Kaunas University of Technology

Faculty of Informatics

Research on Deep Learning Based Applications. Video Inpainting

Master's Final Degree Project

Serhii Postupaiev

Project author

Prof. Rytis Maskeliūnas

Supervisor

Kaunas, 2023

Kaunas University of Technology

Faculty of Informatics

Research on Deep Learning Based Applications. Video Inpainting

Master's Final Degree Project

Artificial Intelligence in Computer Science (6211BX007)

Serhii Postupaiev

Project author

Prof. Rytis Maskeliūnas

Supervisor

Abbreviation of the position, name and surname of the reviewer

Reviewer

Kaunas, 2023

Kaunas University of Technology

Faculty of Informatics

Serhii Postupaiev

Research on Deep Learning Based Applications. Video Inpainting

Declaration of Academic Integrity

I confirm that the final project of mine, Serhii Postupaiev, on the topic „Research on Deep Learning Based Applications. Video Inpainting“ is written completely by myself; all the provided data and research results are correct and have been obtained honestly. None of the parts of this thesis have been plagiarised from any printed, Internet-based or otherwise recorded sources. All direct and indirect quotations from external resources are indicated in the list of references. No monetary funds (unless required by Law) have been paid to anyone for any contribution to this project.

I fully and completely understand that any discovery of any manifestations/case/facts of dishonesty inevitably results in me incurring a penalty according to the procedure(s) effective at Kaunas University of Technology.

(name and surname filled in by hand)	(signature)

Kaunas University of Technology

Faculty of informatics

Task of the Master's final degree project

Topic of the project: Research on Deep Learning Based Applications. Video Inpainting

Requirements and conditions (title can be clarified, if needed):

Supervisor: Prof. Rytis Maskeliūnas

(position, name, surname, signature of the supervisor) (date)

Serhii Postupaiev. Research on Deep Learning Based Applications. Video Inpainting. Master's Final Degree Project/ supervisor Prof. Rytis Maskeliūnas; Faculty of Informatics, Kaunas University of Technology..

Study field and area (study field group): Artificial Intelligence in Computer Science.

Keywords: ..... (type here).

Kaunas, 2023. Number of pages.

Summary

Lorem ipsum dolor sit amet, eam ex decore persequeris, sit at illud lobortis atomorum. Sed dolorem quaerendum ne, prompta instructior ne pri. Et mel partiendo suscipiantur, docendi abhorreant ea sit. Recteque imperdiet eum te.

Eu eum decore inimicus consetetur, cu usu habeo corpora intellegam. Ut antiopam efficiendi deterruisset sit. Mel sint eirmod id, qui quot virtute id, dolor nemore forensibus usu id. Fugit dolore voluptatum cu vim. An vix veniam graecis insolens, sit posse iusto id. Ut vim ceteros percipit, id quo ubique recusabo, eum sint lucilius ea. In sumo inani numquam has.

Serhijus Postupajevas. Giluminiu mokymusi pagrįstų taikomųjų programų tyrimas. Vaizdo dažymas. Magistro baigiamojo studijų projektas/ vadovas prof. Rytis Maskeliūnas; Kauno technologijos universiteto Informatikos fakultetas...

Studijų kryptis ir sritis (studijų krypčių grupė): Dirbtinis intelektas kompiuterių moksle.

Raktiniai žodžiai: ..... (įveskite čia).

Kaunas, 2023. Puslapių skaičius.

Santrauka

Lorem ipsum dolor sit amet, eam ex decore persequeris, sit at illud lobortis atomorum. Sed dolorem quaerendum ne, prompta instructior ne pri. Et mel partiendo suscipiantur, docendi abhorreant ea sit. Recteque imperdiet eum te.

Eu eum decore inimicus consetetur, cu usu habeo corpora intellegam. Ut antiopam efficiendi deterruisset sit. Mel sint eirmod id, qui quot virtute id, dolor nemore forensibus usu id. Fugit dolore voluptatum cu vim. An vix veniam graecis insolens, sit posse iusto id. Ut vim ceteros percipit, id quo ubique recusabo, eum sint lucilius ea. In sumo inani numquam has.

Table of contents

1

List of figures (if needed)

List of tables (if needed)

List of abbreviations and terms

Abbreviation	Description
DL	Deep Learning
CNN	Convolutional Neural Network
GAN	Generative Adversarial Network
FFC	Fast Fourier Convolutions
PSNR	Peak signal-to-noise ratio
SSIM	Structural Similarity Index
LIPS	Local Image Patch Similarity
FID	Fréchet Inception Distance

Introduction

The field of computer vision has seen a significant focus on the research direction of video inpainting, as it is widely used in various applications such as video editing and video restoration. The primary goal of video inpainting is to generate visually consistent structure and texture for the regions that are missing or damaged in images.

With the rapid advancement of DL in recent years, the potential of DL in video processing has become increasingly evident, and has been able to address the limitations of traditional video inpainting algorithms to a certain extent. As a result, video inpainting based on deep learning has become a highly active area of research within the field of computer vision.

Aim and Objectives

The main aim is to …

The report consists of …

1. Video Inpainting Analysis

1.1. General Review

Video inpainting is an advanced technique that utilizes various forms of known information within a video to fill in the missing or unknown information. This can include information such as structural details, statistical data, semantic information, and more. As a result of utilizing different types of information, various techniques have emerged in the field of video inpainting. These techniques have a wide range of applications in research such as video super-resolution, removing obstructions in video, repairing damaged videos, and more. Furthermore, the ongoing research in these applications is also driving the advancement and development of video inpainting technology.

Over the past several decades, digital images and videos have emerged as critical mediums for both recording and disseminating information. To meet the needs of a wide range of video-based applications, numerous video processing technologies have been developed, including but not limited to video denoising, video super-resolution, video colorization, and video inpainting. These technologies are essential to ensure that digital videos are of high quality and can be effectively used in a variety of settings.

Video inpainting, in particular, is a method that aims to create visually convincing restorations for missing regions in damaged videos. These missing regions can be caused by a variety of reasons, such as regular 2 by 2 sampling patterns or physical damage to the video. Examples of this type of damage include scanned old photos or videos with cracks, captured videos with unwanted objects, and videos with damaged paintings. The goal of video inpainting is to repair these missing regions so that the resulting video is as visually similar to the original as possible.

The problem of video inpainting is complex, multi-faceted, and challenging. Firstly, the inputs for video inpainting are incredibly diverse and complex, including traditional gray and color images of nature scenes, line drawings/sketches, textures, texts, and depth images, which all require different inpainting strategies or algorithms. Secondly, the damage to the videos may be extensive, leading to unsatisfactory results with traditional algorithms. And lastly, inpainting is an ill-posed problem, meaning that the inpainting results are not unique, and most algorithms only consider one possible result.

