IMAGE SUPER RESOLUTION USING DEEP LEARNING

A THESIS
Submitted in partial fulfillment of the requirements for the award of degree

of

DOCTOR OF PHILOSOPHY

in

ELECTRONICS AND COMMUNICATION ENGINEERING

by

Ajay Sharma
Scholar Number: 193114013

Under the guidance of
Dr. Bhavana P. Shrivastava

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY
BHOPAL-462003 (INDIA)
SEPTEMBER- 2023

©MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY,

BHOPAL– 462003

ALL RIGHTS RESERVED

MAULANA AZAD NATIONAL INSTITUTE OF TECHNOLOGY, BHOPAL

CANDIDATE'S DECLARATION

I hereby certify that the work which is being presented in the thesis entitled "**Image Super-Resolution Using Deep Learning"** in partial fulfillment of the requirements for the award of the degree of Doctor of Philosophy and submitted to the Department of Electronics and Communication of the Maulana Azad National Institute of Technology, Bhopal is an authentic record of my work carried out during the period from July 2019 to September 2023 under the supervision of **Dr. Bhavana P. Shrivastava**, Assistant Professor, Department of Electronics and Communication, Maulana Azad National Institute of Technology, Bhopal.

The matter presented in this thesis has not been submitted by me for the award of any other degree of this or any other institution.

(Ajay Sharma**)**

This is to certify that the above statement made by the candidate is correct to the best of my knowledge.

Dr. Bhavana P. Shrivastava

Date: (Supervisor)

The Ph.D. Viva-Voice Examination of Mr. Ajay Sharma, Research Scholar, has been held on ……

Signature of Supervisors Signature of External Examiner

ACKNOWLEDGEMENTS

I would like to express my deepest gratitude to all who have been associated with this research entitled “Image Super-Resolution using Deep Learning”.

I am deeply indebted to my supervisors Dr. Bhavana P. Shrivastava for their support, motivation, and guidance in this entire journey. I would like to acknowledge the contribution of Dr. Bhavana P. Shrivastava in my research Chapters/thesis drafting. She did all the strenuous corrections patiently and motivated me to the further progress. Her vast research experience and overall insights in the area of machine learning have made this an inspiring experience for me. Dr. Bhavana P. Shrivastava has always helped me in providing resources, and in organizing my research progress presentations. She has been a great role model for women in academia. I am thankful to my SRPC committee members Dr. Laxmi Kumre and Dr. Saritha Khetwat for their valuable feedback. This journey would not have been possible without their sincere efforts.

The TA duties assigned each semester have also been a great learning experience in my Ph.D. program. The support and helping attitude of my M-tech and Ph.D. colleagues is commendable. I would like to thank Praveen Tyagi, Rajneesh Patel, and Rahul Kumar who were part of my TA duties group. I also wish to express my sincere gratitude to all lab technicians and non- teaching staff. Mrs. Vasanta Kumari and Mr. Pradeep guided me with all the fellowships regarding Chapter work, extension, registration formalities, etc.

Words cannot express my gratitude towards my parents and family members whose unconditional support backed me to soar high in life. My Wife Namrata Joshi always supported me in my journey to Ph.D. In the program, My Father guided me with his experience and instilled confidence to work even harder, whenever Chapter rejections took a toll on me. Last, but not least I would like to express my gratitude to all those who either directly or indirectly helped in the successful completion of my thesis work.

(Ajay Sharma)

ABSTRACT

Image Super-Resolution is the process of enhancing Low-Resolution (LR) images to High Resolution (HR) images, which are widely used in applications such as surveillance, medical diagnosis, and so on. With an increasing number of imaging applications, super-resolution is becoming more efficient and practical. As a result, the goal of this work is to provide a comprehensive study of image super-resolution using a deep convolution neural network. A brief discussion of frame works and different network designs, such as Linear networks, Residual networks, Recursive networks, Gan, and Attention networks, for image super-resolution using deep learning is presented, as well as a comparison of their performances and complexity. This work also highlights the use of various upscaling techniques and loss functions, as well as parametric evaluation, which assists new researchers in designing a new framework. A brief analytical study on standard datasets such as Set5, Set14, BSD100, and Urban100 is also presented, which aids in the gathering of data for less complex network design. Regardless of recent progress, this work identified some flaws in existing models and proposed future research directions to address open issues related to image super-resolution. Because of the growing demand for surveillance camera-based applications, the very low resolution (VLR) issue arises in many face application systems. Face recognition algorithms are unable to work satisfactorily on the VLR face image. While face super-resolution (SR) techniques can be used to improve image resolution, these VLR face images do not respond well to existing learning-based face SR techniques. To address this issue, this research proposes a novel method for learning the link between the high-resolution image space and the VLR image space for face SR. This work aims to address the problem of face image super-resolution and solutions for resolving facial image issues. The main issue with face image super-resolution during reconstruction is the fusion of multiple types of features, which can cause image noise, blind spot generation, low perceptual quality, and checkboard issues. Existing models are intended to improve the perceptual quality of face super-resolution, but they fail to produce a better image perceptual quality due to loss during the reconstruction stage. As a result, we propose a novel progressive face hallucination super-resolution (FHSR) model with a loss-aware upscaling network layer in this work. The upscaling layer incorporates correlation filters and employs a combined loss function. The model is built as a cascade of progressive upscaling modules with dense skip connection layers. To minimize

reconstruction losses, loss functions from each upscaling module are evaluated, cascaded, and fed forward to the next layer. The images were upscaled using three cascading modules that magnified them 2x, 4x, and 8x. This research also presented an ablation study that demonstrated that the designed FHSR model outperformed baseline models. This architecture solves the problem of extracting more features with minimal losses while also constructing high magnification at the reconstruction end. The comparative study demonstrates improved performance over state-of-the-art models. Deep learning architecture is also used for the super-resolution of generalized images and medical images. Generalized image super-resolution aids in information retrieval in computer vision tasks, whereas medical image super-resolution aids in diagnosis. Image resolution, imaging equipment, and imaging parameters all degrade performance in medical image diagnosis. Deep learning is currently gaining popularity among researchers due to its ability to maintain perceptual quality after reconstruction. As a result of the benefits of deep learning, this work proposes a novel model known as the correlation filter interleaved progressive convolution network (CFIPC). Dilated convolution interleaved with a correlation filter in this work expands the receptive field without pixel information loss. The analysis was performed on DRIVE, CHASEB1, MRI, ultrasound, and histopathological dataset medical datasets at 2x/4x/8x/16x upscaling factors, with the histopathological dataset achieving the highest PSNR/SSIM value of 50.54/0.9986 at a 2x upscaling factor. The run time Efficiency of of (CFIPC) is evaluated as 96.77%, the Run time efficiency of Multilevel Progressive Recursive Dilated Networks with Correlation Filter (MPRDNCF) is 97.50 % and for FHSR Network run time efficiency is 97.46%.

In this work, various deep learning architectures are designed for image multi-scaling using a progressively dilated convolution interleaved correlation filter, which improves image quality at the reconstruction end and is very useful in many applications such as medical images, generalized images, facial images, which are also important in computer vision tasks.

TABLE OF CONTENTS

CANDIDATE'S DECLARATION ..... **iii**

ACKNOWLEDGEMENTS ..... **v**

ABSTRACT ..... **vii**

Table of Contents ..... **ix**

List of Figure ..... **xiii**

List of Table ..... **xv**

ABBREVIATIONS ..... **xvii**

Chapter 1: Introduction **1**

1.1 Introduction 1
1.2 Learning Strategy 2
1.2.1 Nearest Neighbour Interpolation 3
1.2.2 Bilinear Interpolation 3
1.2.3 Bi-Cubic Interpolation 4
1.2.4 Deep Learning Upscaling Frameworks. 4
1.2.5 Learning-based Up-sampling 7
1.2.6 Meta-up Scaling Module 7
1.3 Convolutional Neural Network 8
1.3.1 Training of CNNs 10
1.3.2 Advantages of Convolutional Neural Networks 11
1.4 Performance Metric Evaluation 12
1.4.1 Peak Signal-to-Noise Ratio (PSNR) 12
1.4.2 Contrast to Noise Ratio (CNR) 13
1.4.3 Mean Square Error (MSE) 13
1.4.4 Root Mean Square Error (RMSE) 14
1.4.5 Structural Similarity Index Measurement (SSIM) 14
1.5 Loss Function 15
1.5.1 Adversarial loss (AL) 16
1.5.2 Perceptual loss 16
1.5.3 Super-Resolution Loss 17
1.5.4 Pixel Loss 18
1.5.5 Structural Similarity Index Measurement (SSIM) 19
1.5.6 L1 loss function 20

1.5.7 L2 loss function 20
1.5.8 Run Time Effieciency in Image Super-Resolution 20
1.6 Other Improvements in Super-Resolution. 21
1.6.1 Data augmentation –..... 21
1.6.2 Multi-task Learning 22
1.6.3 Context wise Network fusion 23
1.6.4 Self-ensemble 23
1.7 Datasets for Super - Resolution 24
1.8 Use of Correlation Filter in Image Super Resolution 25
1.9 Organization of Thesis 26
Chapter 2: Recent Work 29
2.1 Introduction 29
2.2 Network Designs 29
2.2.1 Linear Networks 29
2.2.2 Residual Network 34
2.2.3 Recursive Network 37
2.2.4 Attention Network 39
2.2.5 Generative Adversarial Network 41
2.3 Application Areas of Image Super Resolution Models 47
2.4 Current Limitations and Challenges in Image Super Resolutions 51
2.5 Problem Formulation 51
2.6 Research Objective 52
2.7 Summary 52
Chapter 3: Medical Image Super-Resolution Using Deep Learning 55
3.1 Medical Image Super-Resolution using Correlation Filter Interleaved Progressive Convolution Network (CFIPC) 55
3.1.1 Introduction 55
3.1.2 Methodology 57
3.1.3 Experimental Results and Discussion: 59
3.1.4 Performance Evaluation: 60
3.1.5 Run Time Efficiency 62
3.2 Complex Wavelet Transform with Progressive Network for Medical Imaging Super Resolution 62
3.2.1 Introduction 62
3.2.2 Related Work 64
3.2.3 Methodology 65
3.2.4 Experimental Results and Discussion 67

3.2.5 Comparison with State-of-the-Art 68
3.2.6 Conclusion 71
3.3 Summary 72
Chapter 4: Image Super-Resolution Using Progressive Network Interleaved Correlation Filter. 73
4.1 Introduction 73
4.2 Related Work 74
4.3 Problem Identification 79
4.4 Methodology 80
4.4.1 Overview of MPRDNCF Network 80
4.4.2 Architecture of MPRDNCF 82
4.4.3 Training Details 83
4.5 Results and Discussions 85
4.5.1 Performance Metrics 85
4.5.2. Structural Similarity Index Measurement (SSIM) 85
4.5.3 Implementation Details 86
4.5.4 Dilated Convolution Analysis 86
4.6 Performance Analysis 87
4.7 Comparison with State-of-the-Arts 88
4.8 Run Time Efficiency 90
4.9 Summary 90
Chapter 5: Facial Image Super-Resolution Using Progressive Network Interleaved Correlation Filter. 93
5.1 Introduction 93
5.1.1 Motivation 96
5.1.2 Key Contribution and Chapter Organizations 97
5.2 Related Work 97
5.3 Methodology 99
5.3.1 Novelty of Work 99
5.3.2 Overview of Progressive FHSR Network 99
5.4 Experiments 105
5.4.1 Dataset Description 105
5.4.2 Ablation Study 106
5.4.3 Result Analysis 110
5.4.4 Comparisons with State of Art techniques. 110
5.5 Run Time Efficiency 114

5.6 Summary 114
Chapter 6. Conclusion and Future Scope 115
Reference 119
List of Publication 135
Authors Contribution 137

LIST OF FIGURE

Figure 1.1 Pre-Sampling Framework **5**
Figure 1.2 Post-Sampling Framework **6**
Figure 1.3 Progressive Up sample Framework **7**
Figure 1.4 CNN Architecture **9**
Figure 2.1 (a) Frameworks for SR, (b) Networks for SR, (c) Applications for SR Bibliometric Analysis for Research Contribution on Image Super Resolution **29**
Figure 2.2 Residual Framework **34**
Figure 2.3 Recursive Framework **37**
Figure 2.4 Attention Framework **40**
Figure 2.5 Generative Adversarial Network **42**
Figure 3.1 Architecture of CFIPC **58**
Figure 3.2 Examples of test results on different medical image datasets. **62**
Figure 3.3 Architecture of PDCN **67**
Figure 3.4 Result using K-10 cross-validation **71**
Figure 3.5 Examples of test results on different medical image datasets. **71**
Figure 4.1 Basic Architecture of Proposed Model **80**
Figure 4.2 Difference between standard and dilation convolution **81**
Figure 4.3 Basic Architecture of dilated_Conv+LReLU Layer **81**
Figure 4.4 Architecture of MPRDNCF **82**
Figure 4.4 Correlation between LR and HR images with varying epochs **87**
Figure 4.5 Super Resolution Results on Set5, Set14, BSD100 and Urban100 images with scale factor 4 ..... **90**
Figure 5.1 Example of VLR in Face Super-Resolution **94**
Figure 5.2 Architecture of FHSR **103**
Figure 5.3 Architecture of UNCF Layer **103**
Figure 5.4 Ablation Study on Baseline Models **109**
Figure 5.5 Robustness Visualization of Progressive FHSR (UNCF). The first row represents the generation of the HR image from the 24x24 LR face image. The bottom row represents the original image. **110**
Figure 5.6 Reconstruction Comparative State-of-Art Performance **112**

LIST OF TABLE

Table 1.1 List of public image datasets for super-resolution benchmarks. 24
Table 2.1 Performance Evaluation on Benchmark Dataset Image SR of different CNN models for 2x magnification..... 47
Table 3.1 Parameter Evaluation for Proposed CFIPC..... 60
Table 3.2 Comparative state-of-art on DRIVE dataset 61
Table 3.3 Comparative state-of-art on Histopathological dataset 61
Table 3.4 Comparative state-of-art on Ultrasound dataset..... 61
Table 3.5 Evaluation of Proposed PDCN Parameters..... 69
Table 3.6 On the DRIVE dataset, state-of-the-art 69
Table 3.7 Histopathological dataset comparative state-of-the-art 69
Table 3.8 Ultrasound dataset comparative state-of-the-art..... 70
Table 3.9 Ablation Study of Proposed Network with Different Loss Functions..... 70
Table 4.1 Comparative Analysis of Different CNN Architectures used for Image Super Resolution..... 84
Table 4.2 PSNR and SSIM Evaluation of MPRDNCF 88
Table 4.3 Comparative State-of-arts..... 89
Table 5.1 Comparative Analysis of Different CNN Architectures used for Image Super-Resolution..... 98
Table 5.2 Description Of network architecture of UN Modules 104
Table 5.3 Model Characteristics for Ablation Study 107
Table 5.4 Ablation Study of Baseline Models..... 108
Table 5.5 Ablation Study of Model IV with Different Loss Functions 108
Table 5.6 Ablation Study of Model IV with Rotation Degree..... 109
Table 5.7 PSNR Evaluation on LFW, HELEN, and CELEB A Datasets 111

ABBREVIATIONS

VDSR	Very Dep Super Resolution.
SRCNN	Super Resolution Convolution Neural Network
EDSR	Enhance Deep Super Resolution
MDSR	Multi-Scale Deep Super Resolution
SRGAN	Super Resolution Generative Adversarial Network
CARN	Cascading Residual Network
BTSRN	Balanced Two-Stage Residual Networks
GPU	Graphical Processing Unit
RDB	Residual Dense Block
RDN	Residual Dense Network
MLP	Multi-Layer Perceptron
PCA	Principal Component Analysis
PreLU	Parametric Rectified Linear Unit
LReLU	Leaky Rectified Linear Unit
PSNR	Peak Signal-to-Noise Ratio
SICNN	Super Identity Convolutional Neural Networks
SSIM	Structural Similarity Index Measure
CFIPC	Correlation Filter Interleaved Progressive Convolution
MPRDNCF	Multilevel Progressive Recursive Dilated Networks with Correlation Filter
IRCNN	Image Restoration Convolutional Neural Network
ESPCN	Efficient sub-pixel convolutional neural network
DnCNN	Denoising Convolutional Neural Network
C-SRIP	Cascaded Super Resolution Identity Prior

CHAPTER 1: INTRODUCTION

1.1 Introduction

The digital representation of visual information stored in a digital format is referred to as a digital image. Images can exist in either two-dimensional or three-dimensional forms. The field of image processing and computer vision is continuously evolving due to the widespread utilization of digital image data across various applications. In recent years, there has been a growing demand for high-quality image magnification in computer vision, necessitating the enhancement of image quality during the magnification process. Super-Resolution (SR) is a technique employed to convert low-resolution (LR) images into high-resolution (HR) images [1]. For the task of single image SR (SISR), an HR image (X2) is typically generated from an LR image (X1) based on a model defined as follows: $X2 = X1 * k \downarrow s + n$, where k represents a low-pass filter, $*$ denotes the convolution operator, $\downarrow s$ represents the down-sampling operator with a down-sampling factor of s, and n represents the random noise parameter. In this process, the HR image (X2) is produced by convolving the LR image (X1) with a low-pass filter, followed by down sampling. These SR techniques push the boundaries of imaging sensors. There are two types of SR variants: single framing and multiple framing, with the latter being more beneficial. SR algorithms are categorized into two domains: frequency domain [2] and space domain [3]. The merging of both domains results in an enhanced-resolution image from the LR input. Frequency-domain SR has limited applications due to increased noise generation during image reconstruction and its dependence on HR images. It finds utility in specific domains such as astronomy, microscopy, and radar [4]. The fundamental concept behind super-resolution is the fusion of a series of LR images to reconstruct an HR image. The information is expanded through aliasing, whereby LR images are sub-sampled and precisely shifted by sub-pixel units. By shifting the pixels of LR images, each image retains similar information, enabling the reconstruction of the HR image using the previous pixels. Therefore, no new information is introduced during the reconstruction process. However, sometimes the reconstructed image fails to provide a genuinely high-resolution image due to the aliasing effect [5]. Previous super-resolution techniques have gained significant popularity due to their wide range of applications, including medical imaging, surveillance, and security [6]-[8]. Super-resolution (SR) presents a valuable opportunity to enhance various applications. SR methods are highly valuable in meeting

the increasing demands of imaging applications that require precise spatial resolution. While the SR technique has been effectively applied to various machine learning and deep learning paradigms, further research is needed to enhance its performance. Convolutional Neural Network (CNN) has emerged as one of the most prominent SR techniques and has demonstrated superior reconstruction capabilities for super-resolution images. Recent Deep-Learning-based approaches, particularly those utilizing deeply and completely convolutional networks, have shown great performance in the challenge of enhancing low-resolution (LR) images to high-resolution (HR) images [9]. The utilization of CNN layers to improve performance has become a popular practice. However, the majority of currently used SR models rely on powerful computers, limiting their practical application. Additionally, many algorithms overlook the exploration of intermediary features that play a crucial role in the final image recovery. In contrast, the CNN structure enables faster computations with minimal information loss and facilitates the direct processing of original images by reducing the size of the output from preceding layers. Convolutional neural networks also extract more comprehensive features [10]. Furthermore, each CNN's layers and filters are optimized to significantly reduce computation costs. As a result, CNN not only achieves cutting-edge performance but also delivers quicker and more effective computations.

