Deep Learning Approaches for Single Image and Video Super-Resolution: Leveraging CNN, RNN and GAN Architectures for Upscaling

by

Moinul Hossain Bhuiyan
20301002
Istiaq Ahmad
20301056
Abdullah Al Mamun
20301062
Shouvik Banerjee Argha
20301118
Labib Sadman Azam
21301643

A thesis submitted to the Department of Computer Science and Engineering
in partial fulfillment of the requirements for the degree of
B.Sc. in Computer Science

Department of Computer Science and Engineering
Brac University
September 2023

Declaration

It is hereby declared that

1. The thesis submitted is our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. We have acknowledged all main sources of help.

Student's Full Name & Signature:

Moinul Hossain Bhuiyan
20301002

Istiaq Ahmad
20301056

Abdullah Al Mamun
20301062

Shouvik Banerjee Argha
20301118

Labib Sadman Azam
21301643

Approval

The thesis titled "Deep Learning Approaches for Single Image and Video Super-Resolution: Leveraging CNN, RNN and GAN Architectures for Upscaling" submitted by

1. Moinul Hossain Bhuiyan (20301002)
2. Istiaq Ahmad (20301056)
3. Abdullah Al Mamun (20301062)
4. Shouvik Banerjee Argha (20301118)
5. Labib Sadman Azam (21301643)

Of Summer, 2023 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of B.Sc. in Computer Science on September 20th, 2023.

Examining Committee:

Supervisor:
(Member)

Dr. Muhammad Iqbal Hossain

Associate Professor
Department of Computer Science and Engineering
BRAC University

Abstract

In a world where visual content plays a crucial role in anything imaginable, the need for sharper, more detailed images and videos has never been more important. This research paper explores innovative approaches to improve the quality of both single images and videos through the application of Deep Learning techniques, specifically Convolutional Neural Networks (CNN) Recurrent Neural Networks (RNN), and Generative Adversarial Networks (GAN). Upscaling is essential because many people out there own older devices that can not properly output high-definition content. These devices struggle to display high-quality videos or images that are stored locally. That is the reason why video upscaling methods are needed to help these devices to first render lower-quality content and then enhance it to the quality people desire to see. This research explores how smart computer systems, using advanced techniques of CNNs, RNNs, and GANs, may remarkably increase the quality of pictures and videos. Imagine converting grainy photos into clear, vivid ones and making visuals smoother and more detailed. The examination delves into how different technologies collaborate to enhance individual photographs and videos. The results not only improve entertainment but also have practical consequences in the medical sector, security, and beyond. By harnessing the power of deep learning, a new level of clarity and richness is added to the pictures the user sees every day. Also, the research paper aims to demonstrate the potential of deep learning techniques to improve the visual quality of any Low Resolution content.

Keywords: Super Resolution (SR), Convolutional Neural Networks (CNN), Recurrent Neural Networks (RNN), Generative Adversarial Networks (GAN)

Table of Contents

List of Figures

Nomenclature

The next list describes several symbols & abbreviation that will be later used within the body of the document

CNN Convolutional Neural Network

$EGAN$ Enhanced Generative Adversarial Network

$FFMPEG$ Fast Forward Moving Picture Experts Group

$FPGA$ Field Programmable Gate Array

GAN Generative Adversarial Network

$MMCNN$ Multi-Memory Convolutional Neural Network

RNN Recurrent Neural Network

$SISR$ Single Image Super Resolution

SR Super Resolution

VSR Video Super Resolution

Chapter 1

Introduction

1.1 Background

In a world filled with visual content, the desire for crisper, more colorful pictures and videos has become a fundamental component of an individual's digital experience. It is a common phenomenon that, when a user has looked at a photo or watched a video that appeared somewhat hazy or lacking the desired sharpness? The number of these people is a lot. Many of us have this difficulty, particularly on older devices that struggle to display high-definition material. Fortunately, this research looks into an interesting field of technology that may convert these less-than-perfect pictures or videos into something absolutely extraordinary

High-definition displays have grown less expensive and more common since the mid-2000s, and can now be found in TVs, computer monitors, and even the devices a common person uses most, which is their smartphones. In the era where online streaming is at its peak, here only video is responsible for 88% of total internet traffic in 2023 [6].