The field of video inpainting poses a multitude of intricate and demanding challenges. One of the key challenges is the diversity and complexity of the inputs that are used for video inpainting. These inputs include traditional gray and color images of natural scenes, line drawings/sketches, textures, texts, and depth images. Each of these inputs require different strategies or algorithms in order to be effectively inpainted. Another major challenge is the extent of damage that videos may sustain, which can lead to poor quality results with traditional algorithms. And lastly, video inpainting is an ill-posed problem, meaning that there are multiple possible results and most algorithms only consider one of them, making it a difficult task to achieve a desired outcome.

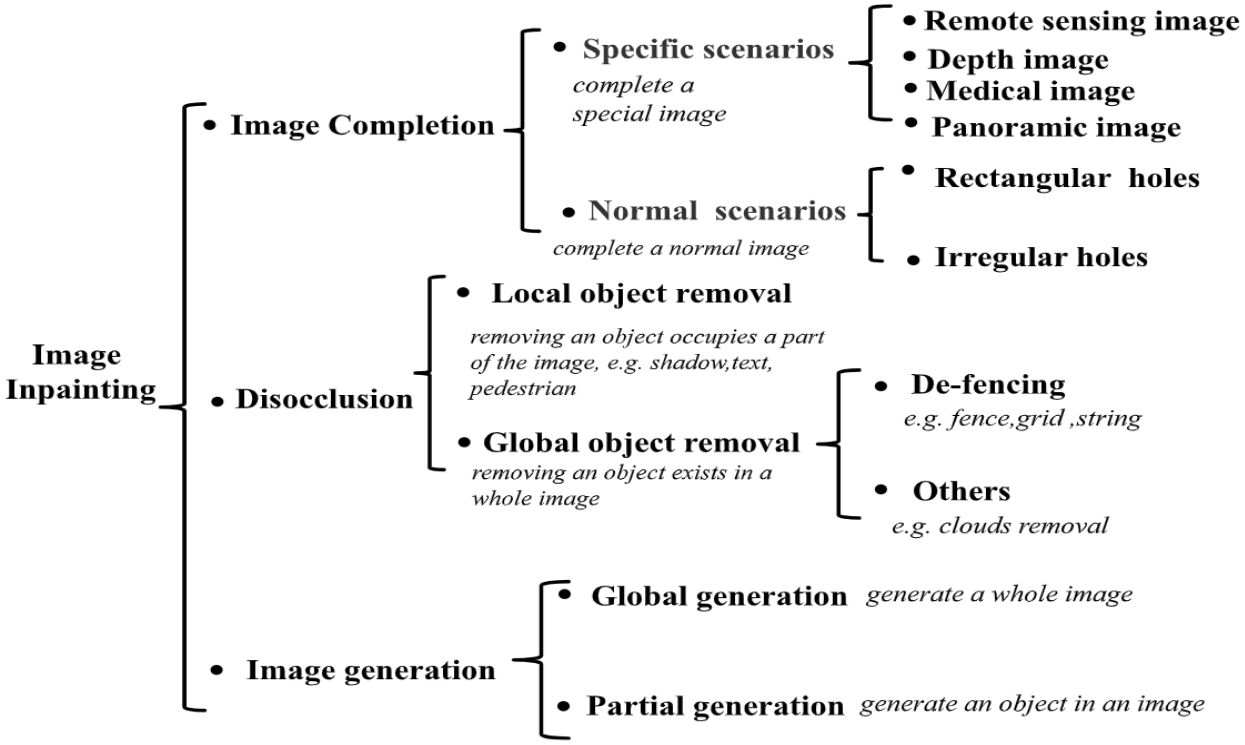
Deep learning, also known as deep structured neural networks, is a form of representation learning that aims to mimic the mechanism of the human brain. This approach learns the characteristics of input data by imitating the way information is transmitted between neurons in the

human brain, and finds the mapping between input and output data. When applied to video inpainting, deep learning technologies can accurately learn the semantic information of videos and then use that information to predict missing content. This greatly overcomes the limitations of traditional inpainting algorithms, resulting in more logical and visually appealing results. The field of deep learning has seen tremendous progress in recent years, with feedforward neural networks, particularly convolutional neural networks, being proven to be highly effective at capturing abstract information in images.

The use of GANs in deep learning has proven to be highly effective in improving the quality of generated results, particularly in the field of video inpainting. This powerful method has been demonstrated to generate missing content in videos with remarkable success.

1.2. Inpainting Tasks

The realm of video inpainting can be divided into three separate categories (Fig 1.1): image completion, object removal, and image generation. The task of image completion is to fill in missing areas of an image with appropriate content. The method described above can be split into two subcategories: the completion of special and ordinary images.



1.1 Fig. Inpainting tasks [1]

Finally, image generation represents a newer application of image inpainting, which involves using a given image as a starting point and then generating an area or an entirely new image. Object removal includes removing a specified object from an image and then filling the gap

with relevant content. Depending on the location of the object being removed within the image, this can be further divided into local and global object removal.

1.3. Application

Deep learning based inpainting algorithms have demonstrated superiority in comparison to traditional algorithms. Since time, deep learning based algorithms continued to improve and played a base role in solving such tasks as user-guided face editing, privacy protection, pose-guided image synthesis, digitization of cultural heritage, remote sensing, virtual and augmented reality, and other fields.

1.3.1. User-Guided Face Editing

User-guided face editing aims to specify which parts of a face in an image or video should be edited or manipulated by providing user inputs and guidance. Usually, the unwanted elements are masked and replaced with user-drawn strokes.

Tiziano Portenier et al. in their project [20] present a novel system for sketch-based face image editing, enabling users to edit images intuitively by sketching a few strokes on a region of interest. The proposed interface runs in real-time and facilitates an interactive and iterative workflow to quickly express the intended edits.

Youngjoo Jo et al. designed a novel image editing system [21] that generates images as the user provides free-form mask, sketch and color as an input. The system responds to the user's sketch and color input, using it as a guideline to generate an image.

1.3.2. Privacy Protection

Since there is more and more data shared over the internet, removing objects in the photos has become an inevitable procedure for privacy problems.

Sun et al. have proposed a technique [22] for obscuring identities by inpainting heads in an image. This technique is effective at defeating face recognition systems. In contrast, Ma et al. have done the opposite [23] by inpainting masks onto face images to enhance the performance of face verification. For removing objects, Upenik et al. [24] propose an object removal technique in a reversible manner. Only those who possess the private decryption key have access to the hidden data.

1.3.3. Pose-Guided Image Synthesis

There has been a lot of research in the field of learning human appearance from a single image in recent years. The task can be introduced as resynthesizing the view of human faces or bodies.

Deng et al. introduced a UV completion network to generate synthetic faces with arbitrary poses [25]. The UV map is inpainted and the completed UV map, together with the corresponding 3D model, is used to synthesize 2D face images. Grigorev et al. [26] performed some more broad research, not only resynthesizing the view of faces, but also focusing on human bodies, including faces and garments.