Therefore, this chapter aims to introduce prior studies on SR techniques, specifically Sparse-based and Deep Learning approaches, which provide valuable insights into the challenges and open issues in SR, such as perceptual quality and loss functions. These studies contribute to improving visionary tasks and perceptual quality. Generally, the SR technique is employed to address challenging tasks such as Viewpoint Variation, Illumination, Deformation, Occlusion, Background Clutter, and Interclass Variation [6], [7]. A literature survey of prior SR techniques has been conducted for these challenges, including Edge-based feature extraction, Patch-based approaches, Sparse-based approaches, and Prediction-based methods.

1.2 Learning Strategy

Interpolation methods play a significant role in resizing digital images and find wide applications in various fields. The commonly employed interpolation techniques include a nearest neighbor, bilinear interpolation, and bicubic interpolation. These methods are

popular due to their ease of implementation and their integration into CNN-based super-resolution models.

1.2.1 Nearest Neighbour Interpolation

Nearest neighbor interpolation involves selecting the pixel value from the original image that is nearest to the desired pixel location in the resized image. This method is straightforward to implement but may result in blocky or jagged edges in the resized image. The nearest neighbor interpolation algorithm is a simple and intuitive method used to determine the pixel value for each position during interpolation. It selects the nearest pixel value from the original image without considering the surrounding pixels. While this technique is computationally efficient, it often results in blocky and low-quality interpolated images. The blocky artifacts occur because the nearest neighbor interpolation does not take into account the smooth transition of pixel values in the original image. As a result, the resized image may appear distorted and lack the fine details present in the original image. This method is primarily used in situations where computational efficiency is prioritized over image quality, such as in real-time applications or situations with limited computing resources. However, for applications that require higher-quality interpolation, such as image resizing for display purposes or image enhancement, more advanced techniques like bilinear or bicubic interpolation are commonly preferred. These methods consider the surrounding pixels and employ interpolation formulas to calculate the pixel values, resulting in smoother and visually more appealing results [11].

1.2.2 Bilinear Interpolation

Bilinear interpolation is an interpolation technique that calculates the new pixel value by taking the weighted average of the four nearest neighbour pixels surrounding the desired location. It provides smoother and more visually appealing results compared to the nearest neighbour interpolation method, while still being relatively simple to implement. Bilinear interpolation performs linear interpolation along one axis of the image at a time. It calculates the pixel values based on the intensity values of the four nearest neighbor pixels in a 2x2 receptive field. By taking the weighted average of these pixel values, bilinear interpolation creates a smoother transition between adjacent pixels, resulting in a more continuous and visually pleasing image. Compared to the nearest neighbor interpolation, bilinear interpolation offers better image quality by considering the local neighborhood of pixels. It reduces the blocky artifacts and provides a more accurate

representation of the original image. Additionally, bilinear interpolation is faster than more complex interpolation techniques like bicubic interpolation, making it a popular choice for applications where efficiency is important. In summary, bilinear interpolation strikes a balance between simplicity and improved image quality. It provides smoother results compared to the nearest neighbor interpolation while maintaining computational efficiency [11].

1.2.3 Bi-Cubic Interpolation

Bicubic interpolation is an interpolation technique that considers a larger neighborhood of pixels, typically 16 pixels, to calculate the new pixel value. It provides even smoother results compared to bilinear interpolation, but it requires more computational resources. In bicubic interpolation, the algorithm uses a weighted average of the 16 surrounding pixels within a 4x4 receptive field to determine the interpolated pixel value. By considering a larger neighborhood, bicubic interpolation can capture more information and produce smoother transitions between pixels. This results in improved image quality, particularly in scenarios such as super-resolution, where higher levels of detail are desired. Bicubic interpolation is widely used in image super-resolution due to its accuracy and ability to reduce aliasing artifacts. It is an anti-aliasing technique that helps mitigate issues related to upscaling images. However, it is important to note that bicubic interpolation is computationally more demanding compared to simpler interpolation methods like bilinear interpolation. While bicubic interpolation offers superior results in terms of smoothness and reduced artifacts, it does come with some drawbacks. These include increased computational complexity, potential blurring effects, and the possibility of amplifying noise present in the original image. To address these challenges and improve performance, alternative techniques such as sampling layers have been developed and integrated into super-resolution models. These techniques aim to balance computational efficiency, image quality, and noise reduction in the context of image upscaling [11].

1.2.4 Deep Learning Upscaling Frameworks.

1.2.4.1 Pre-Up-Sampling Framework.

The process of learning features, also known as mapping, plays a crucial role in image super-resolution. It involves transforming low-dimensional features into high-dimensional features. This challenge is addressed by a framework illustrated in Figure 1.1 (a). Initially, the framework employs traditional techniques such as bicubic, nearest

neighbors, or bilinear transformations to convert the image from low to high resolution. Subsequently, the image is refined straightforwardly using a CNN architecture to generate high-quality images [12]. This framework was first introduced by C. Dong et al. [10], who proposed the use of end-to-end mapping, which has become a widely popular method in image super-resolution. However, this framework does have some side effects, including blurring and noise amplification. These issues arise because most of the operations are performed in high dimensions, increasing in time and space complexity compared to other techniques.

The blurring effect can occur due to the smoothing nature of the CNN operations, which may inadvertently soften the image details. Noise amplification can arise from the fact that the framework may amplify the noise present in the low-resolution input during the process of generating the high-resolution image. While the framework proposed by C. Dong et al. [10] has gained popularity due to its end-to-end mapping approach, it is important to consider its limitations, such as the potential for blurring and noise amplification, as well as the higher computational cost compared to other techniques. Researchers continue to explore and develop methods that address these challenges to improve the overall performance and quality of image super-resolution.

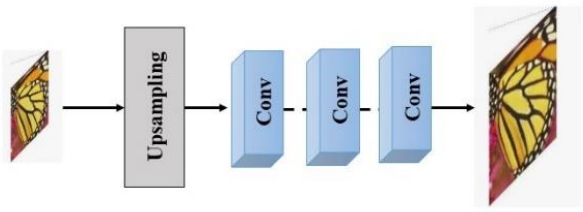


Figure 1.1 Pre-Sampling Framework (referred from [10])

1.2.4.2 Post-Up Sampling Framework.

To address the complexity of network computations, researchers have explored approaches that involve operating in a low-dimensional space. These proposed networks follow a two-step process to upscale the image using a CNN architecture with end-to-end mapping in the low-resolution space. Then, at the end of the network, they apply traditional methods such as bicubic or nearest neighbors' interpolation to further upscale the image. This approach aims to reduce the computational cost of the framework by performing most of the computations in the initial stage. Figure 1.2 illustrates the overall framework. By focusing on the initial stage, the computation time of the framework can be decreased, resulting in improved efficiency. Additionally, this approach can reduce the

overall size of the network architecture, making it more practical and easier to use. As a result, this framework has gained popularity [13]. These models often differ in their learning strategies, the utilization of learnable up-sampling layers, and the specific CNN architectures employed. Researchers explore various techniques and variations to improve the overall performance and quality of the super-resolution models. The goal is to strike a balance between computational efficiency and maintaining high-quality image upscaling.

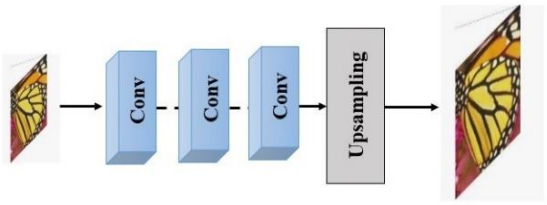


Figure 1.2 Post Sampling Framework (referred from [13])

1.2.4.3 Progressive Sampling Framework.

While the post-sampling framework reduces the complexity and cost of the network, it can become unwieldy when dealing with multiple magnifications. In such cases, a progressive approach has been developed to address this challenge. This approach allows for the use of a single network design in a cascaded manner to handle multiple magnifications, such as 4x and 8x, in a single step. Figure 1.3 illustrates the architecture of this progressive approach. In this progressive framework, the image is initially magnified using a CNN architecture and then upscaled further using various learning strategies. By cascading these steps, the network can effectively handle multiple magnifications within a single pass. This approach was first introduced by models like LapSRN [14] and Csrip [1]. However, these models have their own set of challenges. Designing complex architectures to handle multiple magnifications can be demanding, requiring careful consideration of network structure and parameters. Additionally, training stability can be an issue in these models, as the progressive approach may introduce challenges in convergence and optimization. To overcome these challenges, further guidance and the development of more robust training approaches are needed. Researchers continue to explore and propose new techniques to enhance the performance and stability of progressive super-resolution models for multiple magnifications.

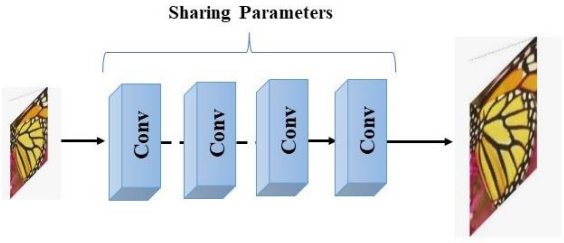


Figure 1.3 Progressive Up sample Framework (referred from [15])

1.2.5 Learning-based Up-sampling

To address the limitations of interpolation techniques, sampling layers have been introduced in image super-resolution. These layers include transposed convolution layers [12], sub-pixel convolution layers, pixel TCL, and Meta upscaling modules. Transposed convolution, also known as the deconvolution layer, is used for up sampling images using learned parameters. It expands the image by introducing zero values to achieve the desired output feature map size. However, a major drawback of transposed convolution in image super-resolution is the occurrence of the "checkerboard effect." This effect manifests as uneven overlapping in certain areas of the image, leading to undesirable visual artifacts. To mitigate this issue, it is recommended to use a kernel size that is divisible by the stride. The subpixel layer helps overcome the problem of uneven overlapping introduced by transposed convolution [16]. It performs convolution while generating an output with s^2 times the number of channels, where s is the scaling factor. This layer has a larger receptive field and can generate realistic details. However, it still has limitations in producing smooth output in adjacent blocky regions. Other sampling layers, such as pixel TCL (Transposed Convolutional Layer) and Meta upscaling modules, have also been proposed to address specific challenges in image super-resolution. These layers aim to improve the quality and smoothness of the upscaled images by incorporating different strategies and techniques. The introduction of sampling layers in super-resolution models offers alternative approaches to traditional interpolation techniques and contributes to enhancing the quality and realism of the upscaled images. Researchers continue to explore and develop new sampling layers and architectures to further improve the performance and address the remaining challenges in image super-resolution.

1.2.6 Meta-up Scaling Module

To address the issue of training required at each scaling factor in super-resolution models, Meta-learning techniques have been introduced. Meta-learning allows for the

simultaneous training of multiple factors or parameters, including the convolution weights of the input image. By utilizing Meta-learning, the model can predict the convolution weights during inference for each targeted pixel. This approach offers several advantages, including reduced execution time and the ability to efficiently predict weights. However, it may be less efficient for larger magnifications, where the scaling factor is significantly higher. The Meta-learning module provides a more efficient way to train super-resolution models by learning multiple factors simultaneously and predicting weights during inference. This helps to streamline the training process and improve the overall efficiency of the model. Researchers continue to explore and refine Meta-learning techniques to enhance their effectiveness in various super-resolution applications.

1.3 Convolutional Neural Network

The convolutional neural network (CNN) architecture is designed specifically for image processing tasks, addressing the problem of having too many connections in traditional neural network architectures. The main features of CNNs are as follows:

Convolutional Layer (CL): The convolutional layer applies convolutional operations to extract local features from the input image. It uses filters to convolve over the input and produce feature maps, which capture different aspects of the image.

Rectified Linear Unit (ReLU): ReLU is commonly used as the activation function in CNN architectures. It sets negative coefficients to zero while preserving positive coefficients, allowing for an efficient representation of local features. The ReLU function is defined as $f(x) = \max(0, x)$.

Pooling Layer (PL): The pooling layer is used to reduce the dimensionality of the feature maps from the previous layer. It selects the most prominent features within local regions and subsamples the image. Max pooling is a commonly used technique, where the maximum value within each pooling region is retained.

Fully Connected Layer (FCL): The fully connected layer is typically placed at the end of the CNN architecture. It takes the features extracted by previous layers and performs classification or regression tasks. In image classification, the output is an array of probabilities representing different categories.

Stride Convolution: Stride convolution is used to reduce the size of the feature maps generated by convolutional operations. It allows for down sampling while maintaining

the desired output size. Utilizing filters, can effectively reduce artifacts and improve processing time.

Neurons Elimination: In CNNs, some neurons that are not contributing significantly to the forward propagation or backpropagation process can be eliminated. This reduces the complexity of the network and the computational load.

Pooling Layer: The pooling layer is used to reduce the dimensionality of the feature maps and make them more suitable for the subsequent layers. The most common approach is to use max pooling, which selects the maximum value within each pooling region.

Fully-Connected Layer: The fully-connected layer is typically used as the last stage of the CNN architecture and is responsible for classifying the images. It outputs an array of probabilities for each category, allowing for image classification.

Overall, CNNs are designed to efficiently process images by leveraging the convolutional layer, ReLU activation, pooling layer, and fully connected layer. These architectural components help extract meaningful features from images and enable accurate image classification and other image-processing tasks. A Hierarchy of Convolutional architecture is shown in Figure 1.4, which shows how we connect the different layers to generate an output image. All the layers are connected to extract the feature of the image.

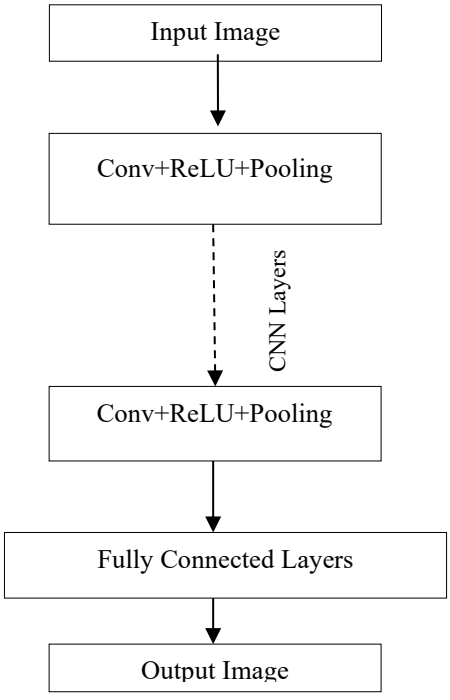


Figure 1.4 Basic CNN architecture

1.3.1 Training of CNNs

The training of a convolutional neural network (CNN) involves optimizing the network's parameters (such as weights and biases) to minimize the difference between the predicted output and the desired output. This is typically done through supervised learning, where a training dataset consisting of input images and their corresponding desired output images is used. The process of training a CNN involves defining a loss function, also known as an objective function, which quantifies the error between the predicted output and the desired output. The choice of loss function depends on the specific task at hand, such as image restoration, image classification, or image segmentation. Commonly used loss functions include mean squared error (MSE), mean absolute error (MAE), or cross-entropy loss. Different algorithms can be employed to optimize the loss function and update the network's parameters during training. Some commonly used training algorithms for CNNs include:

First-order hybrid training methods: These methods combine the use of gradient descent with other optimization techniques, such as momentum, to update the network parameters.

Conjugate gradient methods: These methods use conjugate gradient descent to find the minimum of the loss function by iteratively searching in conjugate directions.