Before the machine learning revolution of the 2010s, most upscaling tasks were accomplished using proprietary algorithms; in recent years, scholars have moved their emphasis towards employing deep neural networks instead, particularly making use of CNNs or GANs. This method has previously discovered widespread usage in industries such as picture editing [7] and video games [16].

While video production has Adapted to changes in display technology by simply recording at higher resolutions, material generated before these resolution shifts is trapped at its prior resolution, and does not look well on today's monitors or displays without the use of some form of scaling.

In this case, the video upscaling method plays a significant role. Using neural networking, there are some techniques that can accomplish the task. Imagine having the ability to take those older, lower-quality videos and pictures and suddenly make them appear as though they were recorded with the newest high-end equipment. That's exactly what this research paper emphasizes on. This paper emerges into the area of deep learning, a form of computer intelligence that's surprisingly effective at learning from examples and improving things. In particular, this paper will investigate three strong technologies: CNNs, RNNs, and GANs.

Additionally, bicubic interpolation is used to upgrade low-resolution information, such as vintage broadcast recordings in PAL resolution (720 x 576) [2]. video filters in FFMPEG [14], with Lanczos [1] filtering being utilized in various tests.But, " In terms of a single image, the CNN model is highly efficient where the model can detect the object and recognize it, opening a new era where images are being processed automatically" [8].

To enhance the resolution of single images, it is seen utilizing the power of CNNs, which excel at recognizing patterns and features in visual data. After training these networks on vast datasets, it enables them to generate high-resolution (HR) images from low-resolution (LR) inputs, allowing a person to witness the final output which has more depth or definition into it.

As a result, when it comes to videos, maintaining consistent high-quality output is a challenging task. Here, RNNs come into play, as they excel at processing sequential data.

Lastly, a little bit of Artificial intelligence touch is needed to properly make an image or video look real. Here, GANs add an extra layer of magic. The abbreviation of GAN stands for generative adversarial networks. They're like artists who make the enhanced images look even more realistic. GANs help fill in the missing pieces and make the upscaled pictures and videos look like they were shot in high definition from the start.

So, eventually, if a user has ever desired sharper, more vivid images on their older devices, this study paper is a ticket to knowing how cutting-edge technology might turn a dream into reality. Welcome to the world of deep learning super-resolution, where photographs and videos are about to receive a stunning makeover.

1.2 Problem Statement

Video upscaling plays a critical role in preserving the quality and interoperability of visual material in today's digital age. When a video is not upscaled as required, it may give rise to a number of difficulties that influence both the watching experience and the overall efficacy of the information. If a video has a low quality and has not been correctly enhanced then the initial problem that can be detected is the loss of details. In a low-quality video, it lacks the essential pixel density. In conclusion, a viewer may find diminished clarity, resulting in it being harder to recognize objects, or facial expressions within the content. Additionally, a low-quality video that is not upscaled might seem hazy or pixelated, particularly when presented on bigger displays or displays that support higher resolutions.

Not only that, content inside the video, such as subtitles, on-screen graphics, or essential comments, might become unreadable when the video is not upscaled. This may hamper understanding, especially in instructional or informative topics. When analyzing the role of a video, it typically acts as a medium for transmitting information, generating emotions, or describing tales. When details are lost owing to the lack of upscaling, the footage's capacity to effectively deliver its intended message is weakened, resulting in misconceptions or distracted viewers. All those aspects add up to a terrible user experience.

Besides, the screens that are used today, these all have precise aspect ratios. For instance, a high-definition (HD) video has to be played on a 16:9 ratio. If it has to scale up to 4K, the aspect ratio would be 16:9 or 21:9. Whenever a video is not upscaled to fit the right aspect ratio, it may seem stretched, deformed, or have black bars on the sides, which is aesthetically unpleasant and breaks the intended composition of the video. Users do not prefer having bezels in the videos, they prefer a much more immersive experience. As a high-resolution display is meant to deliver immersive and cinematic viewing experiences. While the video is not enhanced to make use of these capabilities, consumers might feel disenchanted with the content and lose out on the engaging characteristics that higher resolutions may bring. In terms of media consumption, a person can understand that video upscaling plays a crucial role in one's day-to-day life. Not just media consumption it also plays a significant role in the gaming industry.