1.3.4. Digitization of Cultural Heritage

Natural disasters, economic development, and tourism have put many rare cultural heritages in danger. It is crucial to protect and preserve these cultural heritage sites [27]. Utilizing inpainting algorithms on murals can aid in their digitization. Despite being corrupted and small in scale, Wen et al. trained a network with general datasets and found that it performed well on murals [28]. Moreover, Chen et al. [29] generated enough training data using a sliding window method in the augmentation process, allowing for the original size murals to be divided into smaller patches. To improve the stability of training, Cao et al. proposed [30] an enhanced consistent GAN model, which not only improves the generalization ability, but also speeds up the computation.

1.3.5. Remote Sensing

Led by development of satellite technology, remote sensing images have been widely used in various applications, including scene recognition [32], object detection [33], land-use classification [34]. However, because of the poor working conditions of satellite sensors and weather conditions, remote sensing images often suffer from dead pixels, artifacts, cloud and shadow covers [35]. Removing the covers and restoring the original images, in other words inpainting, are crucial for the subsequent image processing and application.

Shao et al. [35] in their research propose to solve three typical reconstruction tasks: Landsat ETM+ SLC-off; cloud removal; and shadow removal. The network is pretrained on ImageNet [31]. Xu et al. [34] utilized the enhanced consistent GAN model, used reconstruction loss to generate straight edges with reasonable structure, and used adversarial loss to produce complex edges using the results of reconstruction loss.

1.3.6. Virtual reality

The widespread use of mobile devices has created new opportunities for the application of image and video inpainting, particularly in the realm of augmented or virtual reality. For such kinds of applications, a real-time video inpainting pipeline which can be used on mobile devices would be highly useful for a wide range of tasks.

Julie Alfosine et al. [36] made use of two 3D convolutional neural networks to inpaint panoramic videos. Alasdair Newson et al. [37] propose an automatic video inpainting algorithm which is able to deal with many challenging situations which can happen in video inpainting, such as reconstruction of dynamic textures, multiple moving objects, and moving background. Niclas Scheuing [38] inspired by patch-matching algorithms implemented an augmented reality application that removes a tracked object in real-time and runs on mobile platforms.

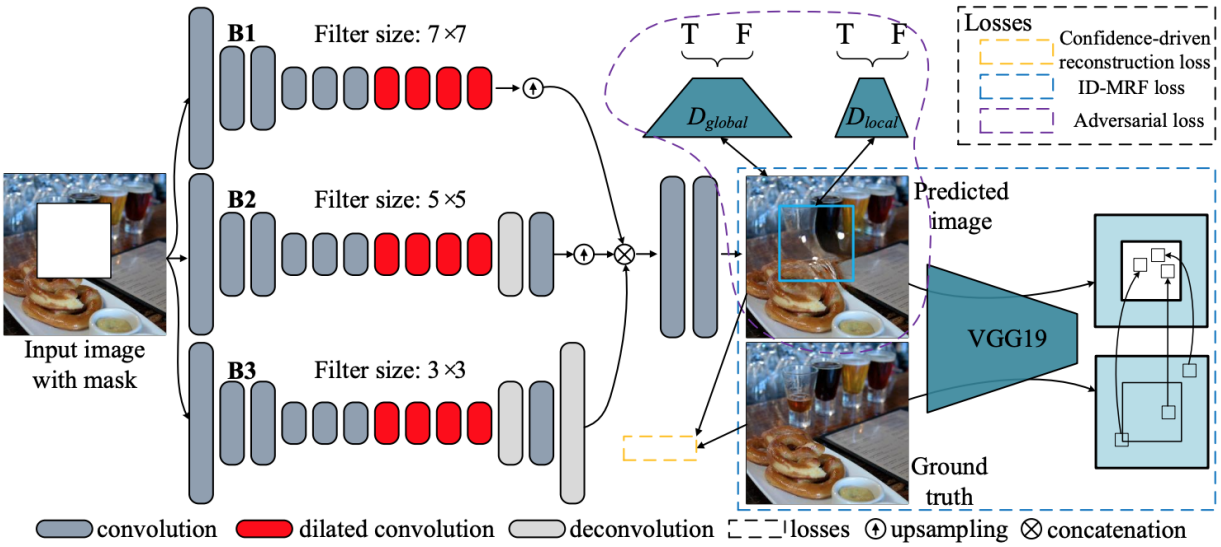
1.4. Inpainting Methods

Judging on the past researches and reviews of video inpainting realm [1], [2], the main deep learning based methods of video inpainting can be divided into convolution based, generative, flow based/copy paste inpainting.

1.4.1. Convolution Based Inpainting

Convolutional based video inpainting methods use CNN in its base. The methods use convolutional principles to analyze the missing or corrupted area in the video and generate a relevant replacement for the missing or deteriorated content based on the patterns and features it learns from the surrounding pixels.

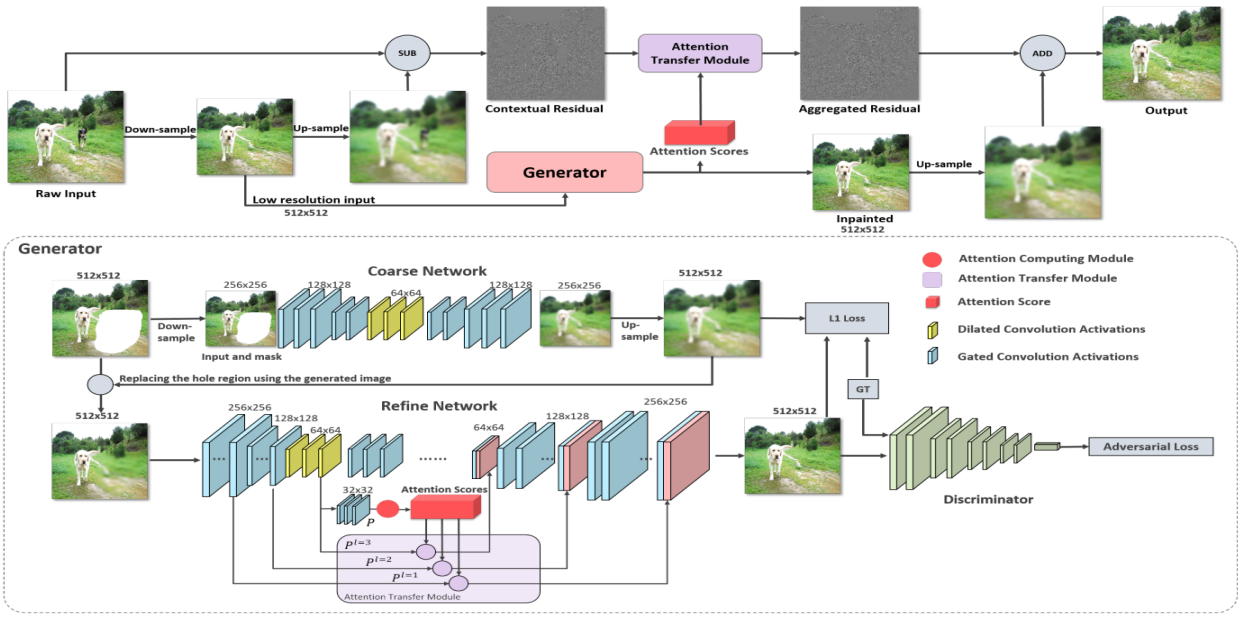
Yi Wang et al. [3] in their research paper the authors propose a generative multi-column network for image inpainting. The multi-column structure is used since it can decompose images into components with different receptive fields and feature resolutions (Fig. 1.2). The method has difficulties dealing with large-scale datasets with thousands of diverse object and scene categories. When data falls into a few categories, the proposed method works best, since the ambiguity removal in terms of structure and texture can be achieved in these cases.