Quasi-Newton methods: These methods approximate the Hessian matrix (second-order derivatives of the loss function) to update the network parameters more efficiently.

Levenberg-Marquardt method and its variants: These methods are specifically used for nonlinear least squares problems and can be effective for certain types of image restoration tasks.

Least Squares Method: This method aims to minimize the sum of the squares of the differences between the predicted and desired outputs.

Gradient-based methods, including stochastic gradient descent (SGD), batch gradient descent, and mini-batch gradient descent, are widely used for training CNNs. These methods calculate the gradient or slope of the loss function concerning the network parameters using the first-order derivative. The network parameters are then updated in the opposite direction of the gradient to minimize the loss function.

Backpropagation is an important technique used in CNN training. It involves propagating the gradient information backward through the layers of the network, allowing for efficient computation of the gradient at each layer. By iteratively adjusting the network parameters based on the computed gradients, CNN learns to improve its performance on the given task.

Overall, training a CNN involves iteratively optimizing the loss function using gradient-based methods and updating the network parameters through backpropagation. The choice of training algorithm depends on the specific problem and requirements of the task at hand.

1.3.2 Advantages of Convolutional Neural Networks

The advantages of using Convolutional Neural Networks (CNNs) for image data can be summarized as follows:

(i) Automatic Feature Learning: CNNs can automatically learn features from the data, eliminating the need for manual feature engineering. This is particularly advantageous in domains such as medical imaging, where handcrafted feature extraction can be challenging, time-consuming, and require domain expertise. CNNs have been successfully applied in medical imaging for tasks like mortality prediction, leveraging their ability to learn relevant features directly from the data.

(ii) Computational Efficiency: CNNs exhibit computational efficiency due to the concepts of weight sharing and dimensionality reduction. Weight sharing allows the network to reuse learned parameters across different spatial locations, resulting in a reduction in the number of parameters compared to fully connected networks. Additionally, the use of local receptive fields and sub-sampling helps in capturing spatial invariances and reducing the computational complexity of the network.

(iii) Availability of Benchmark Datasets: There are several benchmark datasets available for various image processing tasks, including denoising, segmentation, and classification. These datasets provide standardized evaluation metrics and enable fair comparisons between different methods. For example, in medical imaging, datasets with annotated images are available for tasks like diagnosing normal and abnormal conditions using techniques such as texture-based classification.

(iv) Transfer Learning: CNNs can benefit from transfer learning, where pre-trained networks on large-scale datasets are used as a starting point for other applications. By leveraging the learned features from pre-trained models, the initial layers of the network can capture general visual representations, while the latter layers are fine-tuned for the specific task at hand. Transfer learning reduces the training time and can be particularly useful when limited training data is available. This approach has been successfully applied in various applications, including image classification of histopathological images using pre-trained VGG networks.

Overall, CNNs provide automatic feature learning, computational efficiency, availability of benchmark datasets, and the potential for transfer learning, making them well-suited for a wide range of image processing tasks.

1.4 Performance Metric Evaluation

1.4.1 Peak Signal-to-Noise Ratio (PSNR)

Peak Signal-to-Noise Ratio is a widely used metric to evaluate the quality of denoised images. It measures the ratio between the maximum possible power of a signal and the power of the noise present in the image. The formula for calculating PSNR is shown in Equation (1.1) as follows:

$$PSNR = 20log_{10}\left(\frac{Max_f}{\sqrt{MSE}}\right) \tag{1.1}$$

Where, Max_f = maximum pixel value, MSE = Mean Square Error.

PSNR is expressed in decibels (dB), and a higher PSNR value indicates better denoising performance. It represents the ratio of the signal power to the noise power, with higher values indicating a higher signal-to-noise ratio and therefore better image quality.

It is important to note that while PSNR is a commonly used metric, it has some limitations. It is primarily based on pixel-wise differences between images and may not always correlate well with human perception of image quality. Other metrics, such as the Structural Similarity Index (SSIM) and perceptual metrics based on deep learning models, are also used to evaluate image quality in denoising tasks.

1.4.2 Contrast to Noise Ratio (CNR)

Contrast-to-Noise Ratio (CNR) is a metric used to quantify the difference in contrast between the target image and the background noise in an image. It is calculated using the following formula shown in Equation (1.2):

$$CNR = \frac{Contrast}{Noise} = \left| \frac{u_1 - u_2}{\sqrt{{\sigma_1}^2 + {\sigma_2}^2}} \right| \tag{1.2}$$

Where,$u_1, u_2, and\ \sigma_1, \sigma_2$ are the expected values and mean deviation of the signal. CNR measures the relative contrast of the target image concerning the background noise level. A higher CNR value indicates a greater contrast between the target and the background, which implies better visibility of the target.

CNR is commonly used in various imaging applications, including medical imaging, where it helps assess the detectability of structures or abnormalities in the presence of noise. It provides quantitative information about the distinguishability of the target from the background noise, aiding in evaluating the effectiveness of denoising techniques and image enhancement methods.

1.4.3 Mean Square Error (MSE)

The Mean Squared Error (MSE) is a common metric used to evaluate the quality of an image reconstruction or denoising algorithm. It measures the average squared difference between the original image and the reconstructed/denoised image.

While it is true that a lower MSE indicates a smaller average difference between the original and reconstructed images, it is important to note that MSE alone does not necessarily guarantee the best subjective quality of the image. The perception of image quality involves various factors, including visual appearance, texture preservation, sharpness, and overall reliability of the original image.

MSE is sensitive to scaling factors because it compares pixel values directly. If an image is scaled or normalized differently before computing the MSE, it can affect the numerical value of the MSE. However, changing the scaling does not necessarily lead to an image that is visually closer to the original. Therefore, simply minimizing MSE may not always result in the best subjective quality.

To achieve the best image quality, it is important to consider multiple evaluation metrics and subjective assessments. Other metrics, such as Peak Signal-to-Noise Ratio (PSNR)

or Structural Similarity Index (SSIM), can provide additional information about image fidelity, perceptual similarity, and preservation of structural information. Additionally, human visual perception and subjective evaluations are crucial in determining the quality of an image reconstruction or denoising algorithm is shown in Equation (1.3):

$$MSE = \frac{1}{n}\left[\sum_{i=1}^{i=n}\sum (X_i - x_i)^2\right] \tag{1.3}$$

Where n is referred to as data of the variable of all points, X_i and x_i are the observed value and the predicted value respectively.

1.4.4 Root Mean Square Error (RMSE)

RMSE parameter determines the difference in squares between the output and the input and is represented in Equation (1.4) [20]:

$$\text{RMSE} = \sqrt{\text{MSE}} \tag{1.4}$$

Where MSE= Mean Square Error

1.4.5 Structural Similarity Index Measurement (SSIM)

The Structural Similarity Index is a widely used metric in digital image processing for assessing the similarity between images. Unlike traditional metrics such as PSNR and MSE, which focus on pixel-wise differences, SSIM takes into account both structural information and perceived similarity. SSIM measures the similarity between two images based on three components: luminance, contrast, and structure. It considers the statistics of these components within the local windows of the image. The calculation of SSIM involves comparing corresponding windows (of size N x N) in the reference image (x_1) and the distorted image (x_2). The structural similarity formula computes the similarity index for each window, and the average of these indices gives the overall SSIM value. The formula for computing SSIM involves calculating the mean, variance, and covariance of the luminance, contrast, and structure components of the windows. These calculations capture the statistical properties of the image regions and their relationship to each other. By considering these components, SSIM can provide a more comprehensive assessment of image similarity compared to metrics like PSNR or MSE. SSIM has been shown to correlate well with human perception of image quality and is particularly useful in scenarios where the perceived visual quality is important, such as image compression, denoising, and restoration applications. It is worth noting that SSIM is just one of several

quality assessment methodologies, and the choice of which metric to use depends on the specific application and requirements. Different metrics have their strengths and limitations, and it is often beneficial to consider multiple metrics and subjective evaluations to obtain a comprehensive understanding of image quality. The formulation is shown in Equation (1.5):

$$SSIM\ (x1, x2) = \frac{(2\mu_1\mu_2 + k_1)(2\sigma_{12} + k_2)}{(\mu_1^2 + \mu_2^2 + k_1)(\sigma_1^2 + \sigma_2^2 + k_2)} \tag{1.5}$$

Where, μ_1= mean of x_1, μ_2 = mean of x_2, σ_{x1}^2 = variance of x_1, σ_{x2}^2= variance of x_2, σ_{x1x2}= co-variance of x_1 and x_2, k_1 and k_2 are variables to stabilize the separation with a weak denominator.

1.5 Loss Function

Loss functions play a crucial role in training neural networks by quantifying the difference between predicted outputs and the ground truth. The choice of the loss function depends on the specific task and the type of data being processed. For multiclass classification tasks, the commonly used loss function is cross-entropy. Cross-entropy loss measures the dissimilarity between the predicted class probabilities and the true class labels. It encourages the network to assign higher probabilities to the correct classes and penalizes incorrect predictions. In regression tasks where the goal is to predict continuous values, mean squared error (MSE) is often used as the loss function. MSE computes the average squared difference between the predicted values and the ground truth. It provides a measure of the overall discrepancy between the predicted and actual values. For image super-resolution (SR) models, different loss functions can be used to evaluate the quality of the reconstructed images. One commonly used loss function is pixel-wise L2 loss (mean square error), which measures the average squared difference between the reconstructed image and the ground truth. However, it has been observed that relying solely on pixel-wise L2 loss may not capture perceptual differences accurately. To address this issue, other loss functions have been proposed, such as content loss and adversarial loss. Content loss compares the high-level features extracted from the reconstructed image and the ground truth, aiming to match their perceptual similarity. Adversarial loss leverages generative adversarial networks (GANs) to encourage the reconstructed images to be indistinguishable from the ground truth in a discriminator network. These loss functions help in producing more visually pleasing and high-quality results. It's important to note that different loss functions have different mathematical formulations and may be

suitable for different tasks and data types. The choice of the loss function should be carefully considered and evaluated based on the specific requirements and characteristics of the problem at hand.

1.5.1 Adversarial loss (AL)

In an adversarial generative system, there are two main components: the discriminator and the generator. The goal of the generator is to generate synthetic data (i.e., images) that closely resemble the real data, while the discriminator's task is to differentiate between the real data (ground truth) and the generated data. The adversarial loss function, commonly known as the adversarial loss or the GAN loss, is used to train the generator and the discriminator in an adversarial manner. The loss function measures the ability of the generator to deceive the discriminator and the ability of the discriminator to correctly distinguish between real and generated data. The adversarial loss is typically defined as a minimax game, where the generator tries to minimize the loss, while the discriminator aims to maximize it. The loss function encourages the generator to produce synthetic data that is indistinguishable from the real data, while the discriminator aims to correctly classify the real and generated data. By training the generator and the discriminator in this adversarial manner, the generator learns to generate more realistic and high-quality data, while the discriminator becomes more effective at distinguishing between real and generated data. The ultimate objective is to achieve a state where the generator produces synthetic data that is difficult for the discriminator to differentiate from real data.

Overall, the adversarial loss function plays a critical role in the training of Generative Adversarial Networks (GANs) by driving the competition between the generator and the discriminator, leading to the generation of more realistic and high-quality synthetic data.

Therefore, the Jensen Shannon divergence is used in Equation (1.6).

$$AL_{loss} = \sum_{n=1}^{N} -logD_{\theta_D}(G_{\theta_G}(i^{LR})) \tag{1.6}$$

Where n is referred to as data of the variable of all points, D_{θ_D} = Probability function for the similarity between the reconstructed image and natural image G_{θ_G}.

1.5.2 Perceptual loss

The loss function you are referring to, which is based on the difference of pixels between two images, is often called pixel-wise loss or pixel-level loss. It is commonly used in

various image processing tasks, such as image denoising, image super-resolution, and image restoration. The pixel-wise loss measures the discrepancy between the pixel values of the generated image and the ground truth image. It calculates the element-wise difference between the corresponding pixels and typically aggregates these differences using a specific metric, such as mean squared error (MSE) or mean absolute error (MAE). The lower the value of the pixel-wise loss, the more similar the generated image is to the ground truth image. By minimizing the pixel-wise loss during training, the network is encouraged to produce images that closely match the ground truth in terms of pixel values. This loss function is effective when the task requires preserving fine details and accurately reproducing the pixel-level information of the input image. However, it's important to note that relying solely on pixel-wise loss may not capture higher-level semantic information or perceptual quality. In some cases, optimizing pixel-level differences alone may result in overly smooth or blurry outputs, as the network prioritizes matching individual pixels rather than capturing the overall structure or content of the image. To address this limitation, alternative loss functions have been proposed, such as perceptual loss or feature loss, which take into account higher-level features extracted by pre-trained convolutional neural networks (CNNs). These loss functions consider not only pixel-level differences but also the similarity of feature representations in deeper layers of the network, which can better capture perceptual quality and semantic information. In summary, while pixel-wise loss based on the difference of pixels is commonly used and provides a straightforward measure of similarity between images, it may not capture all aspects of image quality. It is often beneficial to consider other loss functions that incorporate higher-level features or perceptual information to achieve more visually pleasing and semantically meaningful results as stated in Equation (1.7):

$$P_{loss} = \frac{1}{w_i h_i} \sum_{X=1}^{w_i} \sum_{Y=1}^{h_i} ((I^{HR})_{X,Y}) - \left(G_{\theta_G}(I^{LR})_{X,Y}\right)^2 + AL_{loss} \tag{1.7}$$

Where, AL loss = Adversial loss, w_i = Width of feature maps, h_i = height of feature maps, I^{HR} = Target HR image, I^{LR} = LR image & X, Y = domain of image.

1.5.3 Super-Resolution Loss

Pixel-wise Euclidean loss, also known as mean squared error (MSE) loss, is a commonly used loss function in super-resolution tasks. It measures the difference between the pixel values of the generated high-resolution image and the ground truth high-resolution image.

The pixel-wise Euclidean loss calculates the squared difference between corresponding pixels in the generated and ground truth images and then computes the mean of these squared differences. By minimizing this loss during training, the network is encouraged to generate high-resolution images that closely resemble ground truth images in terms of pixel values. Super-resolution loss focuses on accurately reconstructing high-frequency details and preserving fine textures in the image. It penalizes large discrepancies between the generated and ground truth pixel values, encouraging the network to produce sharper and more visually pleasing results. While pixel-wise Euclidean loss is a straightforward and commonly used loss function in super-resolution, it is worth noting that it solely emphasizes pixel-level differences and may not capture perceptual quality or higher-level semantic information. As a result, the network trained using this loss function may prioritize pixel-level accuracy at the expense of overall image structure or content. Therefore, additional loss functions, such as perceptual loss or adversarial loss, are often used in combination with pixel-wise Euclidean loss to improve the visual quality and perceptual fidelity of the generated super-resolved images [23].

$$SR_{loss} = \left\| CNN_H(I_i^{LR}) - I_i^{HR} \right\|_2^2 \tag{1.8}$$

Where, I_i^{LR}, I_i^{HR} is a low-resolution and high-resolution image respectively, and $CNN_H(I_i^{LR})$ = output of network with input I_i^{LR}.

1.5.4 Pixel Loss

Pixel Loss is a common approach to measuring differences between images on a pixel level. It involves comparing the pixel values of two images and quantifying the dissimilarity between them. The per-pixel loss, often computed as the absolute difference or squared difference between corresponding pixels, provides a direct measure of pixel-level discrepancies between the generated image and the ground truth image. It indicates how much the pixel values in the generated image deviate from the corresponding pixel values in the ground truth image. By optimizing the network to minimize this per-pixel loss during training, the model is encouraged to generate images that closely resemble the ground truth images at the pixel level. This loss function is particularly suitable for tasks where pixel-level accuracy is crucial, such as image denoising, image inpainting, or image super-resolution. However, it's important to note that relying solely on per-pixel loss may not capture higher-level perceptual or semantic differences between images. Sometimes, even small pixel-level differences may not be visually noticeable or

important for the overall image quality. In such cases, incorporating additional loss functions that consider perceptual similarity, structural similarity, or adversarial discrimination can help improve the overall quality and visual fidelity of the generated images. This loss is estimated as in Equation (1.9):

$$Pixel_{loss} = \frac{1}{k1k2k3}\sum_{n=1}^{n}(\hat{I}_{HR} - I_{LR})^2 \tag{1.9}$$

Where, width (k_1), height (k_2), and many channels (k_3) of the evaluated images, respectively.

1.5.5 Structural Similarity Index Measurement (SSIM)

The per-pixel Loss function is commonly used for comparing the output high-resolution (HR) image generated by a super-resolution model with the original image (ground truth). By calculating the difference on a pixel level, this loss function provides a precise measure of dissimilarity between the two images. The per-pixel loss computes the discrepancy between corresponding pixels in the HR image and the ground truth image. It can be calculated as the absolute difference (L1 loss) or the squared difference (L2 loss) between the pixel values of the two images. The lower the per-pixel loss, the closer the generated HR image is to the original image in terms of pixel-level similarity. Using the per-pixel loss function helps in evaluating the accuracy and fidelity of the super-resolution model. By minimizing this loss during training, the model learns to generate HR images that resemble the ground truth images at a pixel level. This loss function is particularly useful for tasks where precise pixel-level reconstruction is important, such as image super-resolution or image restoration.