Supersampling approaches like DLSS (Deep Learning Super Sampling) and FSR (Fidelity Super Resolution) have become vital tools in the area of gaming. Their value rests in their capacity to dramatically improve the game experience on many platforms.

One of the key advantages of these technologies is their potential to increase the visual quality of video games. DLSS and FSR do this by automatically upscaling lower-resolution visuals in real time. This leads to photos that are not only crisper but also more precise and aesthetically attractive. The upgraded visuals add to a heightened feeling of realism inside the game world, making virtual settings and people more alive and engaging. Gamers may immerse themselves in finely detailed settings and notice subtler subtleties in character design, all of which contribute to the overall attractiveness of the gaming experience.

However, it's not only about appearances. DLSS and FSR also play a crucial part in ensuring games operate smoothly on a broad variety of gaming devices. By decreasing the computing burden necessary for generating high-quality images, these technologies allow games to reach greater frame rates, resulting in smoother gameplay. The importance of this cannot be emphasized, since smoother gameplay not only increases the gaming experience but also gives a competitive advantage in online multiplayer games, where split-second reflexes may make all the difference.

More on that, upscaling is akin to waving a magic wand over older computers and gaming consoles. It breathes fresh life into aged gear by allowing it to run complex, expensive games that were formerly out of reach. Users no longer need to feel confined by their equipment's limits; they can now explore the newest games with increased performance and visuals. Not only in PC or mobile gaming it creates a new dimension in Virtual Reality (VR) gaming as well. In the immersive realm of VR gaming, upscaling acts like a pair of high-tech glasses. VR headsets strive to transport users to compelling virtual worlds, and visual quality is important for a genuine experience. Upscaling boosts the clarity, detail, and realism inside VR settings. As a consequence, users may explore these virtual environments with more ease and interest. Reduced visual abnormalities and enhanced picture quality lead to a more comfortable and motion sickness-free VR experience, making the technology more accessible and attractive to a larger audience. This technology not only increases the game experience but also makes more effective use of hardware resources. By utilizing the potential of available processing power, upscaling eliminates bottlenecks, lowering system strain and overheating concerns. This optimal resource usage leads to longer device lifespans, cheaper maintenance costs, and a smoother gaming experience overall.

By applying deep learning architectures for upscaling, the gaming industry, and the individuals who love to consume content may present the user with more spectacular, detailed, and engaging experiences, finally pushing the boundaries of what's possible in the media and the gaming world.

1.3 Research Objectives

In this research paper, the purpose is to build a hybrid model that utilizes CNN, RNN, and GAN methods. The fundamental purpose of this paper is to examine and develop the use of deep learning methods, including CNNs, RNNs, and GANs, in the arena of single image and video super-resolution. The research objectives of this paper are:

- Investigate the effectiveness of CNNs in enhancing the resolution and quality of single images
- Evaluate the capabilities of RNNs to improve video super-resolution, ensuring temporal coherence between frames.
- Examine the role of GANs in refining and adding realism to the upscaled images and videos
- Explore various combinations and architectures of CNNs, RNNs, and GANs to determine optimal solutions for different scenarios.
- Assess the practical applications of deep learning super-resolution techniques in fields such as medicine, security, and entertainment.
- Provide insights and recommendations for leveraging these technologies to enhance visual content on older or lower-quality devices.

By addressing these goals, this paper will contribute to the evolution of technology that can improve the visual quality of photos and videos, making them sharper, more detailed, and acceptable for a broad variety of real-world applications.

Chapter 2

Related Work

The paper "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network" describes a novel strategy to improve the quality and speed of single-image super-resolution using deep neural networks [4]. Current techniques upscale low-resolution images before improving them, which is ineffective. The authors provide a revolutionary CNN architecture that works directly with low-resolution data and employs a smart upscaling method. This methodology exceeds conventional approaches, delivering better image quality and being significantly faster.