1.2 Fig. Generative Multi-column Convolutional Neural Network [3]

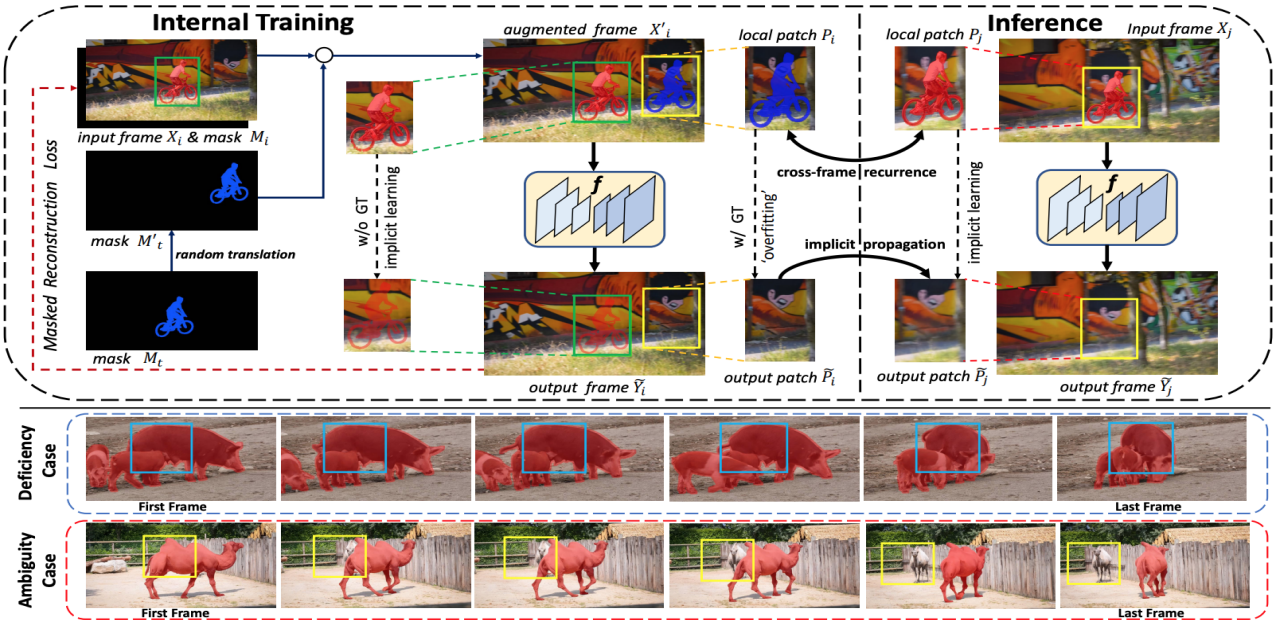
Guilin Liu et al. [4] introduced the use of partial convolutions, where the convolution is masked and renormalized to be conditioned on only valid pixels. The given model can easily process and inpaint holes of any shape, size, location, or distance from the image borders. However, there is one limitation of the authors' method - it performs with poor results for some sparsely structured images and struggles on the largest of holes.

Zili Yi et al. [5] propose a Contextual Residual Aggregation mechanism to enable the completion of ultra high-resolution images with limited resources. Inside it has Light-Weight Gated Convolutions to improve the inpainting quality, computation, and speed (Fig. 1.3). In general the increase of resolutions and hole size does not deteriorate the inpainting quality and does not essentially increase the processing time in our framework.



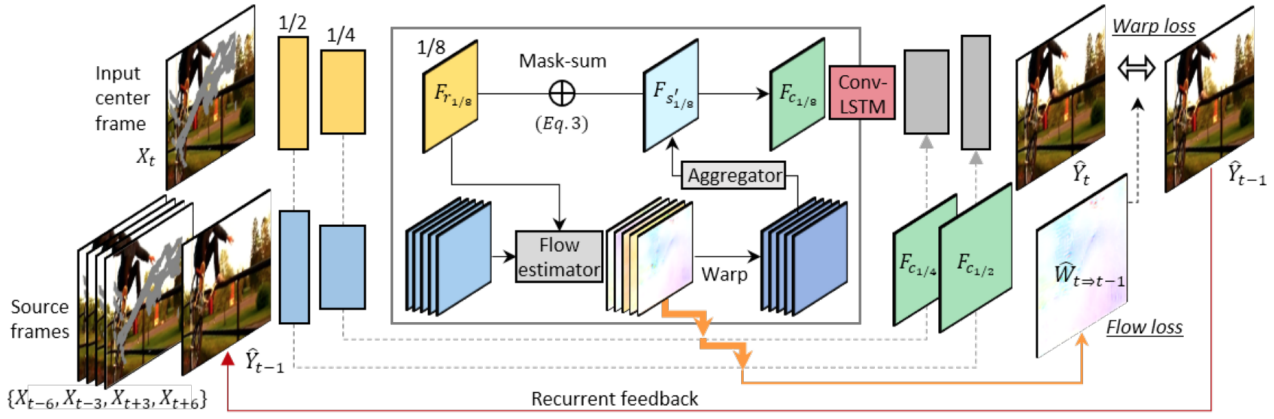
1.3 Fig. Contextual Residual Aggregation General Pipeline [5]

Hao Ouyang et al. [6] made a presentation of an internal video inpainting method, which implicitly propagates information from the known regions to the unknown parts. They implicitly propagate long-range information using an overfitted CNN without explicit guidance like optical flow (Fig. 1.4). It performs well with challenging cases such as large masks, complex motion, and long-term occlusion with the proposed regularization. Nevertheless, the given model has two main drawbacks, the method is not real-time and sometimes it can semantically incorrect restore the deteriorated details in the video.