However, it's important to consider other factors such as perceptual quality, structural similarity, and overall visual fidelity, as relying solely on per-pixel loss may not capture higher-level details or perceptual differences that are important for human perception. Therefore, incorporating additional loss functions or perceptual metrics can help achieve better image quality and closer resemblance to the original image [1]. This loss is stated below in Equation (1.10):

$$SSIM_{loss} = \frac{1}{2}(1 - E_x[SSIM(I_{LR}, \hat{I}_{HR})]) \tag{1.10}$$

Where, Ex [·] = Expectation operator over the spatial coordinates, I_{LR} , $\hat{I}_{HR}$ = LR image & reconstructed image respectively.

1.5.6 L1 loss function

The L1 function is used to decrease the error which is the addition of absolute subtraction between the true value (Y_{ture}) and predicted value ($Y_{predicted}$). It is stated below in Equation (1.11):

$$\sum_{n=1}^{n} Y_{true} - Y_{predicted} \tag{1.11}$$

Where n is referred to as data of the variable of all points [23]. Y_{true} is the original image, and $Y_{predicted}$ is the HR image.

1.5.7 L2 loss function

The L2 loss function is used to decrease the error which is the addition of the square difference between the true value (Y_{ture}) and predicted value ($Y_{predicted}$), as stated in Equation (1.12):

$$\sum_{n=1}^{n} (Y_{ture} - Y_{predicted})^2 \tag{1.12}$$

Where n is referred to as data of the variable of all points, Y_{true} is the original image, and $Y_{predicted}$ is the HR image.

1.5.8 Run Time Effieciency in Image Super-Resolution

The runtime efficiency of a program or algorithm is typically calculated as the ratio of useful work performed by the program to the total time taken, often expressed as a percentage. The formula for calculating runtime efficiency can be defined as:

Run Time Effieciney = (Useful Work / Total Time) × 100% (1.13)

Where-

Useful Work -This represents the amount of meaningful computation or processing performed by the program. It's often specific to the task or application you're measuring. For example, in the context of a deep learning model, useful work could be the number of images processed or the number of calculations performed for a particular task.

Total Time: This is the entire time taken by the program to complete the task, including all processing time, overhead, and any waiting or idle time. The architecture is designed for image super-resolution, aiming to enhance the resolution of images while maintaining or improving their visual quality. To calculate the runtime efficiency of the architecture, we need to measure the time it takes to process an image through the network. Here's a step-by-step guide on how to calculate the runtime efficiency. Analyse the runtime efficiency results to understand how quickly the model can process LR images. We can compare these results to other super-resolution models or architectures to make informed decisions about which model to use in your application. It's essential to consider that runtime efficiency can vary based on factors such as the hardware used (e.g., GPU type), batch size (if we process multiple images simultaneously), and the complexity of the model. Therefore, we have performed experiment with different configurations and hardware setups to optimize runtime efficiency for specific case.

1.6 Other Improvements in Super-Resolution.

1.6.1 Data augmentation –

Data augmentation techniques such as scaling, cropping, rotation, flipping, and color jittering can be beneficial for single-image super-resolution tasks. Data augmentation is a common approach used to artificially increase the diversity and quantity of training data by applying various transformations to the original images. Here are some ways in which augmentation techniques can be useful in single-image super-resolution:

Scaling: Rescaling the input image to different sizes can simulate different levels of degradation and resolution changes. By training the model on images with various scaling factors, it can learn to handle different levels of super-resolution.

Cropping: Randomly cropping patches from the input image and using them as training samples can help introduce spatial variability and improve generalization. Cropping can also simulate different levels of detail in the images, which can be useful for training the model to handle different levels of super-resolution.

Rotation and flipping: Applying random rotations and horizontal/vertical flips to the input image can help increase the diversity of training samples. This augmentation technique can improve the robustness of the model to variations in orientation and direction.

Color jittering: Modifying the color properties of the image, such as brightness, contrast, and saturation, can help the model learn to handle different color variations in the input. Color jittering can make the model more robust to changes in lighting conditions and color distribution.

By applying these augmentation techniques to the training data, the super-resolution model can learn to generalize better and handle different variations and distortions that may occur in real-world scenarios. It can also help prevent overfitting and improve the overall performance and robustness of the model. It's important to note that the choice and combination of augmentation techniques may vary depending on the specific super-resolution task and dataset. Experimentation and careful selection of augmentation strategies are necessary to find the most effective approaches for a particular problem.

1.6.2 Multi-task Learning

Multi-task learning is a machine learning technique where a single neural network is trained to perform multiple related tasks simultaneously. In the context of image processing, such as super-resolution, denoising, and deblurring, multi-task learning can be applied to train a single network to address these different aspects. Here's how multi-task learning can be performed for denoising, deblurring, and super-resolution:

Denoising: The network is trained to take a noisy image as input and generate a denoised image as output. The loss function used for the denoising task can be based on pixel-wise Euclidean distance or other suitable metrics that measure the difference between the denoised output and the clean ground truth image.

Deblurring: The network is trained to take a blurry image as input and generate a deblurred image as output. The loss function for deblurring can be similar to denoising, aiming to minimize the difference between the deblurred output and the clean ground truth image.

Super-Resolution: The network is trained to take a low-resolution image as input and generate a high-resolution image as output. The loss function for super-resolution can be based on pixel-wise Euclidean distance or perceptual loss, which considers high-level features to capture the similarity between the super-resolved output and the high-resolution ground truth image. In multi-task learning, the network is designed to have shared layers that capture common features across different tasks, as well as task-specific layers that focus on learning task-specific representations. The shared layers enable the

network to leverage the shared knowledge and features between tasks, leading to improved generalization and performance. Concatenation can be used as a strategy to merge the outputs of different task-specific layers, allowing the network to produce a combined output that addresses all the tasks simultaneously. By jointly optimizing the loss functions for all the tasks, the network can learn to extract shared representations and effectively perform denoising, deblurring, and super-resolution tasks. The advantage of multi-task learning is that it can exploit the relationships and dependencies between different image processing tasks, leading to improved performance and efficiency compared to training separate models for each task. It can also help in cases where limited training data is available for individual tasks, as the shared representations can be learned from the combined dataset of all tasks.

1.6.3 Context wise Network fusion

In the context of super-resolution, multi-path learning can be applied to fuse different outputs of super-resolution models. This can involve training multiple super-resolution models with different architectures or settings, each focusing on a particular aspect of super-resolution (e.g., texture enhancement, edge sharpening, color restoration). The outputs of these models are then combined or fused to produce a final super-resolved image. The fusion of outputs can be achieved through various methods, such as simple averaging, weighted averaging, or more sophisticated fusion techniques like feature-based fusion or deep fusion networks. The goal is to leverage the strengths of different super-resolution models and combine their outputs to generate a visually pleasing and high-quality super-resolved image. By utilizing multi-path learning and fusing the outputs of different super-resolution models, it is possible to enhance the performance and improve the overall quality of the super-resolution process. Each model can contribute its specialized knowledge and expertise to the final result, leading to better texture details, sharper edges, and improved color representation in the super-resolved image.

1.6.4 Self-ensemble

Self-ensemble, also known as model ensembling or model averaging, is a technique used to improve the predictions of a machine-learning model by combining multiple predictions from different instances or versions of the same model. This approach is often employed in scenarios where the model is applied to variations of the input data, such as rotational images. In the context of rotational images, self-ensemble can be used to enhance the predictions of a model by leveraging the rotational invariance property. Since

the same object or scene can appear in multiple orientations or angles, a model trained on rotational images can benefit from self-ensemble to improve its predictions across different rotations. The self-ensemble process typically involves applying a transformation (e.g., rotation) to the input image and obtaining multiple predictions from the model for each transformed instance. These predictions are then combined or averaged to produce a final prediction. By incorporating self-ensemble into the prediction process, the model can capture the inherent variability introduced by different rotations of the same object. This helps to enhance the robustness and accuracy of the predictions, as the ensemble of predictions provides a more comprehensive representation of the underlying patterns in the data. Self-ensemble can be particularly useful in rotational image tasks, such as object recognition or image classification, where the orientation of the object can significantly impact its appearance. By considering multiple rotations and aggregating the predictions, the model can effectively handle rotational variations and make more reliable predictions regardless of the object's orientation.

1.7 Datasets for Super - Resolution

These days, there are a wide variety of datasets available that can be utilized for image super-resolution. These datasets contain images that vary greatly in terms of quantity, quality, resolution, variety, and other characteristics. While others will only provide you with HR images, some of them will provide you with LR-HR image pairs. If you only get HR images, you can typically get LR images by using the im-resize function in MATLAB with the standard parameters, which are Bicubic Interpolation with Anti-Aliasing. This is the case even if you only get HR images. We have compiled a list of image datasets that are frequently utilized by the SR community and have included them in Table 1.1.

Table 1.1 List of public image datasets for super-resolution benchmarks

Data Sets	Description	Format
BDS 100	Plants, Animals, Food, Building, People, etc.	JPG
BDS500	Plants, Animals, Food, Building, People, etc.	JPG
Div-2k	Environment, fauna, handmade object, flora, people, etc	PNG
L20	Plants, Animals, Food, Building, People, etc.	PNG

Manga 109	Manga Volume	PNG
Set 5	Baby, bird, butterfly, head, woman	PNG
Set 14	Humans, animals, insects, flowers, vegetables, comics, slides, etc	PNG
Urban 100	Structure, architecture, city, urban, etc	PNG

1.8 Use of Correlation Filter in Image Super Resolution

Correlation filters are used in computer vision and image processing to capture inherent properties between low-resolution and high-resolution images. The concept revolves around the idea that there are certain patterns, structures, or features that are common between a low-resolution image and its corresponding high-resolution version. These correlation filters help enhance the resolution of an image by exploiting this inherent relationship. Here is a step-by-step explanation of how correlation filters work to capture inherent properties between low-resolution (LR) and high-resolution (HR) images:

I. Data Preparation - To start, you have a pair of LR and HR images, with the LR image being a down sampled version of the HR image. The LR image has fewer pixels and lower detail compared to the HR image.

II. Feature Extraction - In the context of correlation filters, you typically extract features from both the LR and HR images. These features can be simple or complex, depending on the specific task. Common features include edge information, texture patterns, and colour gradients.

III. Learning the Correlation Filter - The key step in this process is to learn a correlation filter that captures the relationship between the LR and HR features. This filter essentially encodes the common patterns and structures shared between the two images.

IV. Convolution - The LR image is convolved with the learned correlation filter. Convolution is a mathematical operation that measures the similarity between the LR image and the filter. This operation identifies areas in the LR image that match the patterns encoded in the filter.

V. Up sampling - Once the convolution is performed, the result is an intermediate image that captures some of the inherent properties present in the HR image. However, this intermediate image is still at a lower resolution.

VI. Interpolation and Post-processing - To recover the HR image, interpolation techniques are applied to the intermediate image. These techniques use information from the LR image and the intermediate result to estimate the missing high-frequency details and increase the resolution. Post-processing steps may also be applied to enhance the final HR image's quality.

The correlation filter acts as a bridge between the LR and HR images by identifying common features or patterns. It allows the model to "learn" how the LR image relates to the HR image and then apply this knowledge to enhance the LR image's resolution. Correlation filters are commonly used in tasks like super-resolution, where the goal is to generate a high-resolution image from a low-resolution input. By capturing the inherent properties between LR and HR images, they help improve image quality and provide more detailed information, making them valuable tools in various computer vision applications.

1.9 Organization of Thesis

Chapter 1 provides the comprehensive study of image super-resolution is discussed. Upscaling of the image using different interpolation techniques, Deep Learning Strategies, and Loss functions with their parametric measurement to enhance the quality of the image is discussed. Advantages and Disadvantages are also discussed with their different applications.

Chapter 2 shows the details of the review of the literature in the area of image super-resolution. We talk about the usual ways to increase the resolution of an image, including Linear Networks, Residual Networks, Recursive Networks, Attention Networks, and Generative Adversarial Networks. Deep learning techniques that train with large image datasets are becoming the new norm. The CNN model is made with convolutional neural network architecture. The number of convolutional layers, training sets, optimization methods, and loss functions can all be changed. Improvisation of convolutional neural networks can also be done with the prior term design and residual learning. This chapter looks at the architecture and performance of convolutional neural-based super-resolution in terms of peak signal-to-noise ratio. Different deep learning methods, like convolutional

neural networks and generative adversarial networks, are explained in detail. It shows how traditional methods gradually gave way to deep learning methods.

Chapter 3 discuess the medical images super-resolution proposed networks are discussed with their quantities and qualitative analysis. The use of the dilated network increases the receptive field of the network which helps in improving the picture quality of medical images with the LRelu activation function which helps in activating the neurons for 0.1x value when $x<0$. And also discusses the architecture with the enhancement of the prior network by integrating complex wavelet transform with the CFIPC network.

Chapter 4 gives the study of general image super-resolution problems and their solution are discussed with proposed architecture on set 5, set 14, bds 100, bds400, and urban 100. Qualitative and Quantitively analysis is also discussed with comparison and discussion of the prior designed network.

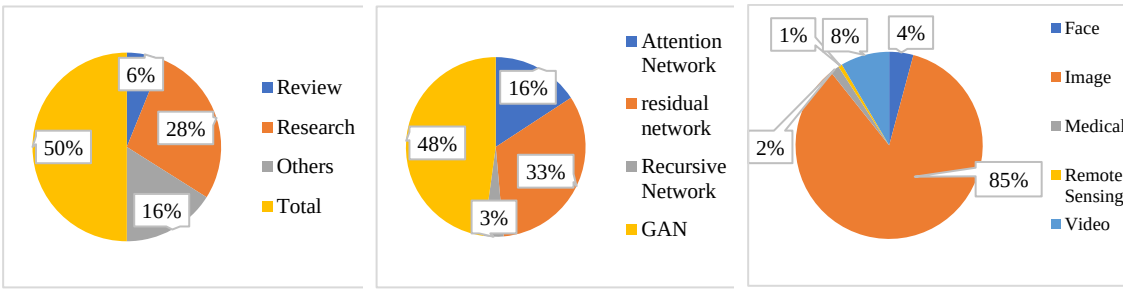
Chapter 5 addresses the issue of face image super-resolution and offers solutions for resolving facial image-related issues. Major issues associated with face image super-resolution during reconstruction end include the fusion of multiple types of features, which can result in image noise, blind spot generation, poor perceptual quality, and checkboard issues. Existing models are intended to improve the perceptual quality of face super-resolution, but they are unable to generate an image with improved perceptual quality due to data loss during reconstruction. Therefore, the newly designed architecture is discussed and compared to previous network architectures, demonstrating that the proposed architecture is superior.

Chapter 6 concludes the thesis along with future research directions.

CHAPTER 2: RECENT WORK

2.1 Introduction

To begin the literature review, this Chapter initially gathered research Chapters by conducting searches using terms like "image processing" and "deep learning" from various sources including Science Direct, Springer, ACM (Association for Computing Machinery), and IEEE (Institute of Electrical and Electronic Engineers), among others. The analysis of the bibliography was conducted based on the frameworks for SR, networks for SR, and applications for SR, as depicted in Figure 2.1. This article presents a collection of deep-learning-based image super-resolution networks, categorizing network designs into residual, recursive, progressive, and attention-based approaches, as illustrated in Figure 2.1 (a-c).



(a) (b) (c)

Figure 2.1 (a) Frameworks for SR, (b) Networks for SR, (c) Applications for SR Bibliometric Analysis for Research Contribution on Image Super-Resolution [23]

2.2 Network Designs

2.2.1 Linear Networks

The network structure of linear networks is characterized by a single path for signal flow, lacking skip connections or multiple branches. In these designs, several convolution layers are stacked sequentially, with the input flowing from the initial to the later layers. The main distinction among linear networks lies in the method of up-sampling, which can either be performed early or late in the network. It is worth noting that some linear networks aim to learn the residual image, representing the difference between the LR and HR images. Such networks are categorized as linear due to their linear network architecture, in contrast to residual networks that incorporate skip connections. The two sub-categories of notable linear network designs.

Early up-sampling designs: These linear networks initially up-sample the LR input to match the desired HR output size and then learn hierarchical feature representations to generate the output. A commonly employed up-sampling technique for this purpose is Bicubic interpolation, which, although effective, is computationally expensive. A seminal work following this pipeline is the SRCNN.

2.2.1.1 SRCNN

The Super-Resolution Convolutional Neural Network (SRCNN) is an important contribution to the field of deep learning-based super-resolution (SR), as the first successful attempt to employ only convolutional layers for this task. It was a ground breaking work that influenced numerous subsequent efforts in the same direction [10]. SRCNN's structure is relatively straightforward, consisting solely of convolutional layers. Except for the final layer, each layer is followed by rectified linear unit (ReLU) non-linearity. The network is linearly composed of three convolutional layers and two ReLU layers. Although all layers are convolutional, the authors gave each layer a unique name based on its function. From the input images, the first convolutional layer, also known as patch extraction or feature extraction, generates feature maps. The second convolutional layer, also known as non-linear mapping, transforms feature maps into high-dimensional feature vectors [10]. The final convolutional layer then combines the feature maps to generate the final high-resolution image. To train the network, a training dataset is created by extracting dense patches of size 32x32 from high-resolution (HR) images. The patches of the low-resolution (LR) input image are down-sampled and then upsampled using bicubic interpolation, resulting in patches of the same size as the high-resolution output image. SRCNN is a trainable end-to-end network to minimize the difference between the reconstructed high-resolution output images and the ground truth high-resolution images. Utilizing the Mean Squared Error (MSE) loss function accomplishes this [20].