So, the main idea of this research paper is to highlight the importance of super-resolution (SR) in digital image processing, where the objective is to enhance the quality and resolution of low-resolution (LR) photos or videos. It highlights that SR has practical applications in numerous areas including HDTV, medical imaging, satellite imaging, face recognition, and surveillance. The research paper analyzes the problems of SR, including the loss of high-frequency information during the LR-to-HR transition and the inherent uncertainty in the mapping from LR to HR space. It additionally provides two main kinds of SR methods: multi-image SR, which depends on numerous LR photographs of the same scene, and single image super-resolution (SISR), which aims to retrieve HR information from a single LR instance exploiting implicit redundancy discovered in natural data. Both methods have to deal with the ill-posed nature of the issue and need restrictions or previous information to guide the reconstruction process.

After that, the paper "Real-time image upscaling with commonly available resources" discusses the rising demand for high-quality video streaming and the challenge of decreasing data transmission while preserving visual quality [11]. Traditional compression algorithms are getting closer to their limitations, increasing the need for alternative techniques. The idea provides a technique comprising reducing picture resolution on the sender side and upscaling on the receiver's side. However, the difficulty is that consumers have variable resources for upscaling, and conventional techniques may not give sufficient outcomes. The idea seeks to solve this using a real-time deep learning network built to operate on a 2 GHz CPU core, giving higher picture quality compared to standard interpolation approaches. This study illustrates the importance of bit rate in video streaming, highlighting the necessity for reducing data transmission while having acceptable picture and audio quality. It illustrates that broadcasting raw 1080p video at 60 FPS needs a high bit rate, which may be reduced by lowering the image resolution and frame rate. Compression algorithms currently assist lower bit rates, but additional reductions while still maintaining acceptable quality might lead to tremendous cost savings for video streaming services. The paper also emphasizes that contrary to expectations, consumers value smooth video delivery over video quality, making choices like lower resolution appealing if they result in smoother video playback.

More on the topic, the paper known as "Deep upscaling for video streaming: a case evaluation at SVT" analyzes the study on how deep learning, specifically a convolutional neural network (CNN), may enhance video quality via super-resolution [10]. A large-scale A/B video quality test was conducted in order to compare CNN-based upscaling to the standard bicubic method. The research results show that viewers usually prefer CNN-scaled video, but not necessarily for content that is generally upscaled. The research reveals that deep upscaling technology offers potential but needs more optimization and flexibility to become suitable for mainstream use. So, the main goal of this paper is to highlight the growth of display technology, the rising prevalence of high-definition monitors, and the growing popularity of video streaming services. It also introduces the idea of super-resolution (SR), which includes upscaling low-resolution pictures or videos to better resolutions. The paper discusses the issues connected with SR and how deep learning, especially CNNs and generative adversarial GANs, have become essential to tackling these challenges. The promise of deep super-resolution is highlighted, including its possible applications to improve the quality of historical material and decreases distribution cost for streaming services. Finally, the paper discusses the particular case study that is being conducted to assess whether viewers prefer video scaled using deep learning over traditional bicubic interpolation, highlighting the significance of subjective evaluation in addition to measurements that are objective.

Moving on, the paper "Generative Adversarial Networks for Image and Video Synthesis: Algorithms and Applications" provides the importance of GANs as a potent framework for image and video synthesis, both unconditionally and with input conditions [9]. GANs have changed the development of high-resolution, photorealistic visuals, a task previously believed tough or unachievable. Their emergence has prompted various unique applications in the creation of content. This paper presents an overview of GANs, highlighting their algorithms and applications in visual synthesis. It dives into critical strategies geared at stabilizing the famously hard GAN training process. Additionally, it covers GAN applications in image translation, image processing, video synthesis, and neural rendering. This paper addresses the fundamental concept of Generative Adversarial Networks (GANs) in the area of deep learning and its major impact on several aspects of visual content synthesis. GANs comprise a generator and discriminator network engaged in a competitive training process, resulting in the development of synthetic data that resembles actual data. GANs have successfully replaced hand-designed components in computer vision pipelines, especially for generation tasks, by deriving objective functions from training data. However, GANs are tough to train because of the changing nature of the generator's output distribution, which demands careful control of training dynamics. Various ways have been proposed to stabilize GAN training over time.