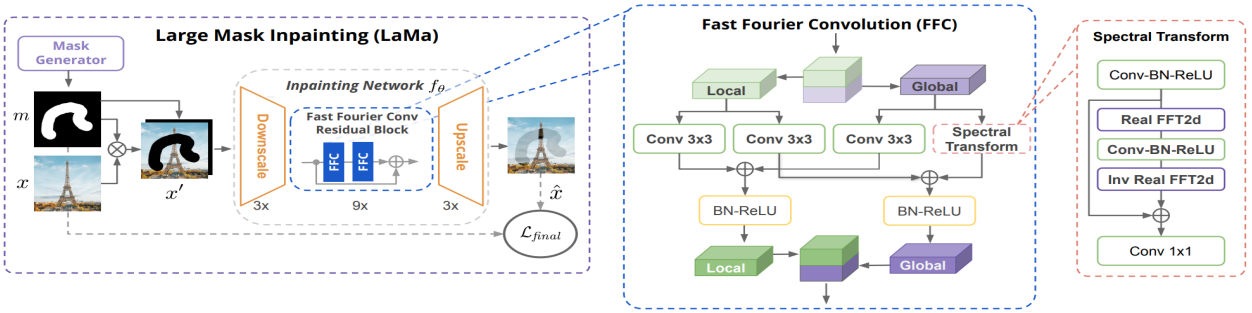
1.4 Fig. Implicit Long-range Propagation Pipeline [6]

Dahun Kim et al. [7] casted video inpainting problem as a sequential multi to single frame inpainting task and present a brand new deep 3D-2D encoder-decoder network (Fig. 1.5). The given method collects features from the neighbor frames and generates missing content based on them. But the model results with color saturation artifacts when the video has a large and long occlusion in its frames.



1.5 Fig. Implicit Long-range Propagation Pipeline [7]

Naejin Kong et al. [8] developed a new inpainting network architecture that uses fast Fourier convolutions, which have the image wide receptive field (Fig. 1.6). Used FFC improves perceptual quality and parameter efficiency of the presented network, the inductive bias of FFC allows the network to generalize to high resolutions that are not introduced to the model during the training.



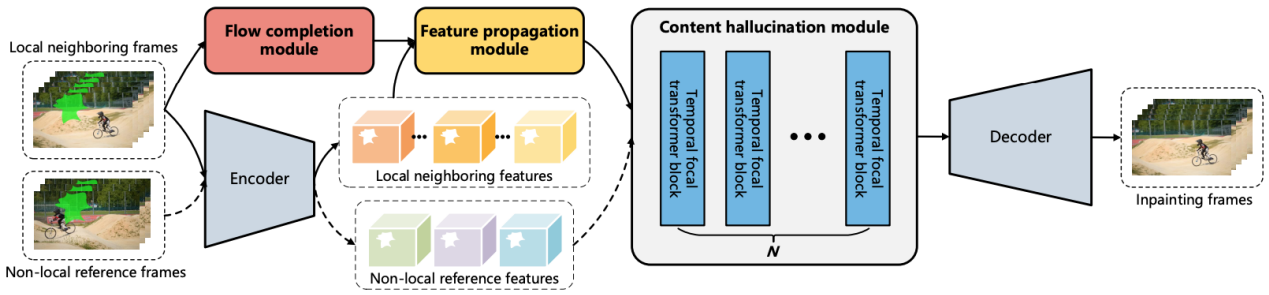
1.6 Fig. Large Mask Inpainting Schema [8]

1.4.2. Flow Based Inpainting

Optical flow is used to determine the movement of objects and background in the video by tracking the displacement of particular pixels from frame to frame. The flow can be forward or backward, when the displacement is computed from the first frame forward, or from the last in reverse order.

Riu Xu et al. [9] in this paper utilizes high computational flow with Deep Flow Completion Network which consists of three sub-networks. Each of the subnetworks takes inputs resized to 1/2, 2/3 and 1 of the original size. Deep Flow Completion network is introduced to handle arbitrary missing regions, complex motions, and maintain temporal consistency. But the given method fails when the completed object flow is inaccurate and the propagation process cannot amend that.

Zhen Li et al. [10] proposed an End-to-End framework for Flow-Guided Video Inpainting, which includes three modules in it, including flow completion, feature propagation, and content hallucination modules (Fig. 1.7). The given model performs with a state of the art accuracy and high efficiency. However it has some limitations, the same as the previous work. With large motion or a large amount of missing object details across through the frames, the introduced solution may produce a lot of artifacts of masked regions.



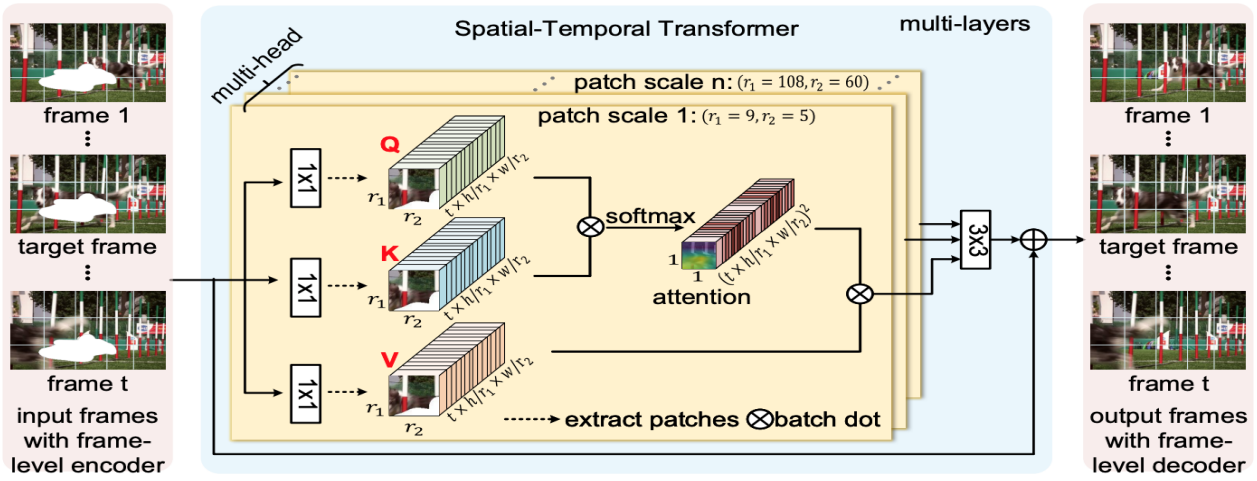
1.7 Fig. Overview of the Flow-Guided Video Inpainting Framework [10]

Sungho Lee et al. [11] presented a network based on three stages - alignment, copy, and paste. The alignment step includes aligning reference frames with test frames, in other words matching similarities between the frame pixels. The copy step acts with encoder and context matching modules. The encoder module passes the aligned frames and their masks through the feature encoders. Then the results are passed to the context matching module to produce the weight of each pixel in contributing to the missing region. Finally, the paste module takes the weights and concatenated frame features and outputs resultant images. So basically this approach was introduced as opposed to optical flow but has the same core idea, the given solution is able to extract valid pixels from distant frames giving accurate scene reconstruction.

