeatureSetx} \mid\mid \sum_{i=1}^{n} \text{FeatureSety}[y])$$

Step 6 : Return Weight

IV. RESULTS AND DISCUSSIONS

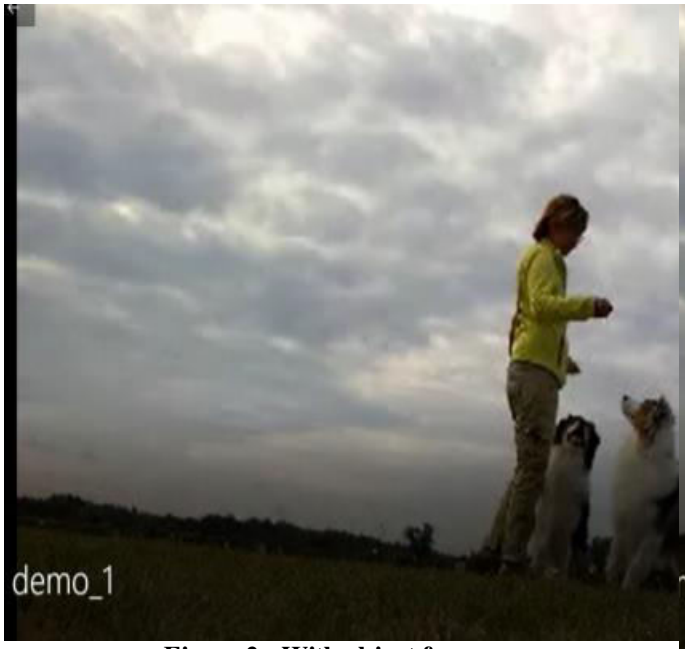

Figure 2 : With object frame

Figure 3 : after object removing frame

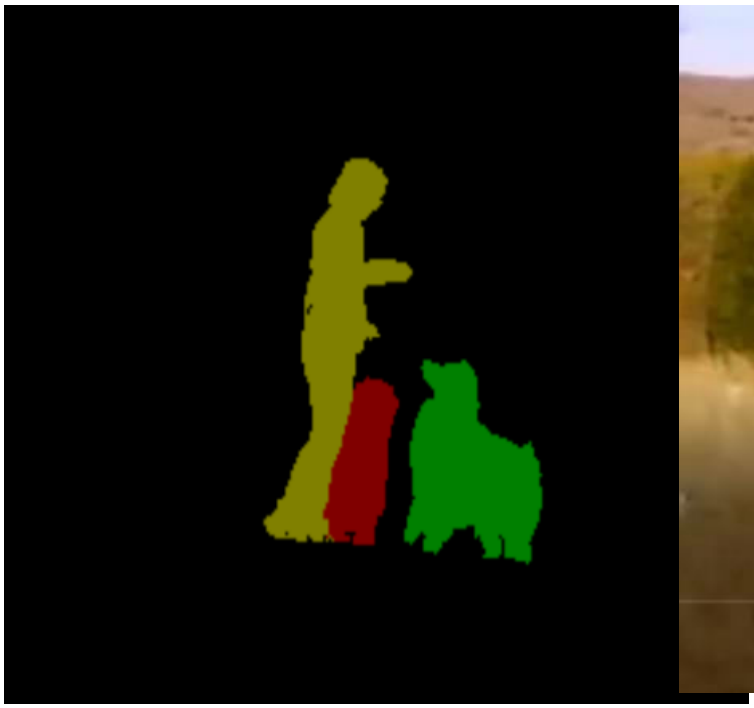

Figure 4 : Mask Information Result

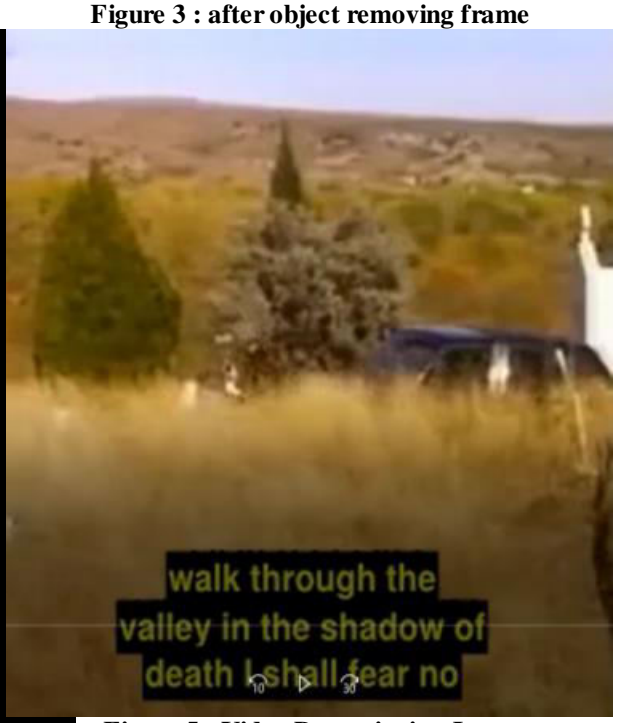

Figure 5 : Video Decaptioning Input

| e-ISSN: 2319-8753, p-ISSN: 2320-6710| www.ijirset.com | Impact Factor: 7.512|

|| Volume 10, Issue 6, June 2021 ||

| DOI:10.15680/IJIRSET.2021.1006256 |

Figure 6 : Decaption output

V. CONCLUSION

We propose a novel framework for video inpainting based on two guiding principles: aggregating temporal information and recurrently connecting past predictions for temporally coherent generation. To deal with different video inpainting scenarios, we propose two network designs : caption removal network and object removal network.Our extensive experiments aims to demonstrate that both caption removal and object removal achieve significantly better visual quality than the per-frame deep CNN based competitors.

VI. FUTURE WORK

Working on multiple data-sets from a network data-set, including NLP and in deep learning, in a large data environment

REFERENCES

[1] Kim, Dahun, et al. ”Recurrent temporal aggregation framework for deep video inpainting.” IEEE Transactions on Pattern Analysis and Machine Intelligence 42.5 (2019): 1038-1052.

[2] Chu, Mengyu, et al. ”Learning temporal coherence via self-supervision for GAN-based video generation.” ACM Transactions on Graphics (TOG) 39.4 (2020): 75-1.

[3] Wang, Haodi, et al. ”New inpainting algorithm based on simplified context encoders and multi-scale adversarial network.” Procedia computer science 147 (2019): 254-263.