2.2.1.2 VDSR

In contrast to the shallow network architectures employed in SRCNN [25] and FSRCNN [26], Very Deep Super-Resolution (VDSR) is built upon a deep convolutional neural network (CNN) architecture [27]. This architecture, commonly referred to as VGG-net, utilizes fixed-size convolutions (3×3) across all layers of the network. To address the challenge of slow convergence in deep networks, especially when dealing with 20-weight layers, the authors propose two effective strategies [28]. Firstly, instead of directly generating a high-resolution (HR) image, the network learns a residual mapping that

produces the difference between the HR and low-resolution (LR) image. This approach simplifies the objective and enables the network to focus primarily on capturing high-frequency information. Secondly, to expedite the training process, the authors employ gradient clipping within the range of [-θ, +θ]. This technique allows for very high learning rates, which aids in accelerating training. By implementing these strategies, VDSR aims to demonstrate that deeper networks have the potential to enhance contextual understanding and acquire generalizable representations that can be applied to multi-scale super-resolution tasks. The results obtained by VDSR support the argument that deeper networks can offer improved contextualization capabilities and learn representations that are applicable across various scales in super-resolution tasks.

2.2.1.3 DnCNN

In VDSR, the network is trained to directly predict the high-frequency residual instead of generating the complete super-resolved image. The residual image represents the difference between the low-resolution (LR) and high-resolution (HR) images. This approach allows the network to focus on capturing the specific details needed for super-resolution. On the other hand, the architecture of the Denoising Convolutional Network (DnCNN) [29] is simple and resembles that of SRCNN, consisting of stacked convolutional layers, batch normalization layers, and ReLU activation functions. Both SRCNN and DnCNN have reported promising results, but their performance is highly dependent on accurate noise estimation. These models lack knowledge of the underlying structures and textures present in the image, which can affect their performance. Additionally, the use of batch normalization operations after every convolutional layer in both SRCNN and DnCNN makes them computationally expensive. These operations contribute to the overall computational load of the models.

2.2.1.4 IRCNN

The Image Restoration CNN (IRCNN) [30] introduces a set of CNN-based denoisers that can be used together for low-level vision tasks like image denoising, deblurring, and super-resolution. The goal of this method is to combine discriminative CNN networks with model-based optimization techniques to make image restoration tasks more general. IRCNN uses the Half Quadratic Splitting (HQS) method to separate the regularisation and fidelity terms in the observation model. Then, a CNN is used to learn discriminative denoising prior, taking advantage of its better modeling ability and efficiency during testing. The CNN denoiser is made up of 7 dilated convolution layers stacked on top of each other, with batch normalization and ReLU non-linearity layers in between. The dilation operation makes the receptive field bigger, which lets more context information be modeled. The architecture uses residual image learning, which is similar to methods like VDSR [27], DRCN [31], and DRRN [9] which were used in the past to speed up the learning process. Also, small-sized training samples with no padding are used to avoid boundary artifacts caused by the convolution operation. A group of 25 denoisers is trained to deal with noise levels from 0 to 50, and they are then used together to fix images. This unified method works well for image denoising, deblurring, and super-resolution all at the same time. When it comes to up-sampling, linear networks usually do early up-sampling on images that are sent in. But this operation can be hard to do on a computer because the structure of the network needs to grow proportionally to handle bigger inputs. Post-up-sampling networks solve this problem by learning from low-resolution inputs and then up-sampling the features near the network's output. This plan leads to approaches that work well and uses little memory. In the next section, we'll talk more about these kinds of designs.

2.2.1.5 FSRCNN

The Fast Super-Resolution Convolutional Neural Network (FSRCNN) [26] is made to make super-resolution faster and better than SRCNN [10]. The main goal is to get 24 frames per second (fps) of real-time computation, which is much faster than SRCNN's 1.3 fps. FSRCNN has a simple structure with four layers of convolution and one layer of deconvolution. Even though all four layers do convolutions, FSRCNN gives each layer a name based on what it does. The feature extraction layer, the shrinking layer, the non-linear mapping layer, and the expanding layer are some of these. The step of extracting features is very similar to that of SRCNN. The only difference is the size of the inputs

and the size of the filters. SRCNN takes a bicubic patch that has been upsampled as its input, but FSRCNN takes the original patch without upsampling it. The second convolutional layer, which is called the "shrinking layer," reduces the number of parameters by using a smaller filter size (f=1) to reduce the size of the features. This helps make computers work better. The next convolutional layer is the non-linear mapping step. This step is important in both SRCNN and FSRCNN because it lets the network learn non-linear functions and has a big effect on performance. In the non-linear mapping layer of FSRCNN, the filter size is set to three, but the number of channels stays the same as in the previous layer. The fourth convolutional layer, which is called the "expanding layer," is the opposite of the "shrinking step" and adds more dimensions. The performance is improved by 0.3 dB because of this layer. The last part of the network is an up-sampling and aggregating deconvolutional layer. This is the opposite of the convolution process. In deconvolution, the image is convolved with the deconvolution filter, and the stride is used to increase the size of the output. In convolution, the stride is used to decrease the size of the output. Another difference from SRCNN is that after each convolutional layer, Parametric Rectified Linear Units (PReLU) are used instead of Rectified Linear Units (ReLU). PReLU adds learnable parameters that allow the slope of the activation function to be changed adaptively. This makes the network better at modeling. The mean-square error cost function is used to train both SRCNN and FSRCNN. The authors used the 91-image dataset [32] and added 100 more images they found on the internet to it. Using techniques like rotation, flipping, and scaling, the number of training images was increased by a factor of 19.

2.2.1.6 ESPCN

The Efficient Sub-Pixel Convolutional Neural Network (ESPCN) is a fast super-resolution (SR) method that can work on both images and videos in real-time. ESPCN takes a different approach than traditional SR methods, which first map the low-resolution (LR) image to a higher resolution using techniques like bicubic interpolation and then learn the SR model in the higher-dimensional space. ESPCN knows that this traditional pipeline makes the amount of computing needed much higher. Instead, ESPCN suggests that feature extraction should be done right in the LR space. By extracting features in the LR space, memory and processing needs are cut down by a large amount. At the end of the network, after the features have been pulled out, ESPCN uses a sub-pixel convolution layer. This layer combines the Low-Resolution (LR) feature maps and projects them at

the same time into the High-Dimensional (HD) space to make the High-Resolution (HR) image. The sub-pixel convolution operation in ESPCN is similar to convolution transpose or deconvolution, where a fractional kernel stride is used to improve the spatial resolution of the input feature maps. In ESPCN, each feature map is mapped with its own upscaling kernel. This gives the LR-to-HR mapping modeling more freedom. An L1 loss function is used by ESPCN to train the whole network. This loss function measures the difference between the reconstructed HR image and the original HR image, pixel by pixel. ESPCN has competitive SR performance while still being very efficient. This means that 1080p videos can be processed in real-time on a GPU. This is made possible by using LR feature processing, the sub-pixel convolutional layer, and an L1 loss function when training.

2.2.2 Residual Network

Residual learning is different from linear networks in that it uses skip connections to deal with the problem of gradients that go away. This method makes it possible to make deep networks by letting information flow directly from earlier layers to later layers. The importance of residual learning was first shown through image classification tasks. Recently, several networks have used residual learning to improve the performance of super-resolution (SR). In this method, algorithms learn the high-frequency parts between the input and the ground truth, which are shown by the residual information. By focusing on what it has already learned, the network can effectively pick up and piece together the missing details. Based on how many stages they use, residual learning approaches in the field of SR can be put into two groups: single-stage networks and multi-stage networks. Single-stage networks usually only have one set of skip connections, while multi-stage networks have more than one set of skip connections to improve information flow and learning across different scales or levels of abstraction. These residual learning methods have shown big improvements in SR performance by letting deeper networks be trained and letting high-resolution reconstructed images capture fine details and textures well. Figure 2.2 shows an example of how residual architecture is used to upscale an image and share information about each pixel. This is the standard architecture of residual learning.

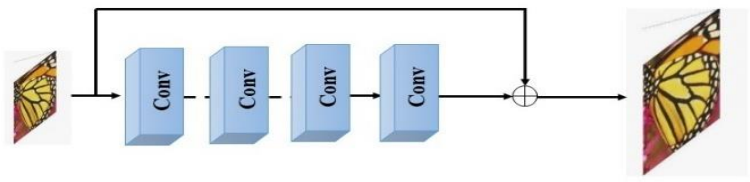


Figure 2.2 Residual Framework (referred from [27])

2.2.2.1 Single-stage Residual Nets

In single-stage designs, there is a single network that performs the super-resolution task. Here are some examples.

Enhanced Deep Super-Resolution (EDSR)

EDSR [33] changes the ResNet architecture, which was first made for image classification so that it can be used for super-resolution. By taking out the Batch Normalisation layers from each residual block and the ReLU activation outside of the residual blocks, they get big improvements. By adding scale-specific layers near the input and output blocks, EDSR expands its single-scale approach to work with more than one scale. These layers learn representations that depend on the scale, but most of the parameters are the same across scales. The network is trained using L1 loss and uses data augmentation techniques like rotations and flips for self-ensemble, where transformed inputs are passed through the network, reverse-transformed, and averaged together to make a single output. Older architectures like SRCNN, VDSR, and SRGAN don't do as well as EDSR and its multi-scale version, MDSR when it comes to quantitative measures like PSNR.

Cascading Residual Network (CARN) ResNet Blocks are used by Cascading Residual Network [31] to figure out how low-resolution inputs lead to high-resolution outputs. It adds both local and global modules that work together in a chain. In the local cascading connections, simple residual blocks are used. In the global cascading connections, features from intermediate layers are cascaded and converged on an 11-convolutional layer. Through multi-level representations and many shortcut connections, this method makes it easy to spread information quickly. CARN is trained with 64x64 patches from datasets like BSD and DIV2K [35], and the Adam optimizer is used to find the best fit with a '1 loss. The learning rate starts at 10-4 and gets cut in half every 4105 steps. These single-stage designs do better at super-resolution tasks than older architectures like SRCNN, VDSR, and other ResNet-based models like SRGAN that are closely related.

2.2.2.2 Multi-Stage Residual Nets

It seems like you have provided additional information about various super-resolution architectures. Here is a summary of the architectures mentioned:

Recursive Restoration Network (DRRN)

As proposed by Tai et al. [35], DRRN is an SR model that recursively restores the image frame. It utilizes residual learning and is trained using the L1 loss function.

Enhanced Deep Super-Resolution (EDSR)

EDSR [33] is a change to the ResNet architecture that improves the resolution of a single image. Each residual block has its Batch Normalisation layers taken away, and then the ReLU activation function is used. EDSR is trained with the L1 loss function and does a better job than previous architectures.

Multi-scale Deep Super-Resolution (MDSR)

The number of parameters in EDSR can be reduced through the use of shared parameters in MDSR, which is an extension of EDSR. To learn scale-dependent representations, scale-specific layers are utilized near the input and output blocks. In addition to that, it makes use of the L1 loss function for training purposes.

Cascading Residual Network (CARN)

CARN [31] uses ResNet Blocks to learn the relationship between low-resolution inputs and high-resolution outputs. This is accomplished through reinforcement learning. To efficiently propagate information, it makes use of both local and global cascading modules. CARN is trained with data augmentation and the L1 loss function, using 64x64 patches taken from datasets such as BSD and DIV2K.

Form ResNet

The DnCNN model is used as a foundation for the Form ResNet [44] algorithm. It is made up of two networks: a "Formatting layer" that takes into account Euclidean and perceptual loss, and a Different ResNet that is capable of learning structured regions. Form ResNet improves upon DnCNN results.

Balanced Two-Stage Residual Networks (BTSRN)

The BTSRN [45] consists of two stages: a stage with a low resolution and a stage with a high resolution. It is trained using the Adam optimizer with '2 distance serving as the loss function, and it employs a variant of the residual block known as projected convolution.

Residual Encoder-Decoder Network (REDNet)

Encoder-decoder architecture is utilized in REDNet [46], which is composed of convolutional and symmetric deconvolutional layers. These layers will also have skip connections added between them. REDNet comes in a few different flavors, each with a

unique number of layers. It learns using the MSE loss function as the primary tool for instruction.

Super-Resolution Image Convolutional Neural Network (SICNN)

SICNN [16] consists of feature extraction, deconvolution, and reconstruction stages, but further details about the architecture are not provided. These architectures employ different techniques and variations to improve super-resolution performance. The Loss function used by SICNN is super-resolution loss and is given by, super-resolution Euclidean loss given in equation 2.1 as:

$$\boldsymbol{Super\ Resolution_{loss}} = \left\| \boldsymbol{CNN_H}\left(\boldsymbol{I_j^{LR}}\right) - \boldsymbol{I_j^{HR}} \right\|_2^2 \quad (2.1)$$

Where, $\boldsymbol{I_j^{LR}}, \boldsymbol{I_j^{HR}}$ = with low-resolution and high-resolution images respectively, and $\boldsymbol{CNN_H}\left(\boldsymbol{I_j^{LR}}\right)$ = output of network with input $\boldsymbol{I_j^{LR}}$. The Cascaded Super-Resolution Identity Prior (CSRIP) model is a robust approach for super-resolution upscaling. It outperforms traditional interpolation techniques like bicubic and nearest-neighbor by leveraging the identity prior. The model utilizes a cascaded architecture to estimate the residual between the low-resolution and high-resolution images. The CSRIP model is particularly effective even for small images with dimensions as low as 24 x 24 pixels. It employs identity priors to enhance the reconstruction process and achieve high-quality results. Additionally, the CSRIP model demonstrates robustness in scenarios where there is object rotation in the image, making it suitable for tasks like image recognition.

2.2.3 Recursive Network

The framework you mentioned involves the use of recursive networks which is shown in Figure 2.3, where convolution layers are applied recursively or units are recursively linked. The main idea behind these designs is to break down the complex super-resolution problem into simpler sub-problems that are easier to solve.

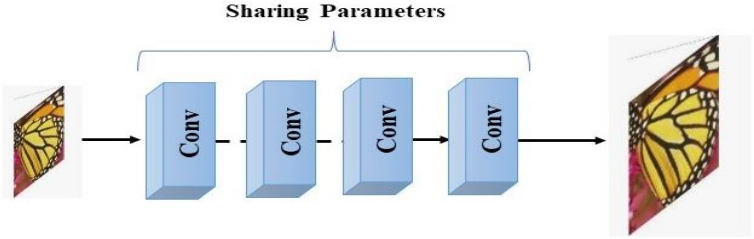


Figure 2.3 Recursive Framework (referred from [29])

Deep Recursive Convolutional Network (DRCN)

Multiple instances of the same convolution layers are applied. This method helps to maintain a constant number of parameters despite increasing recursion depth. The DRCN is composed of three sub-networks: the embedding net, the inference net, and the reconstruction net. The embedding network transforms the input image into feature maps, the inference network performs super-resolution by applying recursively a single layer of convolution and ReLU, and the reconstruction network converts the high-resolution feature maps back into grayscale or color images.

Deep Recursive Residual Network (DRRN)

Introduces a more sophisticated CNN model with conservative parameter complication. It combines residual image learning with local identity connections to produce a model with a deeper architecture and fewer parameters than previous models. Utilizing recursive learning and shared parameters substantially reduces memory and computational costs. DRRN employs multiple recursive blocks and achieves superior performance in comparison to other algorithms.

Memory Network (MemNet)

It consists of a block for feature extraction and several memory blocks. The memory block, which contains a recursive unit and a gate unit, plays a crucial role in the network. MemNet is trained with the MSE loss function and demonstrates promising results for super-resolution and other image restoration tasks, such as denoising and JPEG deblocking.

Super-resolution Feedback Network (SRFBN)

The low-resolution input is refined recursively to produce a high-resolution output using a recurrent architecture design. The feedback block (FB) includes projection groups that generate both high- and low-resolution features. Dense connections between low- and high-resolution representations within each FB and a global residual connection aid in the learning of the residual signal. SRFBN is trained with curriculum-based learning and performs well for various low-resolution image degradations. These frameworks emphasize information sharing and recursive learning to enhance super-resolution outcomes. Each model's design and training methods are distinct to improve performance and address the challenges associated with recursive architectures.

$$H_{-1} = max\ (0,\ W_{-1}*x + bias) \tag{2.2}$$

$$H_0 = max\ (0,\ W_0 * H_{-1} + bias) \tag{2.3}$$

$$F_1 = H_0 \tag{2.4}$$

An interference network was used for mapping using the convolution, Relu layer, and the output generated as :

$$F_2 = g^d(H) \tag{2.5}$$

The output of the network is given as input to the interfering network which consists of a W_{d+1} filter with Convolution + Relu layers which is recursive with b_{d+1} as post-biasing. I^R is referred to as a reconstructed image. *d-1, d+1* is the pre/post connections, as:

$$H_{d+1} = max\ (0,\ W_{d+1}*H_d + b_{d+1}) \tag{2.6}$$

$$I^R = max\ (0,\ W_{d+2}*H_d + b_{d+1}) \tag{2.7}$$

$$F_3 = I^R \tag{2.8}$$

2.2.4 Attention Network

The attention network is a technique used in super-resolution to prevent the influence of low-level information and focus on discriminative semantic portions of the images. It helps to enhance the performance and preserve important details during the super-resolution process and it is shown in Figure 2.4. In Single-Image Super-Resolution based on a Self-Attention Deep Neural Network [43], involves designing an attention network using feature map blocks. These feature maps are extracted from low-resolution images and concatenated to form a feature vector. This attention network guides the generator to focus on the most important parts of the image and prevents unwanted changes in the overall color tone. Another study [44] proposes an attention-based network for super-resolution. The specific details of this network are not provided in your description, but the main idea is to use attention mechanisms to selectively emphasize relevant image features and improve the quality of the super-resolved images.