1.4.3. Generative Inpainting

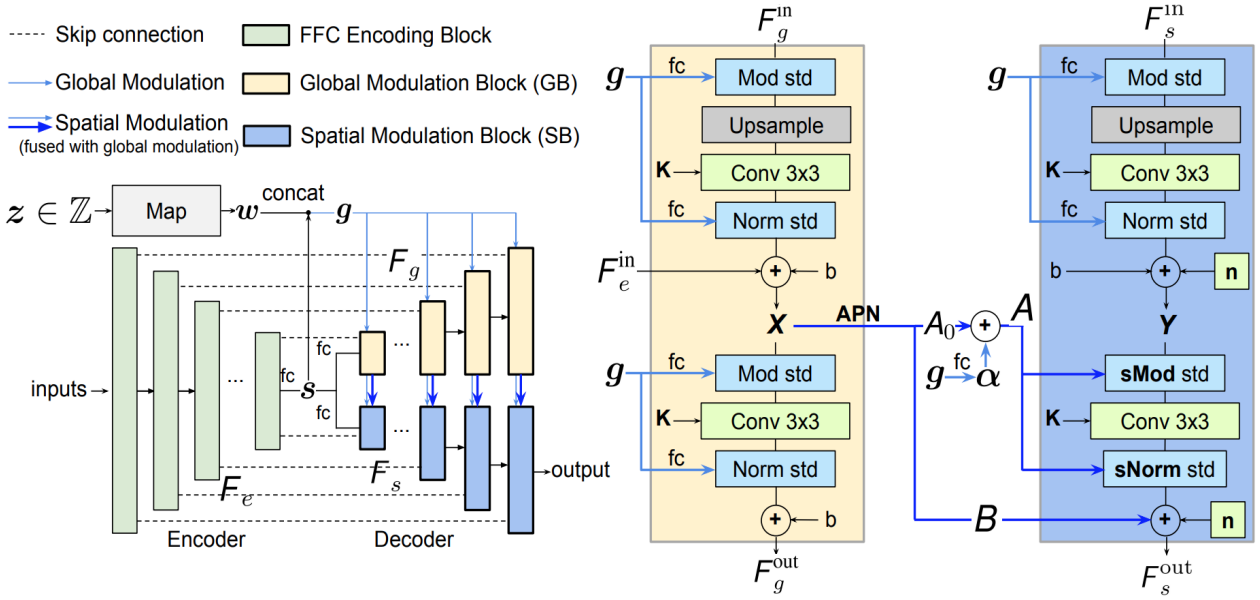
Generative video inpainting is a method of filling deteriorated parts of the video by using generative models such as Generative Adversarial Networks. The model is used to generate new, appropriate frames to replace the corrupted or masked parts of the video. GAN based models cam perform well both with image and video inpainting. However, the best performing models are considered to be transformer based models, but they have the massive drawback - very high memory requirements in attention layers, limiting their input size, which must be effectively handled to perform with high accuracy.

Yanhong Zeng et al. [12] designed a model, which is based on a multi scale and multi head transformers, called Spatial-Temporal Transformers Network. Spatial-Temporal Transformations for takes as input a window of frames distributed across the target frame and frames uniformly sampled from the rest part of the video (Fig. 1.8). Instead of processing full frames, the presented model uses patches of different scales. The developed network has three parts, frame-level encoder, multi-layer multi-head spatial-temporal transformers, frame-level decoder. In other words, the network simultaneously fills the missing regions in all input frames by self-attention. But the given solution has one disadvantage.



1.8 Fig. Overview of the Spatial-Temporal Transformer Networks [12]

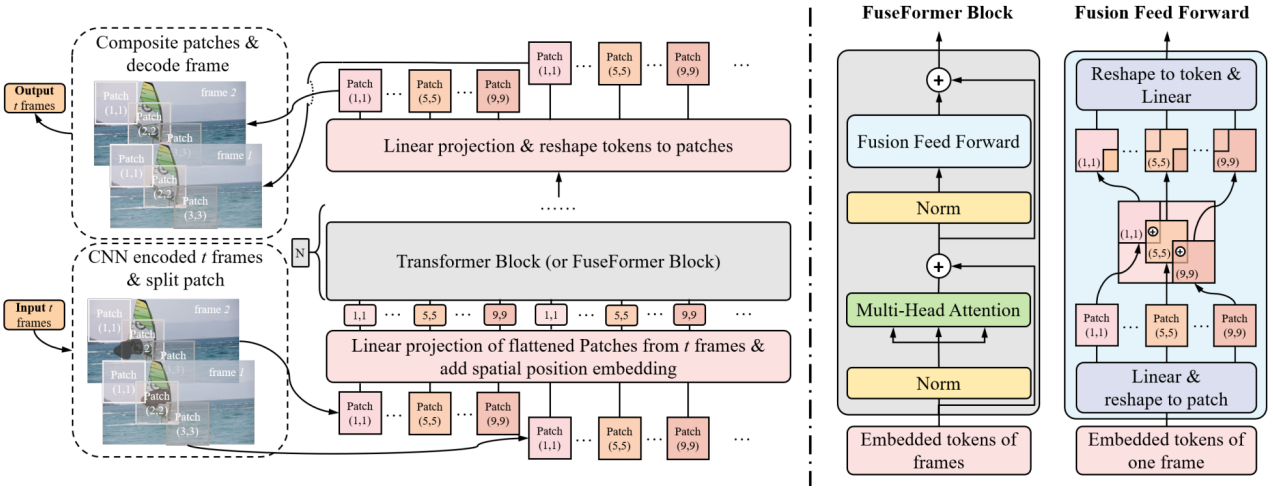
Haitian Zheng et al. [13] in their research proposed a cascaded modulation GAN (Fig. 1.9). The neural network design presented with an encoder which utilizes Fourier convolution blocks that extract multi-scale feature representations from the input image with holes and a dual-stream decoder with a cascaded global-spatial modulation block at each scale level. But the model is limited in restoring humans or animals, which can be solved by training the network for the needed type of objects.



1.9 Fig. Overview of the Cascaded Modulation GAN [13]

Rui Liu et al. [14] presents decoupled spatial transformers, which are different from the described above Spatial-Temporal Transformers Network. The transformer model decouples the task of learning spatial-temporal attention into 2 sub-tasks: attending temporal object movements on different video frames at same spatial locations and attending similar background textures on the same frame of all spatial positions. The method relies on a stacked feature encoder (CNN), one spatially decoupled and one temporarily decoupled transformer and a decoder. Aggregated features for a particular single frame are then divided into patches and processed via the spatially decoupled transformer.

Rui Liu et al. [15] also designed the FuseFormer model (Fig. 1.10). In general when transformers are used in video inpainting that requires fine-grained representation, the existing methods are suffering from blurry edges in detail due to the hard patch splitting. So the authors focused in this research to tackle the problem by introducing the FuseFormer model, it is a transformer model which was designed for video inpainting via fine-grained feature fusion based on novel Soft Split and Soft Composition operations. The soft split operation divides the feature map into many patches, while the soft composition performs by binding different patches into a whole feature map where pixels in overlapping regions are summed up.