Moreover, the paper also differentiates between unconditional and conditional GAN frameworks, where conditional GANs utilize control signals for more precise generation tasks. This breakthrough has led to numerous intriguing applications in semantic picture synthesis, image-to-image translation, image processing, video synthesis, and neural rendering. The overall topic is that GANs have become an essential tool in the area of computer vision, allowing numerous creative visual content-generating applications.

After that, the paper “Algorithm and Architecture Design of High-Quality Video Upscaling Using Database Free Texture Synthesis” introduces a low-complexity super-resolution (SR) algorithm and hardware architecture that improved the quality of TV pictures in real time without costly resources [3]. The algorithm employs a texture synthesis method without having a database, producing detailed visuals by evaluating the input itself. It also preserves temporal consistency. The hardware design decreases computation by 76% utilizing a partial-sum reuse method and optimizes memory utilization using a tile-based processing approach. Experimental findings reveal that this technique produces high-quality output in real-time at a reasonable hardware cost, addressing difficulties observed in traditional scalers such as zigzag and blurring effects. The main objective of this paper is to emphasize the relevance of TV scalers in enhancing the viewing quality of low-resolution content on high-resolution screens, particularly with the increasing resolution gap between content sources and display devices. This paper’s primary goal is to highlight the significance of TV scalers in improving the viewing experience of low-resolution information on high-resolution displays, especially given the increasing resolution gap between content sources and display devices. The zigzag effect, blur effect, and flickering are a few common abnormalities in TV scaling that are examined in this study. To address this problem, interpolation-based methods and super-resolution (SR) algorithms are presented. The goal of the project is to improve image quality while addressing hardware constraints and real-time demands by integrating SR methods into TV scalers. It underlines the need for an effective SR method and architecture that is hardware-friendly in order to close the resolution gap. A summary of the paper’s organization and structure is also provided.

A new single-image upscaling method that improves picture quality and efficiency was studied in the work “Image and Video Upscaling from Local Self-examples” [15]. This approach emphasizes local self-similarity in the picture as opposed to other approaches that rely on external databases or a totally input image. It reduces search time while maintaining quality by extracting patches from small, localized portions of the input picture. Particularly effective for lesser scaling factors is the approach. It applies specialized filters for these small scalings, providing high-resolution outcomes compatible with the original picture. The algorithm is basic, efficient, and can be implemented in parallel on a GPU. It shows high-quality resolution enhancements, works perfectly with video sequences, and is capable of real-time enhancement of low-resolution videos into high-definition formats. The basic goal of this paper is to illustrate the importance and difficulties of image upscaling, a fundamental image-editing procedure. The paper highlights the limits of traditional upscaling approaches, which frequently result in artifacts and image abnormalities. It presents an innovative method that uses local scale invariance in real images, concentrating on small, localized patches in order to improve the accuracy and efficiency of upscaling. The author underlines that this creative method works better for small scaling factors and includes multiple upscaling steps employing specific filter banks to accomplish high-quality resolution enhancement. Additionally, it demonstrates the advantages of the proposed technique for both video and image sequences, along with its efficient implementation on GPUs for real-time performance.

The paper "Learning Spatio-Temporal Downsampling for Effective Video Upscaling" highlights the challenges connected to downsampling in image processing, especially in the context of videos, where improper downsampling may lead to aliasing problems including moire patterns and the wagon-wheel effect [13].To solve this challenge, the researchers offer a framework for neural networks that concurrently learns spatio-temporal downsampling and upsampling. The idea is to keep necessary patterns in the source video while enhancing reconstruction during upsampling. To maintain compatibility with standard image and video storage formats, the downsampling results are encoded as uint8 using a differentiable quantization layer. Two new modules for explicit temporal propagation and space-time feature rearrangement are presented to make more use of spatio-temporal correspondences.

Eventually, results from experiments reveal that this method considerably enhances the quality of space-time reconstruction by maintaining spatial textures and motion patterns throughout both downsampling and upsampling. Additionally, the suggested framework enables many applications, including video resampling, blurry frame reconstruction, and efficient video storage. The fundamental goal of this paper is to solve the issues of scaling high-resolution, high-frame-rate movies in small devices, such as mobile phones and glasses, where constraints in memory and bandwidth for transfer demand a trade-off between spatial and temporal resolution. The research paper notes that traditional nearest-neighbor downsampling, which is commonly used in these kinds of situations, may lead to aliasing concerns owing to high-frequency information being folded over in the downsampled frequency domain.