In Multi-channel Network with Dense Attention for Image Super-Resolution [45], an attention-guided generator is suggested to perceive discriminative semantic portions between the source and target super-resolution images. This helps in preserving important details and preventing color tone changes. Additionally, a multi-scale discriminator is

used to enhance the rich texture details in the generated super-resolved images. To evaluate the performance of the suggested attention-based super-resolution methods, four datasets (Set5, Set14, Urban100, and BSDS100) are commonly used. The evaluation includes quantitative metrics such as PSNR (Peak Signal-to-Noise Ratio), which measures the fidelity of the reconstructed image compared to the ground truth. The reported PSNR values typically range from 23.54 to 29.52, indicating the improvement achieved by the attention-based methods. The qualitative evaluation includes visual comparisons, which show visually encouraging results and demonstrate competitive performance in perceptual measures such as NIQE (Natural Image Quality Evaluator) and PI (Perception Index). These perceptual measures are designed to align with human visual perception and indicate the superiority of attention-based approaches compared to state-of-the-art technologies.

Overall, attention-based super-resolution techniques show promise in improving the quality of super-resolved images by selectively focusing on important image features and preserving important details.

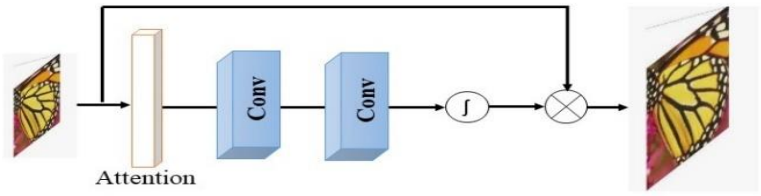


Figure 2.4 Attention Framework (referred from [52])

In the field of underwater image super-resolution, Zhang et al. [45] proposed an attention-guided multi-path cross-convolution neural network (AMPCNet). Their work aimed to improve the model's ability to learn and expand abstract feature representation for underwater image super-resolution. They introduced a multi-path cross (MPC) module that incorporates residual blocks and dilated blocks to enhance the network's performance. Additionally, an attention block was used to adaptively rescale the channel-wise characteristics for better discriminative representations. The up-sampling block was employed to aid in the reconstruction of high-resolution (HR) images. The evaluation of the AMPCNet network was conducted using quantitative measures such as PSNR (Peak Signal-to-Noise Ratio) and SSMIM (Structure SIMilarity Index Measure). The reported PSNR values ranged from 29.54 to 25.90 to 23.83, indicating improvement in the fidelity of the super-resolved images. Similarly, the SSMIM values ranged from 0.80 to 0.66 to

0.62, suggesting an improved structural similarity between the super-resolved images and the ground truth. The results of the research experiments demonstrated that the AMPCNet network outperformed other methods in terms of both quantitative measures and visual quality. However, one limitation of the study mentioned is that it did not utilize perceptual loss and adversarial loss, which are commonly used in super-resolution tasks to improve the visual appeal of the generated high-resolution images. This could be a direction for future research to further enhance the visual performance of the AMPCNet network. In another study by Chen et al., the focus was on generating perceptually appealing high-resolution images using a multi-attention generative adversarial network (MA-GAN) [21]. They proposed a GAN-based framework for image super-resolution and specifically applied it to remote-sensing scene images. The experimental results demonstrated the effectiveness of the MA-GAN approach, with PSNR and SSIM (Structural SIMilarity) performance indicators reported as 31.53 and 0.9007, respectively. However, one drawback mentioned in the study was the difficulty in creating tiny textures, which is a common challenge in super-resolution systems. This indicates a limitation of the MA-GAN approach in accurately reconstructing fine details in the super-resolved images. The AMPCNet and MA-GAN approaches show promise in improving underwater image super-resolution results, but there are still areas for further improvement, such as incorporating perceptual loss and addressing the challenge of generating fine textures.

2.2.5 Generative Adversarial Network

In the work by Goodfellow et al. [34], they proposed a model called GAN (Generative Adversarial Network) that is made for image super-resolution using an unsupervised learning method. The GAN model is based on a two-pair structure with a Generative network and a Discriminator network. The job of the Generative network is to take noise as input and make an image approximation that looks a lot like the high-resolution ground truth image. The generator's goal is to make high-quality images that are super-resolved. On the other hand, it is the job of the Discriminator network to tell the difference between real images (RI) and fake images (GI). It takes in an image, which could be real or made up, and figures out how likely it is that it is real. If the output probability of the discriminator for an image, F(k), is close to 1, the image is considered to be real. On the other hand, if the output probability of the discriminator is close to 0, it means that the image is generated and is closer to the ground truth. To improve the quality of the image reconstruction, the generator, and the discriminator are trained separately. The generator

tries to make super-resolved images that look exactly like real images. The discriminator, on the other hand, learns to tell the difference between real and made images. By training the generator and the discriminator in a way that makes them compete with each other, the GAN model learns to make high-quality, super-resolved images that are close to the ground truth. The model makes the generator better at making more realistic images, while the discriminator gets better at telling the difference between real and fake images. Figure 2.5 shows the architecture of the generator and discriminator. To get unsupervised image super-resolution, the GAN model uses a dual-pair framework with a generator and a discriminator. It uses adversarial training to improve the quality of the images it makes, making them more like high-resolution images from the real world.

$$F(k) = \begin{matrix} 0, k = GI \\ 1, k = RI \end{matrix} \quad (2.9)$$

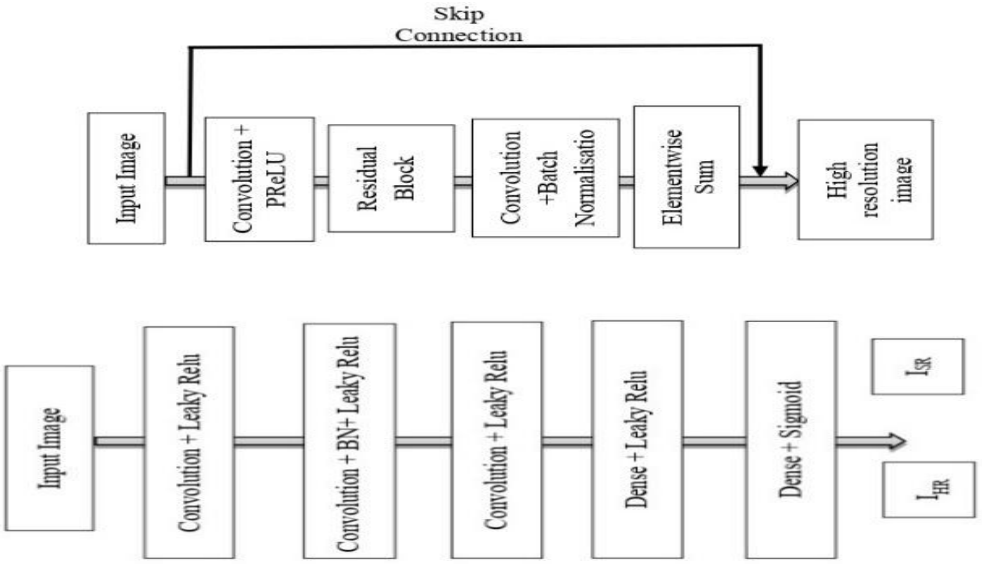


Figure 2.5 Generative Adversarial Network

In Generative Adversarial Networks (GANs), one of the limitations is the potential instability in training, which can lead to oscillations or difficulties in convergence. The specific issue you mentioned, where the discriminator is unable to effectively distinguish between real and generated images (resulting in an output probability of 0.5), is known as the "mode collapse" problem. Mode collapse occurs when the generator produces samples that are limited to a few modes or variations, ignoring the full diversity of the target distribution. When the discriminator is unable to differentiate between these limited modes and real images, it provides ambiguous feedback to the generator. This ambiguity can cause the generator's gradient to become unstable, leading to oscillations and

difficulties in learning **[34]**. To mitigate mode collapse and instability in GAN training, researchers have proposed several techniques. Here are a few commonly used methods:

Improved architectures

Modifying the generator and discriminator architectures can improve stability. For example, using deep convolutional networks and employing techniques like batch normalization can help alleviate mode collapse. Training tricks: Techniques like label smoothing (assigning fewer extreme labels, e.g., 0.9 for real images and 0.1 for generated images) and feature matching (encouraging the generator to match the statistics of real images' features) can improve training stability.

Regularization techniques

Adding regularization terms, such as gradient penalties (e.g., Wasserstein GAN with gradient penalty) or spectral normalization, can help stabilize the training process and prevent mode collapse.

Adaptive learning rate

Adjusting the learning rates of the generator and discriminator differently or using learning rate schedules can influence the training dynamics and improve stability.

Evaluation metrics

Employing additional evaluation metrics, such as Fréchet Inception Distance (FID) or Inception Score (IS), can provide more insight into the quality and diversity of the generated samples, helping to detect and mitigate mode collapse. It's important to note that GAN training can still be challenging, and finding the right balance between the generator and discriminator is an ongoing research topic. Different variations and improvements to the GAN framework have been proposed to address these limitations and enhance stability during training.

The loss functions of the discriminator are given as:

$$Loss(d) = -Ek \sim Ir\ x\ logF\ (k) - E\ k \sim Ig\ x\ log(1-F(k)) \tag{2.10}$$

& Loss function of the generator is given as:

$$Loss(g) = - Ex \sim Ig\ x\ log\ (1 - F(k)) \tag{2.11}$$

Where Ir and Ig are known as real data and the generated data.

Wasserstein GAN (WGAN) is indeed an improved variation of the GAN model that addresses some of the limitations and instability issues. WGAN introduces the Earth Mover's Distance (EMD), also known as Wasserstein distance, as a metric to measure the difference between the real and generated data distributions. Instead of using the Jensen-Shannon divergence or Kullback-Leibler divergence, which can lead to instability during training when the distributions do not overlap, WGAN focuses on minimizing the Wasserstein distance. The Wasserstein distance provides a more meaningful and continuous measure of the dissimilarity between distributions. The WGAN model employs a critic network, which plays the role of the discriminator, and a generator network. The critic network aims to estimate the Wasserstein distance by learning a function that approximates the difference between the true and generated data distributions. The generator network then tries to minimize this estimated distance, leading to improved stability and better convergence properties. The training objective in WGAN involves minimizing the difference between the expected critic scores for real and generated samples, rather than maximizing a traditional adversarial loss. This change in objective function helps avoid mode collapse and encourages the generator to produce diverse and high-quality samples. By using the Wasserstein distance and modifying the training procedure, WGAN provides a more stable and meaningful way to train GANs. It has been successful in generating higher-quality images and addressing some of the challenges associated with traditional GAN models. It's worth mentioning that apart from WGAN, other variations of GANs, such as LSGAN (Least Squares GAN), have also been proposed to improve stability and address issues like mode collapse by modifying the loss functions and training procedures. These improved GAN models continue to be an active area of research with ongoing developments and refinements.

$$Loss(d) = Ek \sim Ig\ x\ F\ k - Ek \sim Ir\ x\ F\ k \tag{2.12}$$

and the generator of WGAN tries to reduce the loss function as follows:

$$Loss(g) = - Ek \sim Ig\ x\ F\ k \tag{2.13}$$

In LSGAN (Least Squares GAN), the traditional sigmoid cross-entropy loss function used in the discriminator of a GAN is replaced with the least-squares loss function. This modification aims to provide more meaningful and stable gradients during training. The least-squares loss function is a regression-based loss that penalizes the discrepancy between the discriminator's output and the target values (1 for real samples and 0 for generated samples). By using this loss function, LSGAN encourages the discriminator to assign higher values to real samples and lower values to generated samples, pushing the generated samples closer to the decision boundary. The training objective in LSGAN involves minimizing the mean squared error (MSE) between the discriminator's output and the target values. This means that the discriminator network is trained to output values close to 1 for real samples and values close to 0 for generated samples. By utilizing the least-squares loss function, LSGAN aims to address the problem of vanishing gradients and mode collapse that can occur in traditional GANs. It provides more continuous and meaningful gradients to both the generator and discriminator networks, facilitating more stable and efficient training. LSGAN is an alternative approach to improve the training and stability of GAN models. It has been shown to generate higher-quality samples and exhibit better convergence properties compared to traditional GANs. It's important to note that both WGAN and LSGAN are just two examples of the many variations and improvements proposed for GAN architectures. Each approach tackles different challenges and objectives, and researchers continue to explore and develop new techniques to enhance the performance and stability of GAN models.

$$Loss(d)=12Ek\sim Irx[(F(k)-b)2] + 12Ek\sim Igx[(F(G(k))-a)2] \tag{2.14}$$

Where generator loss is trying minimized as follows,

$$Loss(g) =12Ek\sim IgxF\ G(k)-c2 \tag{2.15}$$

The UR-DGN (Ultra-Resolution Deep Generative Network) proposed by Prokili et al. [50] utilizes a fully connected layer and deconvolution in the generator network to enable end-to-end mapping of images and generate local details, particularly for face images. The goal is to reconstruct the image with a higher peak signal-to-noise ratio (PSNR). A combination of mean squared error (MSE) loss and adversarial loss functions are used to train the UR-DGN. The adversarial loss is determined by evaluating the output changes of the discriminator network when it is fed the output of the generator. The G network aims to produce high-resolution images from low-resolution input (ILR). The adversarial

loss of the generator is the output differences or discrepancies of the discriminator network, denoted by F when it receives the generator's output, G(ILR). The generator seeks to minimize this adversarial loss, which encourages the generated images to resemble actual high-resolution images and fool the discriminator. The UR-DGN model aims to improve the fidelity and visual quality of the reconstructed images by combining the adversarial loss and the MSE loss, which measures the pixel-wise differences between the generated images and the ground truth high-resolution images. The training procedure involves optimizing the generator and discriminator networks iteratively. The generator aims to produce realistic and high-quality images, whereas the discriminator seeks to distinguish between real and generated images with precision. This adversarial training framework enables the generator to improve its image generation capabilities over time by facilitating the learning of high-level image features. The original architecture proposed by Prokili et al. [50] would include the specific architecture and implementation details of UR-DGN.

$$Loss\ adversarial = -\ EILR \sim IL\ l\ log\ F\ G\ (ILR) \tag{2.16}$$

The loss function of the generator G in UR-DGN can be denoted as the combination of the mean squared error (MSE) loss and the adversarial loss. Given the input low-resolution face image ILR, the generator aims to generate a high-resolution face image G(ILR). By combining the MSE loss and the adversarial loss in the loss function, the generator is trained to minimize the pixel-wise differences with the ground truth images and to produce high-resolution face images that can deceive the discriminator network into classifying them as real. This dual-objective training helps the generator to learn the mapping from low-resolution to high-resolution face images while preserving the visual details and improving the overall quality. The specific values of λ_{MSE} and λ_{adv}, as well as the architecture and other implementation details, would be provided in the original Chapter by Prokili et al. [50].

$$lg = \lambda\ l_{MSE} + ladv \tag{2.17}$$

where λ is a weight to balance adversarial loss and MSE loss.

And the loss of discriminator D can be described as

$$ld = -\ EIHR \sim IHR\ h\ logD\ IHR - EILR \sim IL\ llog\ 1 - F\ G\ (ILR), \tag{2.18}$$

The loss function of the discriminator in ESRGAN is formulated using the Relativistic Discriminator approach, which compares the real high-resolution image Xr with the generated fake image X_f. The discriminator network F is designed to predict the probability of an image being real or fake. The standard discriminator function is expressed as F(x) = σ(x), where σ represents the sigmoid function applied to the input x, and (x) denotes the non-transformed discriminator output. The Relativistic Discriminator extends the standard discriminator by incorporating a relative comparison between the real and fake images. The loss function of the discriminator aims to minimize the difference between the discriminator's output for the real image and the mean output for the fake images. This is achieved by maximizing the relative relation between the real and fake images. The specific formulation of the loss function for the discriminator in ESRGAN would be provided in the original Chapter by Wang et al. [8]. It's worth noting that the architecture and training methodology of ESRGAN differ from the standard GAN network [34]. ESRGAN introduces changes such as using a residual in the residual dense block network instead of batch normalization, which improves the training process and enhances the capacity to capture fine texture details. These modifications contribute to the overall performance and quality of the generated high-resolution images.

$$L_D = -Ex_f [log (F (X_r, X_f))] – Ex_f [log (1 − F (X_r, X_f))] \quad (2.19)$$

Adversarial generator loss is also formulated as,

$$L_G = -Ex_r [log (1- F (X_r, X_f))] – Ex_r [log (F (X_r, X_f)] \quad (2.20)$$

This modification gives more sharp edges and fine texture details.

2.3 Application Areas of Image Super Resolution Models

Image super-resolution using Convolutional Neural Networks (CNN) has various applications across different domains. Some common applications include:

Medical Imaging

In medical imaging, high-resolution images are crucial for accurate diagnosis and analysis. CNN-based super-resolution techniques can enhance the resolution of medical images, such as MRI scans, CT scans, ultrasound images, and histopathological images, improving the visibility of fine details and aiding in better medical decision-making.