1.10 Fig. Overview of the FuseFormer [14]

All in all, creating a full fledged application for solving video inpainting tasks is a complex engineering challenge. During this section lots of impressive, accurate and performative methods were reviewed, but all of the methods have several limitations, which result in noticeable artifacts in the output video. Some of these gaps are problems with processing videos in high resolution, inpainting videos with large motion of the target object on the video, text hallucination, semantic background recreation, shades, reflections inpainting, etc.

One of the recent papers - Omnimate, produced by Google, Oxford and Weizmann institute of Science [16] helps to struggle with the shadows or reflections of the inpainted object. The model includes a 2D UNet [17] that processes the video frame by frame. For each frame, the rough object mask is computed using off-the-shelf techniques to mark the major moving objects in the scene. During the inpainting process the model has to infer: omnimattes - pairs of continuous-valued opacity maps (mattes) and RGB images that capture not only the i-th moving object but also all the scene elements that are correlated with it in space and time (e.g., reflections, shadows, attached objects, etc.), and a refined optical flow field for each layer, and finally a background RGB image.

1.5. Loss Function

Loss function measures the error between the predicted and actual output during the neural network training. It is used to optimize the network's parameters to increase the accuracy and reduce the error, in other words they penalize the deviation between network predictions and true data labels. There are two basic loss functions, reconstruction and adversarial loss.

1.5.1. Reconstruction loss

Reconstruction loss calculates the pixel-level difference between the neural network prediction and the ground truth image. reconstruction loss can be split into weighted reconstruction

loss and multi-scale reconstruction loss. The first one pays more attention to the missing regions that are close to the boundaries of holes, while the second one progressively refines the network prediction for the missing regions at each scale of the decoder.

1.5.2. Adversarial loss

Adversarial loss utilizes a discriminator network to detect inpainted video from the source video. The inpainting network is trained to generate videos that cannot be distinguished from the source videos by the discriminator network. There are some types of adversarial loss functions, some of them are the global and local loss, the first discriminator takes the whole input image, the second one takes the patch located around the missing regions of the image.

There are some other types of loss function like perceptual loss, focused on the difference between the high-level representations of input images. Total variation regularization loss. It is generally applied in image denoising. TV loss computes the difference between the adjacent pixels in the missing regions.

Generally, in real life tasks the multi-loss model is created to train a video inpainting model, with different weights to balance their relative importance.

1.6. Quality metrics

Quality metrics are used to indicate the effectiveness of the chosen method, e.g. of the trained model on unseen data. The main metrics used during video and image inpainting are peak signal-to-noise ratio (PSNR), Structural Similarity Index (SSIM), Local Image Patch Similarity (LIPS), Fréchet Inception Distance (FID).

PSNR is widely used in image compression to measure the quality of reconstruction, SSIM measures the similarity within two videos, perceiving the change in structural information, LIPS measures the performance of the image restoration models, FID stands to compare the statistics of the generated samples with the real samples.

1.7. Datasets

There are two main generally available datasets for video inpainting, DAVIS (Densely Annotated VIdeo Segmentation) [18] and Youtube-VOS [19].

DAVIS dataset consists of video sequences with fully annotated objects and backgrounds.

Youtube-VOS consists of a Training set: 3471 videos, 65 categories and 6459 unique object instances. Validation set: 507 videos, 65 training categories, 26 unseen categories and 1063 unique object instances. Test set: 541 videos, 65 training categories, 29 unseen categories and 1092 unique

object instances. Also it has a video inpainting dataset which contains a set of videos with artificially introduced holes or missing regions.

The described datasets are widespread and have the general purpose, have lots of object classes inside, all the analyzed researches use them to test the accuracy. However, there are exists some more datasets, but they are not easily available and more specified for the particular task.

1.8. Analysis summary

The analysis of various deep learning based video and image inpainting methods was performed in this part. The following conclusions are generated from analysis above:

- deep learning based inpainting is relatively new approach to restore deteriorated parts of the video, most of the breakthrough researches has been introduced in recent years;
- most of the methods perform with high accuracy, but all of the methods have several limitations, which result in noticeable artifacts in the output video;
- some of these gaps are problems with processing videos in high resolution, inpainting videos with large motion of the target object on the video, text hallucination, semantic background recreation, shades, reflections inpainting, etc;
- most successful researchers are using flow-based or generative approaches.

Based on the given analysis the the limitations of the modern deep learning based inpainting methods were revealed, the most essential of them are incapability to inpaint shades or reflections of the target objects, poor accuracy while inpainting the object with high motion.

After research the main problem was formulated, it is a gap in research projects which focus on processing high motion videos and inpainting of reflections and shades of the objects.

Based on the problem the tasks were formulated, these are to develop an algorithm, which will be capable to handle relevant limitations of deep learning inpainting methods such as inpainting shades of the object in the video better than the reviewed solutions in real-time, implement the system and set up experiments to tune the developed model pipeline.

The goal is to develop the system which will inpaint the unwanted objects and their shades from video scenes during the live sport events.

Conclusions

1. Lorem ipsum dolor sit amet, eam ex decore persequeris, sit at illud lobortis atomorum. Sed dolorem quaerendum ne, prompta instructior ne pri. Et mel partiendo suscipiantur, docendi abhorreant ea sit. Recteque imperdiet eum te.
2. Eu eum decore inimicus consetetur, cu usu habeo corpora intellegam. Ut antiopam efficiendi deterruisset sit. Mel sint eirmod id, qui quot virtute id, dolor nemore forensibus usu id. Fugit dolore voluptatum cu vim. An vix veniam graecis insolens, sit posse iusto id. Ut vim ceteros percipit, id quo ubique recusabo, eum sint lucilius ea. In sumo inani numquam has.

List of references

1. Image inpainting based on deep learning: A review [online] September 8, 2022 [viewed 01/10/2022] available from https://www.sciencedirect.com/science/article/pii/S1566253522001324?casa_token=RuScUag8uk8AAAAA:Tu7tkwyddq-cZjs1YJwMaP4fYWs34Cf71-kzhTPdFZSE1tuI_cctcZxbM0X9dsaRSX-mhjbJFf_Z.

2. Deep learning for image inpainting: A survey [online] September 20, 2022 [viewed 05/10/2022] available from https://www.sciencedirect.com/science/article/pii/S003132032200526X?casa_token=ydw-XP_Lji4AAAAA:PgX_mjvRbpgy4APsuUPHU-zKko-5q0W9tVC0psGdTe1LZul50q8KxNogR7aWr6UmB4Roe0rtW9Np.

3. Image Inpainting via Generative Multi-column Convolutional Neural Networks [online] October 20, 2018 [viewed 17/10/2022] available from https://arxiv.org/pdf/1810.08771.pdf.

4. Image Inpainting for Irregular Holes Using Partial Convolutions [online] 2018 [viewed 18/10/2022] available from https://openaccess.thecvf.com/content_ECCV_2018/html/Guilin_Liu_Image_Inpainting_for_ECCV_2018_paper.html.