So, to address aliasing, the research paper recommends using deliberately spatial and temporal anti-aliasing filters, such as optical blur and motion blur, to smudge high-frequency information, allowing the reconstruction of fine features during post-capture processing. These anti-aliasing filters may be pre-designed and deployed with the downsampler during capture. The fundamental uniqueness of this study is the notion of concurrently learning both downsampling and upsampling, which is typically handled separately in traditional methods. By simultaneously learning a downsampler and an upsampler, this paper intends to retain and recover high-frequency information in both space and time during image and video processing. This technique provides advantages for video restoration tasks by enabling the co-design of an upsampler to restore missing details. In summary, the core aim of the study is to offer a unified framework that jointly learns spatiotemporal downsampling and upsampling, highlighting the necessity of keeping high-frequency details in both spatial and temporal dimensions. The paper also covers possible applications of this framework, including video resampling, fuzzy frame restoration, and efficient video storage.

The paper “Super Resolution of Videos using E-GAN” provides a unique strategy called EGAN (Enhanced Generative Adversarial Network) for video super-resolution, which comprises enhancing the details and upscaling the resolution of videos1 [12]. EGAN is highlighted as a more accessible and reliable alternative to conventional neural networks, giving greater performance with fewer artifacts It specializes at maintaining high-frequency details and delivering visually stunning outcomes from substantially downsampled input videos. EGAN varies from standard Generative Adversarial Networks by eliminating Batch Normalization layers. The research paper outlines the benefits of this decision and presents significant improvements in human perceptual quality based on thorough Mean Opinion Score (MOS) and Video Multimethod Assessment Fusion (VMAF) testing compared to state-of-the-art methods. The fundamental concept of this study is to suggest an enhanced method of super-resolution, a technique utilized to improve the resolution and visual quality of videos and images. The research paper explores the use of neural networks, especially GANs, for super-resolution and illustrates the limits and benefits of diverse neural network structures.

Besides, the research offers an enhanced version of SRGAN (Super-Resolution Generative Adversarial Network) by including Residual-in-Residual Dense (RRDB) blocks and eliminating batch normalization layers. These modifications seek to enhance training efficiency and shorten the time necessary for super-resolution. The authors also underline the need to keep internal textures and improve visual quality in the super-resolved images and movies. Additionally, the study describes adjustments made to the discriminator architecture, proposing the idea of a Realistic Average GAN to improve the comparison of visual and realistic quality in images. The suggested method is claimed to provide better textures and improve the frame generation rate compared to SRGAN.

Moreover, the research paper evaluates the suggested method’s performance using multiple quality evaluation measures such as VMAF, PSNR, SSIM, and MOS, while noting that training GANs consumes significant time and suggests a powerful graphics processor for quicker frame generation. In summary, the research paper focuses on enhancing super-resolution methods, notably via modifications in neural network architectures to generate higher-quality super-resolved images and videos.

More on this topic, the paper “Multi-Memory Convolutional Neural Network for Video Super Resolution” analyzed Multi-Memory CNN (MMCNN) for video super-resolution (SR), and emphasizes on improving the quality of high-resolution (HR) frames from low-resolution (LR) video sequences [5]. Unlike previous methods, which generally employ a direct connection and single-memory module inside convolutional neural networks (CNNs), MMCNN utilizes spatio-temporal complementary information across LR frames more efficiently.

The MMCNN design contains an optical flow network and an image-reconstruction network, connected in a way that cascades. It employs a sequence of residual blocks for feature extraction and reconstruction, emphasizing intra-frame spatial correlations. Notably, instead of a single-memory module, convolutional long short-term memory is integrated into the residual block, generating a multi-memory residual block. This approach gradually isolates and keeps inter-frame temporal correlations between consecutive LR frames.

Furthermore, extensive trials on multiple testing datasets with varied scaling factors illustrate the superiority of MMCNN over state-of-the-art approaches. MMCNN delivers greater Peak Signal-to-Noise Ratio (PSNR) and improved visual quality, exceeding the best available approach by up to 1 dB.