Surveillance and Security

In surveillance and security systems, low-resolution images captured by surveillance cameras or CCTV cameras can be enhanced using CNN-based super-resolution. This helps in improving the quality of the captured images, enhancing the ability to identify objects, people, or events of interest.

Remote Sensing

In remote sensing applications, such as satellite imagery and aerial photography, CNN-based super-resolution can enhance the spatial resolution of low-resolution images, providing more detailed and accurate information for applications like land cover classification, urban planning, and environmental monitoring.

Photography and Image Editing

Super-resolution techniques can be used in photography and image editing software to enhance the resolution and details of low-resolution images, improving the overall image quality. It allows photographers and image editors to upscale images without significant loss of quality.

Video Compression and Streaming

CNN-based super-resolution can be applied to improve the visual quality of compressed videos during video compression and streaming processes. By enhancing the resolution of low-resolution video frames, the perceived quality of the video can be improved, providing a better viewing experience for users.

Art and Entertainment

Super-resolution techniques can be utilized in the restoration of old or low-resolution artworks and photographs, preserving and enhancing their visual quality. It can also be applied in the gaming industry to generate high-resolution textures and graphics for immersive gaming experiences. These are just a few examples of the applications of image super-resolution using CNN. The technology has broad utility across various fields where enhancing image resolution is essential for analysis, visualization, and decision-making. Some of the articles related with the application are as follows –

Li An et al. [65] proposed a face image super-resolution method based on Canonical Correlation Analysis, which achieved a PSNR of 32.45 and SSIM of 0.843. This approach

utilizes the correlation between low-resolution and high-resolution face images to enhance the resolution.

Jiang et al. [66] introduced SRLSP, a CNN model for face image super-resolution. This model focuses on smooth regression and incorporates local structure prior information. However, it has a higher computational time requirement compared to other methods.

Yang et al. [67] presented a deep learning model for face image super-resolution based on sparse representations. This approach aims to reconstruct high-resolution images from sparse representations of low-resolution images.

Dwarikanath et al. [68] proposed a progressive GAN network with a triplet loss function for medical image super-resolution. This model utilizes a multi-stage architecture to progressively enhance image quality. The triplet loss helps improve the discriminative power of the generated high-resolution images.

Mansoor et al. [69] investigated the application of SRGAN [22] for medical image super-resolution. SRGAN is a popular GAN-based approach that has shown promising results in generating high-quality images. Mansoor et al. explored its effectiveness in the context of medical images.

Yanshan et al. [69] introduced an attention residual network called an Optical-guided super-resolution network (OGSRN) specifically designed for SAR (Synthetic Aperture Radar) images. This model incorporates an attention mechanism and aims to recover high-frequency information, leading to better visual quality. However, it may struggle with retrieving low-frequency information.

Shaoli et al. [70] presented an Efficient Video Super-Resolution Network (EVSRNet) for video super-resolution. This model achieves a good balance between quality and efficiency, catering to the needs of real-time video processing. However, it may exhibit lower PSNR values, suggesting the need for further improvements in image fidelity.

Table 2.1 Performance Evaluation on Benchmark Dataset Image SR of different CNN models for 2x magnification

	Method	**Set5**	**Set14**	**BSD100**	**Urban100**	**DIV2K**
2x	Bicubic [1]	33.68/0.93	30.24/0.869	29.56/0.843	29.56/0.843	32.45/0.90

	SRCNN [26]	36.66/0.954	32.45/0.906	31.36/0.887	31.36/0.887	34.59/0.93
	FSRCNN [21]	36.98/0.955	32.62/0.908	31.50/0.890	31.50/0.89	34.74/0.93
	VDSR [27]	37.53/0.958	33.05/0.912	31.90/0.896	31.90/0.896	35.43/0.94
	DRCN [29]	37.63/0.958	33.06/0.912	31.85/0.894	31.85/0.894	35.45/0.94
	DCSR [31]	37.58/0.959	33.03/0.912	31.90/0.896	31.90/0.896	-
	EDSR [25]	38.11/0.960	33.92/0.919	32.32/0.901	32.32/0.901	35.03/0.969
	CARN [24]	37.76/0.959	33.52/0.916	32.09/0.897	32.09/0.897	36.04/0.945
	SRGAN [22]	37.86/0.961	33.63/0.918	32.12/0.898	32.12/0.898	36.20/0.946
	ESRGAN [36]	38.10/0.962	33.75/0.919	32.26/0.902	32.26/0.9	36.39/0.951
	CSRIP [1]	39.12/0.963	33.82/0.959	32.45/0.910	32.45/0.91	37.13/0.963
4x	Bicubic [1]	28.43/0.8109	26.00/0.7023	25.96/0.6678	23.14/0.6574	28.11/0.775
	SRCNN [26]	30.48/0.8628	27.50/0.7513	26.90/0.7103	24.52/0.7226	29.33/0.809
	FSRCNN [21]	30.70/0.8657	27.59/0.7535	26.96/0.7128	24.60/0.7258	29.36/0.811
	VDSR [27]	31.35/0.8838	28.02/0.7678	27.29/0.7252	25.18/0.7525	29.82/0.824
	DRCN [29]	31.53/0.8854	28.03/0.7673	27.24/0.7233	25.14/0.7511	29.83/0.823
	DCSR [31]	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.98/0.827
	EDSR [25]	32.05/0.8910	28.53/0.7804	27.57/0.7354	26.07/0.7839	28.92/0.896
	CARN [24]	32.50/0.8973	28.72/0.7857	27.72/0.7418	26.67/0.8041	29.26/0.901
	SRGAN [22]	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	30.43/0.837
	ESRGAN [36]	32.73/0.9011	28.99/0.7917	27.85/0.7455	27.03/0.8153	-
	CSRIP [1]	32.893/0.9103	29.03/0.8123	27.90/0.7505	27.13/0.8224	-
8x	Bicubic [1]	24.4/0.65800	23.10/0.5660	23.67/0.5480	20.74/0.5160	-
	SRCNN [26]	25.33/0.6900	23.76/0.5910	24.13/0.5660	21.29/0.5440	-
	FSRCNN [21]	20.13/0.5520	19.75/0.4820	24.21/0.5680	21.32/0.5380	-
	VDSR [27]	25.93/0.7240	24.26/0.6140	24.49/0.5830	21.70/0.5710	-
	EDSR [25]	26.96/0.7762	24.91/0.642	24.81/0.5985	22.51/0.6221	-
	SRGAN [22]	26.98/0.7812	24.96/0.6445	24.89/0.6010	22.67/0.6377	-
	ESRGAN [36]	27.30/0.7846	25.20/0.6524	24.98/0.6120	23.29/0.6389	-
	CSRIP [1]	27.99/0.7910	25.86/0.662	25.89/0.6453	23.89/0.6431	-

2.4 Current Limitations and Challenges in Image Super Resolutions

Indeed, parametric evaluation metrics such as PSNR (Peak Signal-to-Noise Ratio) and SSIM (Structural Similarity Index) are commonly used to assess the performance of image super-resolution techniques. Higher PSNR values indicate lower reconstruction error and better similarity between the reconstructed image and the ground truth. These metrics help in quantitatively evaluating the quality of super-resolved images and comparing different super-resolution models. Several popular super-resolution models, including SRCNN, VDSR, EDSR, CARN, EUSR, SRGAN, and CSRIP, have shown their efficiency in terms of computational speed, network complexity, multiple upscaling support, and data augmentation support. However, each of these models has its limitations. SRCNN exhibits a slow convergence rate toward a solution. VDSR faces challenges in restoring images with high-frequency details and is not well-suited for blurred images, often requiring denoising as a pre-processing step. CARN does not address the occlusion problem in images, and the reconstructed images may not be visible for distant objects. SRGAN is less suitable for medical images and struggles with reconstructing text or structured scenes with high-frequency information. C-SRIP may struggle to reconstruct images with high levels of noise during the down-sampling process. One advantage of GAN-based super-resolution techniques is their ability to produce visually pleasing results by incorporating perceptual assessment rather than relying solely on simple estimation metrics. However, GANs are challenging to train due to the mode collapse issue, where the generator focuses on a limited number of modes rather than exploring the entire data space. It is worth noting that GAN-based super-resolution techniques are still evolving, with researchers continuously investigating and developing network architectures and loss functions to further improve visual quality. GANs have the potential to achieve superior reconstruction results compared to CNN-based methods due to their ability to capture a broader range of SR patterns with less explicit knowledge. In conclusion, GAN-based super-resolution techniques have shown promising results and are capable of achieving better reconstruction quality than CNN-based methods. However, GAN training can be challenging, and there is ongoing research to address its limitations and improve the overall performance of GAN models for super-resolution tasks

2.5 Problem Formulation

From the previous section, it becomes evident that there is a need to design a real-world image super-resolution network with unsupervised learning and variation in the

architectural pipeline to improve the performance metrics of peak signal-to-noise ratio and structural similarity. These are the issues that degraded the image quality at high resolution which is presented in the prior network.

i. The structure of Resnet uses 40% of the memory and it just normalizes the feature so it decreased the speed. The recovery of fine texture is required at larger upscaling.
ii. There is a problem is the restoration of images of high frequency and can't be suited for blurred images so denoising is required.
iii. They did not use geometric self-ensemble for an unknown downscaling category.
iv. The prior approach is less suitable for medical images and surveillance.
v. Reconstructing a text or structured scene at high frequency is a challenging task.

2.6 Research Objective

The objectives of the thesis are as follows:

(i) To develop a custom convolutional neural network-based upscaling model that de-blur the image with variation in the architectural pipeline, activation functions, and several layers.
(ii) To integrate the correlation filter module in the image super-resolution network. The image quality performance metrics can be improved if the images are super-resolved.
(iii) To improve High resolution generated images by reducing loss at the training time.
(iv) Reduce the size of the network layer for fast processing.
(v) To reduce the memory issue used at the training time. To improve the perceptual quality of the image.

2.7 Summary

This chapter gives a thorough look at and analysis of deep learning models for super-resolution images. Different CNN-based models with residual networks and GAN-based models with recursive networks are put into different groups. For a better understanding, we show how the PSNR results of different network designs on some benchmark datasets compare. It has been seen that adding analytical techniques to machine learning models can make the results even better. Even though there are many networks designed for artificial datasets, image super resolution in the real world is still hard. For practical

applications, the future lies in designing networks with learning frameworks that don't need to be watched. Some things that need more research are the transfer learning approach, adding graph theory to the neural network, prior design, and improving the receptive field.

CHAPTER 3: MEDICAL IMAGE SUPER-RESOLUTION USING DEEP LEARNING

In this Chapter, the following CNN models are designed for medical image super-resolution using correlation filter interleaved progressive convolution network (CFIPC) The CNN model is based on the progressive residual learning concept with principal component analysis on the intermediate feature maps and its enhanced architecture using CWT is also discussed in section 3.2 where it shows more improvement by integrating complex wavelet transform with the CNN architectures. The performance metrics analysis is done on Drive, Histopathological, and Ultrasound Datasets for both architectures.

3.1 Medical Image Super-Resolution using Correlation Filter Interleaved Progressive Convolution Network (CFIPC)

3.1.1 Introduction

Magnetic resonance imaging (MRI), Computed Tomography (CT), and Ultrasound are examples of common medical imaging techniques used for an expert diagnosis. On the other hand, these images are frequently lower quality and lack intrinsic information [51]. Due to hardware and existing imaging technology constraints, medical practitioners and researchers favor image super-resolution (SR) processing technology for medical diagnosis. Super-resolution is a technique for enhancing image features from a series of low-resolution (LR) images into a single high-resolution (HR) image [52]. Medical imaging [53], data analysis and visualization [54], object identification [55], and other applications all benefit from super-resolution. The aim behind super-resolution methods is to combine information from numerous distinct images to generate a single image. Algorithms attempt to reconstruct subsequent frames by extracting features from each image in a series [56]. The super-resolution relies heavily on information gained from LR images for the reconstruction of SR images, and researchers [57] propose a Bayesian-based super-resolution approach for application in MR brain imaging. Another way, in [58] researcher introduces the random forest method, which does not need data training and converts LR patches to HR patches directly, then uses the best regression model for image reconstruction using single or multiple LR images. Other techniques, such as the kernel estimate model [59] and the dictionary learning approach used in PET imaging [60], have proven accurate for single-image super-resolution. The approach described above has three drawbacks: it requires complicated optimization to extract HR images

and is time-consuming. Another issue is that excellent performance requires manually tuning parameters, making it inflexible. Various filtering techniques, such as the Kalman Filter and deep learning models, have been suggested to address this. [61] provided a remarkable study employing the Kalman filter for robust ultrasonic imaging that increased imaging performance. The suggested technique significantly increases the resilience and cumulative time and decreases the SNR ratio. The use of filtering algorithms to get SR images necessitates many computational and storage load problems. Apart from that, the existence of blind spots in the SR images restricts the usage of such approaches. In [62], the author developed enhanced Kalman filtering methods and a regression model to handle this problem. They employed an estimation theory-based framework with three kernel functions. Empirical findings imply that AISUKF with SK regression is better suited to the SR situation at hand, resulting in a denoised and detail-preserved output. Nowadays, because of the enormous capacity (payload) of the NN model and holistic learning, researchers are increasingly interested in using a neural network, a deep learning-based strategy, to solve the SR issue. These neural networks aid in acquiring previously applied functionality and the improvement of several deep learning techniques [63-65]. These upgraded Deep neural network DNN techniques are cost-effective and provide appropriate quality while also cost-effective [66], which is comparable to the three-layer hidden-layer-based super-resolution convolutional neural network (SRCNN) approach inspired by wavelet domain-based SR images. Many of these methods in [67] [68] offer a wavelet transform-based network architecture that is efficient and time-saving, with image super-resolution (SR) processing done in the wavelet domain. The vast majority of these algorithms provide great outcomes. However, the expenses of such calculations are substantial. Acceptable quality rises as deep learning algorithms improve and computing costs decrease. Coarse-to-fine SR CNN (CFSRCNN) was adopted in [69]. The proposed CFSRCNN uses a stack of multi-uses blocks to build a robust SR model. Compared to state-of-the-art SR models, our CFSRCNN model has high efficiency and strong performance on benchmark datasets. However, it has a poor convergence rate and a high training cost. This drawback of CNN is addressed in [70], which presents a lightweight enhanced SR CNN (LESRCNN) that uses a model to produce a high-quality image of various sizes. Similarly, some researchers worked on residual networks for SR [71]-[75]. In [72], the author presented the residual network based on the VDSR model. In this model, spatial location is extracted as an input feature. The model used the L1 loss function to train the model. The main drawback of this model

was that it takes long training epochs to reach optimal results. In [75], the author used the GAN model as the base model as it is known that GAN results in more effective results, but it suffers from issues such as model collapse and training issues. However, training an SR model with such a large number of convolutions and parameters generally requires a high computational cost and greater memory storage, limiting their applicability to SR with resource-constrained devices in the real world. Similarly, in [71] author presented the dilated CNN architecture, but the activation function was ReLU, and the loss function was MSE. But occlusion problem was observed at a small scaling factor. This was removed by applying a progressive network with interleaved correlation filter [55] layers proposed in this work. Therefore, this work is motivated to develop a super-resolution for medical imaging systems to improve resolution because it is essential for a more precise knowledge of anatomy, early diagnosis of anomalies, and correct evaluation of the size and shape of organs and diseases.

3.1.2 Methodology

This chapter is dedicated to designing an image super-resolution model for medical image diagnosis termed Correlation Filter (CF) Interleaved Progressive Convolution Network (CFIPC) for variable upscaling factors. Model architecture is presented in Figure 3.1. The model uses cascading of recursive dilated convolution layers with parametric-ReLU (pReLU) activation function and is termed a Dilated Convolution PReLU Network (DCPR). CFIPC used the dilated convolution layers due to its capability to increase the receptive field in SR images. But dilated convolution $\mathrm{Conv}_{\mathrm{di}}$May cause some blind spots if used directly in SR applications [71]. Therefore, to resolve this issue, the work designed a model, which slightly increases the kernel weight of the convolution layer by increasing the dilation factor. This leads to a linear increase in the receptive field without any loss in resolution. This is mathematically formulated as in Equation (3.1):

$$\mathrm{Conv}_{\mathrm{di}} = \mathrm{I} *_{\mathrm{d}} \mathrm{k}_{\mathrm{t}} \tag{3.1}$$

Where, I = discrete signal, k_t = kernel, $*$ = convolution, and d = dilation factor.

The recursive dilated convolution layers in CFIPC are formulated as Equation (3.2):

$$[\mathrm{k}_{\mathrm{t}} * \mathrm{x}, \alpha, \mathrm{k}_{\mathrm{t}} *_2 \mathrm{x}, \alpha, \mathrm{k}_{\mathrm{t}} *_3 \mathrm{x}, \alpha, \mathrm{k}_{\mathrm{t}} *_2 \mathrm{x}, \alpha, \mathrm{k}_{\mathrm{t}} * \mathrm{x}] \tag{3.2}$$

Where, k_{t} is kernel weight, $*_{\mathrm{n}}$ is dilation factor x = receptive field, and α= correlation filter factor.

The output of each CFIPC layer is feedbacked to minimize losses that occurred while training. An ADAM optimizer performs the minimization of loss. The model is in the recursive stage until the minimum loss is not achieved.