5. Contextual Residual Aggregation for Ultra High-Resolution Image Inpainting [online] May 19, 2020 [viewed 22/10/2022] available from https://ieeexplore.ieee.org/document/9156377.

6. Internal Video Inpainting by Implicit Long-range Propagation [online] August 17, 2021 [viewed 27/10/2022] available from https://tengfei-wang.github.io/Implicit-Internal-Video-Inpainting/index.html.

7. Deep Video Inpainting [online] May 5, 2019 [viewed 30/10/2022] available from https://ieeexplore.ieee.org/document/8953258.

8. Resolution-robust Large Mask Inpainting with Fourier Convolutions [online] September 15, 2021 [viewed 05/11/2022] available from https://github.com/saic-mdal/lama.

9. Deep Flow-Guided Video Inpainting [online] May 8, 2019 [viewed 07/11/2019] available from https://ieeexplore.ieee.org/document/8954458.

10. Towards An End-to-End Framework for Flow-Guided Video Inpainting [online] April 7, 2022 [viewed 8/11/2022] available from https://github.com/MCG-NKU/E2FGVI.

11. Copy-and-Paste Networks for Deep Video Inpainting [online] August 30, 2019 [viewed 11/11/2022] available from https://www.researchgate.net/publication/339555483_Copy-and-Paste_Networks_for_Deep_Video_Inpainting.

12. Learning Joint Spatial-Temporal Transformations for Video Inpainting [online] July 20, 2020 [viewed 21/11/2022] available from https://arxiv.org/pdf/2007.10247.pdf.

13. Image Inpainting with Cascaded Modulation GAN and Object-Aware Training [online] July 21, 2022 [viewed 22/11/2022] available from https://paperswithcode.com/paper/cm-gan-image-inpainting-with-cascaded/review.

14. Decoupled Spatial-Temporal Transformer for Video Inpainting [online] April 14, 2021 [viewed 22/12/2022] available from https://arxiv.org/pdf/2104.06637.pdf.

15. FuseFormer: Fusing Fine-Grained Information in Transformers for Video Inpainting [online] September 7, 2021 [viewed 23/12/2022] available from https://ieeexplore.ieee.org/document/9710180.

16. Omnimatte: Associating Objects and Their Effects in Video [online] October 1, 2021 [viewed 15/01/2023] available from https://ieeexplore.ieee.org/document/9577599.

17. U-Net [online] September 8, 2022 [viewed 15/01/2023] available from https://en.wikipedia.org/wiki/U-Net.

18. DAVIS: Densely Annotated VIdeo Segmentation [online] 2017 [viewed 15/01/2023] available from https://davischallenge.org/index.html.

19. YouTube-VOS: A Large-Scale Benchmark for Video Object Segmentation [online] 2017 [viewed 15/01/2023] available from https://youtube-vos.org/.

20. FaceShop: Deep Sketch-based Face Image Editing [online] April 7, 2018 [viewed 15/01/2023] available from https://www.researchgate.net/publication/324744575_FaceShop_Deep_Sketch-based_Face_Image_Editing.

21. SC-FEGAN: Face Editing Generative Adversarial Network with User's Sketch and Color [online] February 18, 2018 [viewed 15/10/2023] available from https://ieeexplore.ieee.org/document/9010058.

22. Natural and Effective Obfuscation by Head Inpainting [online] March 16, 2018 [viewed 15/10/2023] available from https://ieeexplore.ieee.org/document/8578628.

23. Contrastive attention network with dense field estimation for face completion [online] April 14, 2022 [viewed 15/10/2023] available from https://www.sciencedirect.com/science/article/abs/pii/S0031320321006415.

24. Inpainting in Omnidirectional Images for Privacy Protection [online] April 17, 2019 [viewed 15/10/2023] available from https://ieeexplore.ieee.org/document/8683346.

25. Coordinate-based Texture Inpainting for Pose-Guided Image Generation [online] January 9, 2020 [viewed 15/10/2023] available from https://ieeexplore.ieee.org/document/8953923.

26. UV-GAN: Adversarial Facial UV Map Completion for Pose-invariant Face Recognition [online] December 16, 2018 [viewed 15/10/2023] available from https://ieeexplore.ieee.org/document/8578839.

27. Investigation of correlation between sub-scale and full-scale models of simulator [online] October 10, 2021 [viewed 15/10/2023] available from https://www.sciencedirect.com/science/article/pii/S0376042121000981.

28. Ancient mural restoration based on a modified generative adversarial network [online] January 29, 2020 [viewed 15/10/2023] available from https://heritagesciencejournal.springeropen.com/articles/10.1186/s40494-020-0355-x.

29. Journal of Physics: Conference Series Image Inpainting for Digital Dunhuang Murals Using Partial Convolutions and Sliding Window Method [online] March 20, 2019 [viewed 15/10/2

023] available from https://iopscience.iop.org/article/10.1088/1742-6596/1302/3/032040/meta.

30. Application of Enhanced Consistent Generative Adversarial Network in Mural Repairing [online] February 13, 2020 [viewed 15/10/2023] available from https://www.scopus.com/record/display.uri?eid=2-s2.0-85092081967&origin=inward.
31. Deep Learning Based Feature Selection for Remote Sensing Scene Classification [online] November 12, 2015 [viewed 15/10/2023] available from https://www.scopus.com/record/display.uri?eid=2-s2.0-84947127828&origin=inward.
32. Hierarchical Semantic Propagation for Object Detection in Remote Sensing Imagery [online] January 15, 2020 [viewed 15/10/2023] available from https://ieeexplore.ieee.org/document/8960460.
33. Missing data reconstruction in VHR images based on progressive structure prediction and texture generation [online] January 17, 2021 [viewed 15/10/2023] available from https://www.sciencedirect.com/science/article/abs/pii/S0924271620303270.
34. Context-Based Multiscale Unified Network for Missing Data Reconstruction in Remote Sensing Images [online] May 15, 2020 [viewed 15/10/2023] available from https://ieeexplore.ieee.org/document/9198931.
35. Reconstruction by inpainting for visual anomaly detection April 21, 2021 [viewed 15/10/2023] available from https://www.sciencedirect.com/science/article/abs/pii/S0031320320305094.
36. Time-consistent panoramic video inpainting with 3D convolutional neural networks July 12, 2019 [viewed 15/10/2023] available from https://koasas.kaist.ac.kr/handle/10203/282940.
37. Video Inpainting of Complex Scenes May 19, 2018 [viewed 15/10/2023] available from https://epubs.siam.org/doi/abs/10.1137/140954933.
38. Real-time Hiding of Physical Objects in Augmented Reality May 30, 2018 [viewed 15/10/2023] available from https://www.research-collection.ethz.ch/handle/20.500.11850/283601.
39.
40.
41.

List of information sources (if needed)

1. Information source
2. Information source
3. Information source
4. Information source