The basic goal of this work is to emphasize the relevance of super-resolution methods in computer vision, especially in the context of the rising demand for high-definition videos in formats like 4K and 8K. The paper highlights the significance of video super-resolution (Video SR), which seeks to rebuild high-resolution video frames from low-resolution input frames. Besides, the study examines the evolution of super-resolution methods, from interpolation-based techniques to learning-based approaches, with an emphasis on Convolutional Neural Network (CNN)-based methods that have attracted interest owing to their capacity to learn features from various image samples.

Moreover, it also analyzes the difference between single-image super-resolution (SISR) and video super-resolution (Video SR), emphasizing the necessity to utilize temporal correlations between low-resolution frames in Video SR. Traditional Video SR methods are given, however, they often require computational costs that are higher and struggle with large scaling factors and motions. The research paper covers current CNN-based Video SR techniques and their limitations, notably in terms of computational complexity and the usage of inter-frame temporal information. Finally, the core concept of the paper's suggested method, the Multi-Memory Convolutional Neural Network (MMCNN), is described. The MMCNN contains an optical flow network and an image-reconstruction network, intending to use both spatial and temporal information for enhanced video super-resolution. The paper's contributions and the arrangement of the future parts are also detailed.

In summary, it sets the stage for the paper by emphasizing the importance of Video SR in the era of high-definition displays, highlighting the limitations of existing methods, and introducing the innovative MMCNN framework as a solution to enhance both the quality and efficiency of video super-resolution.

Chapter 3

Work Plan

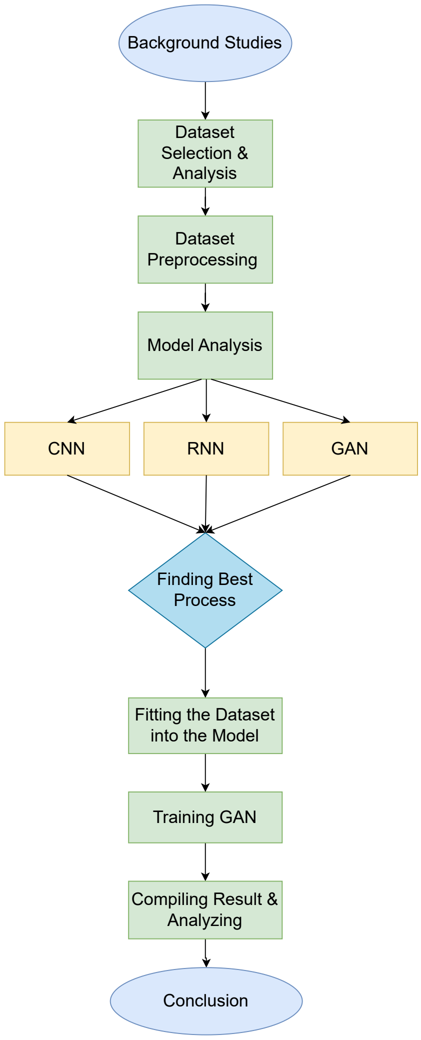

Figure 3.1: Workplan in a flowchart diagram

The research work plan 3.1 starts with a rigorous selection of an appropriate dataset including high-quality photos and videos. This basic step is critical to guarantee that our models get the greatest training data available. After we have obtained our dataset, we will continue to preprocess it and improve it for usage as inputs in our training process.

Our major emphasis focuses on building and improving models for the process of upscaling single photos and videos to attain super-resolution. This key step will include training our algorithms using the carefully produced datasets, allowing them to learn and increase their capacity to boost picture and video resolution.

After completing the training process, we will start the key step of upscaling testing. Here, we will methodically tweak our code and incorporate many layers to refine our models, seeking to obtain the greatest potential performance. This iterative refining process will enable us to examine alternative techniques and combinations to maximize our models.

Ultimately, our study will conclude with a complete examination of the outcomes gained from each model. We will examine how they perform based on multiple metrics and benchmarks to establish their usefulness in super-resolution jobs. Through this comprehensive review, we want to give significant insights and ideas for enhancing the upscaling of single photos and videos.