[4] Wang, Chuan, et al. ”Video inpainting by jointly learning temporal structure and spatial details.” Proceedings of the AAAI Conference on Artificial Intelligence.Vol. 33. 2019.

[5] Antoine Rosset et al. ”OsiriX: An Open Source Software for Navigating in Multidimen- sional DICOM Images”, Journal of Digital Imaging, September 2004, vol.17, no.3, pp 205-216,.

[6] Bardera.A et al. ”Image Segmentation Using Access Entropy”, “Journal Signal Processing Systems”, 2009, vol.54, no.1-3, pp 205- 214. [7] Carlos Albarola et al., ”diSNei: A Collaborative Environment for Medical Images Analysis and Visualization” Medical Image Computing and Computer- Assisted Intervention (MIC-CAI 2000), Third International Conference, Pittsburgh, Pennsylvania, USA, October 11- 14, 2000, LNCS1935, pp 814-823.

[8] DaminiDey, ”Mixing Reality Merging Endoscopic Images and 3D surfaces”, Medical Im- age Computing and Computer-Assisted Intervention (MICCAI 2000), Third International Conference Pittsburgh, Pennsylvania, USA, October 11-14, 2000, LNCS1935, pp 796-805.

[9] Jason R. Swedlow, ”The Open Microscopy Environment: A Collaborative Data Modeling and Software Development Project for Biological Image Informatics”, Imaging Cellular and Molecular Biological Functions, Edited by Shorte, Spencer L.; Frischknecht, Friedrich, Springer, September 2007, vol. 02, no.03, pp 71-92.

[10] Sun K. Yoo et al. ”3D Visualization for Tele-medicine Visualization”, The 2006 Inter- national Conference on Computational Science and Applications (ICCSA 2006), LNCS, Glasgow, UK , May 8-11, 2006, 3980, pp 335-345.

INNO SPACE
SJIF Scientific Journal Impact Factor

Impact Factor:
7.512

ISSN INTERNATIONAL STANDARD SERIAL NUMBER INDIA

निस्केयर
NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN SCIENCE | ENGINEERING | TECHNOLOGY

9940 572 462 6381 907 438 ijirset@gmail.com

www.ijirset.com

Scan to save the contact details