In summary, our study tries to contribute to the improvement of super-resolution approaches by employing carefully selected datasets, robust model training, and iterative optimization. By thoroughly studying and comparing the outputs of our models, we seek to give a superior strategy for upscaling single photos and videos, thereby boosting the quality of visual information.

Chapter 4

Conclusion

In the world of photographs and Videos, where clarity and detail matter, our research on deep learning algorithms for super-resolution has uncovered a route to dramatic visual improvement. Through the synergy of CNNs, RNNs, and GANs, we've observed the potential of technology to breathe new life into the ordinary.

After preprocessing the selected dataset for the Super Resolution the CNN, RNN GAN models were implemented. Analyzing the models for Single Image and Video Super-Resolution we found GAN is dominating the CNN RNN in terms of overall upscaling.

From improving single photos to enhancing video quality, this strategy has broad implications across numerous areas, which is visible in our everyday lives. Our venture into upscaling has proved that, with the correct tools or algorithms, we can expand our digital experiences and enable older devices to go beyond their boundaries.

As we come to the end of this paper, we embrace the combination of artificial intelligence and creativity, opening possibilities for a future where every pixel tells a clearer, more colorful story. With deep learning as our guide, we encourage you to visualize a future where visual perfection knows no limitations.

Bibliography

[1] C. E. Duchon, "Lanczos filtering in one and two dimensions," *Journal of Applied Meteorology and Climatology*, vol. 18, no. 8, pp. 1016–1022, 1979.

[2] R. Keys, "Cubic convolution interpolation for digital image processing," *IEEE transactions on acoustics, speech, and signal processing*, vol. 29, no. 6, pp. 1153–1160, 1981.

[3] Y.-N. Liu, Y.-C. Lin, Y.-L. Huang, and S.-Y. Chien, "Algorithm and architecture design of high-quality video upscaling using database-free texture synthesis," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 24, no. 7, pp. 1221–1234, 2014.

[4] W. Shi, J. Caballero, F. Huszár, *et al.*, "Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 1874–1883.

[5] Z. Wang, P. Yi, K. Jiang, *et al.*, "Multi-memory convolutional neural network for video super-resolution," *IEEE Transactions on Image Processing*, vol. 28, no. 5, pp. 2530–2544, 2018.

[6] U. Cisco, "Cisco annual internet report (2018–2023) white paper," *Cisco: San Jose, CA, USA*, vol. 10, no. 1, pp. 1–35, 2020.

[7] E. Chan, "Pack more megapixels into your photos with adobe super resolution," *Adobe Blog*, 2021.

[8] Z. Li, F. Liu, W. Yang, S. Peng, and J. Zhou, "A survey of convolutional neural networks: Analysis, applications, and prospects," *IEEE transactions on neural networks and learning systems*, 2021.

[9] M.-Y. Liu, X. Huang, J. Yu, T.-C. Wang, and A. Mallya, "Generative adversarial networks for image and video synthesis: Algorithms and applications," *Proceedings of the IEEE*, vol. 109, no. 5, pp. 839–862, 2021.

[10] F. Lundkvist, *Deep upscaling for video streaming: A case evaluation at svt.* 2021.

[11] N. Tovar, *Real-Time Image Upscaling with Commonly Available Resources.* California State University, Long Beach, 2021.

[12] A. Rakvi, J. Shah, P. Singh, and S. Malik, "Super resolution of videos using e-gan," *Available at SSRN 4012374*, 2022.

[13] X. Xiang, Y. Tian, V. Rengarajan, L. D. Young, B. Zhu, and R. Ranjan, "Learning spatio-temporal downsampling for effective video upscaling," in *European Conference on Computer Vision*, Springer, 2022, pp. 162–181.

[14] *FFmpeg a complete, cross-platform solution to record, convert and stream audio and video.* https://ffmpeg.org/, Accessed: 2023-09-18.

[15] G. F. Freedman, "R. 2011. image and video upscaling from local self-examples," *ACM Transaction on graphics*,

[16] *NVDIA dlss 3*, https://www.nvidia.com/en-us/geforce/technologies/dlss/?nvid=nv-int-gfhm-55050, Accessed: 2023-09-18.