After reaching minimum loss, the output features are deconvoluted and passed on to the next recursive CNN layer. The minimized parameters are shared among recursive modules, and optimal parameters are passed from one layer to another. This increased the capacity of the Super-Resolution network and balanced the computational complexity. The loss function used here is mathematically represented as Equation (3.3) and Equation (3.4):

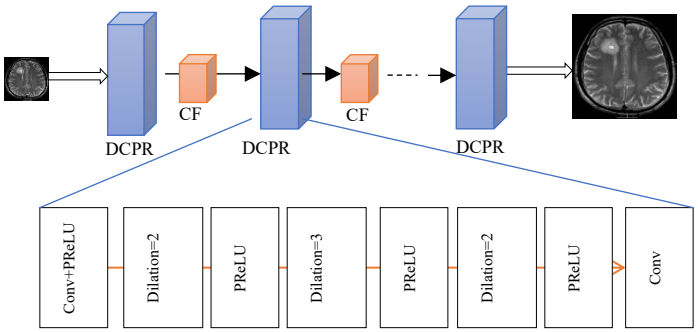
$$\text{loss}_{\text{CFIPC}} = \text{loss}_{\text{mse}} + \frac{(1 - \text{loss}_{\text{ssim}})}{2} \tag{3.3}$$

$$\text{loss}_{\text{mse}} = \frac{\sum_{i=1}^{N}((x_t - x_r)^2}{N}$$

Where, x_t=target value, x_r=reconstructed value, and N= Number of samples.

$$\text{loss}_{\text{SSIM}} = \frac{1}{2}(1 - E_o[\text{SSIM}(I_{LR}, \hat{I}_{HR})] \tag{3.4}$$

Where, SSIM= Structural Similarity, E_o [·] = Expectation operator over the spatial coordinates, I_{LR} , $\hat{I}_{HR}$ = LR image & reconstructed image respectively.



DCPR= Dilated Convolution PReLU Network, CF= Convolution Filter

Figure 3.1 Architecture of CFIPC

The memory usage complexity was reduced by decreasing the number of filters from 128 to 64 with increasing network capability. A deconvolution layer with a correlation filter (CF) is applied to optimize and upscale the feature maps at the output of each CFIPC module. The training algorithm is represented in Algorithm 3.1. The Gaussian kernel is

used to smooth the input image, and these smooth images are subtracted from the original image to generate a residual image with pixel information.

Algorithm 3.1 CFIPC Training

Input: LR_i = LR images; Training dataset,

$T_d = \{LR_i^n, HR_i^n\}$, where $n \in size(T_d)$;

k_t = kernel weight;

$*_n$ = dilation factor;

α= correlation filter factor

Output: SR_i^n, Super-resolution images.

For i=1: epoch

Do

$loss_{CFIPC}$

While $loss_{CFIPC}$ reaches convergence

end

$HR_i^n = [k_t * LR_i, \alpha, k_t *_2 LR_i, \alpha, k_t *_3 LR_i, \alpha, k_t *_2 LR_i, \alpha, k_t * LR_i]$

if

$loss_{CFIPC}$ = min

Return HR_i^n

end

end

3.1.3 Experimental Results and Discussion:

In this Chapter, we have conducted our experiments on medical image datasets, i.e., DRIVE [76], CHASEB1[77], MRI [78], and Ultrasound [79]. The proposed CFIPC SR network has been implemented on MATLAB 2020a with the MatConvNet package on Windows 10. To train the CFIPC network, 150 training images are randomly taken from DRIVE [76], CHASEB1[77], MRI [78], and Ultrasound [79] datasets. The training sample taken here is converted into image patches of size 48×48. The input images are taken by applying down-sampling on HR images. The HR images are considered to be labeled data. The filter size was initially set to 128 and slightly decreased to 32. The CFIPC network is designed for image super-resolution with the down-sampling x/2, x/4, x/8, and x/16. The input image patch size is 48 which is used to train the model. The images are degraded using a bicubic down sampler by x/16 during training. The training

is done for the 300 epochs. For testing the model, DRIVE [76], CHASEB1[27], MRI [78], Ultrasound [79], and Histopathological dataset [75] are used. The result findings are evaluated in parameters such as Peak Signal Noise Ratio (PSNR), Structural Similarity Index (SSIM), and T-Test Observer. The high PSNR and SSIM show better reconstruction, whereas a smaller score on T-Test.

3.1.4 Performance Evaluation:

In this Chapter, the performance analysis of the CFIPC SR model. The model is tested on five datasets as discussed above. In Table 3.1, we have presented the PSNR, SSIM, and T-Test evaluation on input image for scale factors 2x, 4x, 8x, and 16x, respectively. The HR image size for the DRIVE dataset taken for testing is 1024×1024, as a reference with IGAN [74]. While for the Histopathological dataset, the input LR image size taken for testing is 96×96, concerning SAI [75]. While for other datasets, the input LR image was considered to be of size 128×128. The result shows that by introducing the dilated recursive model, pixel information [80] is preserved which can be proved by achieving good performance evaluation values.

Table 3.1 Parameter Evaluation for Proposed CFIPC

	PSNR			**SSIM**			**T-Test**		
	s=2	**s=4**	**s=8**	**s=2**	**s=4**	**s=8**	**s=2**	**s=4**	**s=8**
DRIVE [76]	52.89	47.038	42.208	0.9952	0.984	0.9677	0.038	0.095	0.145
CHASEB1[77]	43.20	39.57	37.91	0.97	0.942	0.923	0.575	0.453	0.366
MRI [78]	32.605	29.491	29.007	0.9517	0.8804	0.8514	0.716	0.657	0.5590
Ultrasound [79]	41.782	40.396	40.30	0.9822	0.9635	0.966	0.6277	0.4179	0.2369
Histopathological dataset [75]	50.54	48.27	46.61	0.9986	0.9972	0.9972	0.0113	0.00005	0.0000021

Table 3.2 Comparative state-of-art on DRIVE dataset

Methods	PSNR/SSIM		
	4x	8x	16x
Bicubic	46.41/0.9992	41.16/0.9975	37.11/0.9943
VDSR [72]	43.04/0.9991	37.92/0.9975	33.43/0.9940
SRGAN [73]	30.13/0.9741	26.43/0.9501	-
IGAN [74]	47.78/0.9995	41.64/0.9982	38.00/0.998
OURS	47.038/0.984	42.20/0.9866	38.45/0.99

Table 3.3 Comparative state-of-art on Histopathological dataset

Methods	PSNR/SSIM	
	2x	4x
VDSR [72]	30.288/0.9594	-
SAI [75]	32.754/0.9509	30.314/0.8628
OURS	50.54/0.9986	48.27/0.9972

Table 3.4 Comparative state-of-art on Ultrasound dataset

Methods	PSNR/SSIM	
	2x	4x
SRCNN [76]	40.4189/0.980	32.21/0.8858
MSCNN [74]	40.4213/0.982	32.53/0.90210
OURS	41.782/0.9822	40.396/0.9635

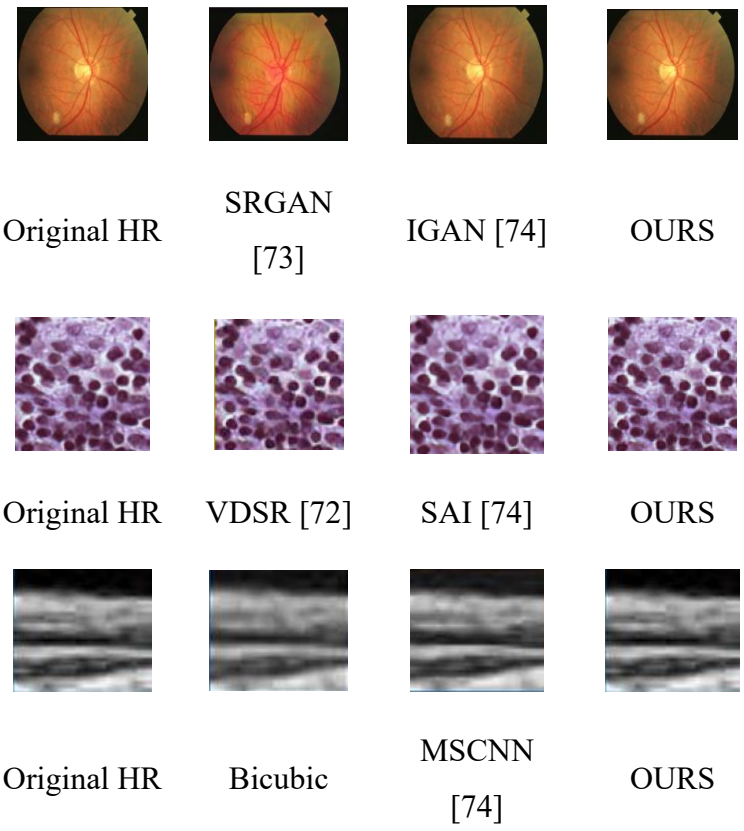


Figure 3.2 Examples of test results on different medical image datasets

3.1.5 Run Time Efficiency

Run Time Effieciency is defined as the ratio of total useful work (input image) trained w.r.to total time which is required for the training. Here, the input image used for the training is 150 images and the total time is 155 minutes for the completion of the training of image.

Run Time Efficiency = (Useful Work / Total Time) × 100%

= (150/155) × 100 % = 96.77%.

The run time Efficiency of of (CFIPC) is evaluated as 96.77%

3.2 Complex Wavelet Transform with Progressive Network for Medical Imaging Super Resolution.

3.2.1 Introduction

Computed Tomography (CT) [78], Glaucoma images, and Magnetic resonance imaging (MRI) are commonly used medical images for an expert diagnosis [79]. These images of good quality are necessary for accurate diagnosis [51]. Typically, medical images like spatial resolution are degraded by some imaging modalities like high-frequency detail or

intrinsic information lost. Medical images are one of the most important tasks in computer vision applications. To enhance the quality of the medical image super-resolution is used which helps in converting a series of low-resolution images in high resolution [52]. The main super-resolution is used to integrate all the frequency feature vectors of low resolution (LR) to generate high-resolution images which improve the quality and help medical practitioners for accurate diagnosis. Super-Resolution depends deeply on information obtained from LR images for SR image reconstruction. Traditional Techniques try to generate high-resolution images with the help of Interpolation (Bi-Cubic, Nearest neighbor, Bi-linear Transformation), Spatial Transform, and nonlinear transform. But still, some distortions like high-frequency details are missing, texture details are missing, edges are not smoothly generated and texture details are not fine. But in Trending some research with deep learning produces enormous results in visionary tasks like Object detection, classification, denoising, deblurring, and de-hazing [53][54]. There are many methods used to solve this task are pre-up sampling super-resolution, post-up sampling super-resolution, Residual Networks, Recursive Networks, Attention Networks, and Progressive and Generative Models. All the methods are generating high-resolution images with the high performance of PSNR and SSIM. But photorealistic quality is not improved[59]. In this work, we have tried to overcome the problem in multimodal medical imaging by proposing a progressive network with Complex wavelet transform (CWT). Complex wavelet transform is having the very unique ability to extract feature information in sub-bands that cover all the information of the image (horizontal, vertical, edges, and diagonal). Complex wavelet transforms firstly by our network integrated with the proposed progressive network with dilated convolution and correlation filter. This proposed network shows its potential in medical imaging. Initially, a Low-resolution image is fed as an input to the CWT which extracts the features from low-resolution space, and information on the feature is subdivided into four bands and then fed into the progressive dilated convolution network which is again used for mapping of feature vectors to gain more information and combined loss function is there to calculate losses during training after that output is sent to the subpixel convolution layer for upscaling at the reconstruction end after generating the high-resolution image and Correlation Filter is integrated into it to compare the image with the ground truth to ensure the performance of the network.

The major Contribution of the Chapter are as follow,

- The major use of CWT is to restore the image using rich pixel information and also provide multidirectional information details like horizontal, vertical, diagonal, and edges.
- Firstly, the Proposed Integrated network of CWT and Dilated progressive network with Correlation filter.
- Dilated Convolution is used to collect more pixel information by increasing spaces.
- Combined Loss Function we boost network performance during super-resolution.

3.2.2 Related Work

In the medical imaging domain of deep learning, SRCNN [10] was the first to introduce a network for super-resolution tasks Following many networks like nonlinear mapping, feature extraction, and reconstruction, hence, they failed to generate multifeatured of the input low-resolution image [57]. Zhou et al [63] give the significance of edge-based features in improving the super-resolution task by training the deep learning model with the sober edge of a low-resolution input image. After these residual learning approaches were introduced like VDSR [88], and Enhanced medical image super-resolution [84], but failed to give better perceptual quality and parametric evaluation like SSIM and PSNR also decreases and also give over smooth output. Ledig et al (2019)., proposed a generative adversarial network (GAN) that improves the perceptual quality by using the combination of networks like a generator and discriminator with using combined loss functions like adversarial and content loss [90]. After that DBPN improved the results by using a back projection network. All these networks are working well on natural image datasets but in medical imaging, their results are not admissible. These all networks are based on pixel space optimization and feature-based optimization which could not generate realistic output. Because of this reason, they are not suitable for medical imaging. There are many studies like IGAN [84], CylceGAN [22], Recursive networks [95], DCSR [96], and Progressive networks but still not acceptable in medical image clinical ability. Moreover, while using these networks on new datasets interference occurs [19][20]. In this proposed work we try to overcome the problems of the above network by designing a progressive dilated convolution network with a correlation filter with an integrated CWT network and combined loss function by using a transfer learning approach our model-generated high-resolution output is clinically acceptable and also works on the multiple medical imagining.

3.2.3 Methodology

Image super- resolution and Image denoising are low-vision tasks in computer vision therefor, it is very difficult to choose wavelet transform. Because in medical diagnosis high-resolution images need more features to extract the disease to provide better treatment for patients. Therefore, CWT stores more high frequency of feature information in many orientations which helps in the reconstruction end. In every step, CWT decomposes images into 4 bands with different frequency components like Horizontal, Vertical, Edge, and Diagonal information and each sub band is having different characteristics that are very important to give accurate and precise features. CWT sends the input signal with the help of L(p) and H(p), low pass, and high pass filters respectively. Mostly, an image I (x1, x2), where x1 and x2 pixel values are arranged in columns. The normal components of the CWT are represented by vertical, horizontal, diagonal, LV, LH, LD, and LE. When we use 1-level CWT to an LV input image to predict the LH, LD, and LE sub bands, we get the missing information features of the LV image, as shown in Figure 2. To get the super-resolution results, Progressive architecture is used in this with cascading recursive dilated convolution layers with parametric-Rectified Linear Unit (P-ReLU) activation function. Dilated convolution $dil.Conv_{di}$ the layer is used because of its capability to enhance the approachable field in the image. But it may cause a few unsighted spots if it is used in a straight line in a super-resolution application. To address this issue, a network is designed that increases the kernel weight of the convolution layer by increasing the dilation factor. This direct increase in the receptive field results in some loss of image resolution. Equation (3.5) expresses this mathematically:

$$dil.Conv_{di} = D *_d KW_t \tag{3.5}$$

Where, D = Discreate Time Signal, W_t = Kernel,

$*$ = Convolution, and d = Dilation Factor.

The recursive dilated convolution layers used in PDCN which is used to increase the receptive field are formulated as equation (ii):

$$[KW_t * X, \varphi, KW_t *_2 X, \varphi, KW_t *_3 X, \varphi, KW_t *_2 X, \varphi, KW_t * X] \tag{3.6}$$

Where, KW_t is kernel weight, $*_n$ is dilation factor X = receptive field, and φ= correlation filter factor. Element-wise sum is used to reconstruct the image generated by the Inverse Complex Wavelet Transform. The output of each PDCNCF layer is

feedbacked to reduce training losses using equation (iii) and equation (iv). The loss is reduced by the ADAM optimizer [68]. The network is in recursive mode until the minimum loss is reached. The generated features are deconvoluted by the subpixel layer after achieving the lowest possible loss before being passed to the next recursive convolution layer. Recursive networks share reduced parameters, and the best parameters are accepted from one layer to the next. This increased the dimensions of the super-resolution network while balancing the computational complexity. The mathematical representation of the combined loss function is shown in Equation (3.7).

$$loss_{\text{PDCNCF}} = loss_{mse} + \frac{(1 - loss_{ssim})}{2} \tag{3.7}$$

$$loss_{mse} = \frac{\sum_{i=1}^{N}((X_t - X_r)^2}{N}$$

Where, X_t=target image value, X_r=reconstructed image value, and N= Number of samples.

Loss SSIM is represented in equation iv.

$$loss_{SSIM} = \frac{1}{2}(1 - E_O[SSIM(\text{lr}_\text{i}\,, \text{hr}_\text{i}\,]) \tag{3.8}$$

Where, SSIM = Structural Similarity Index Measurement,

$E_O[\cdot]$ = Expectation operator over the spatial coordinates, lr_i , hr_i = low resolution image & high resolution respectively.

By increasing the network capacity using dilated convolution with different dilation factor complexity of memory use decreases by decreasing the filter size from 128 to 64. The output of each PDCN module, a sub-pixel layer with a correlation filter (CF), is used to improve feature map upscaling and is represented in Figure 3.3. The training algorithm is described in Algorithm 3.3. The Gaussian kernel is used to level the input image, which is then subtracted from the original image to yield a residual image containing pixel information.

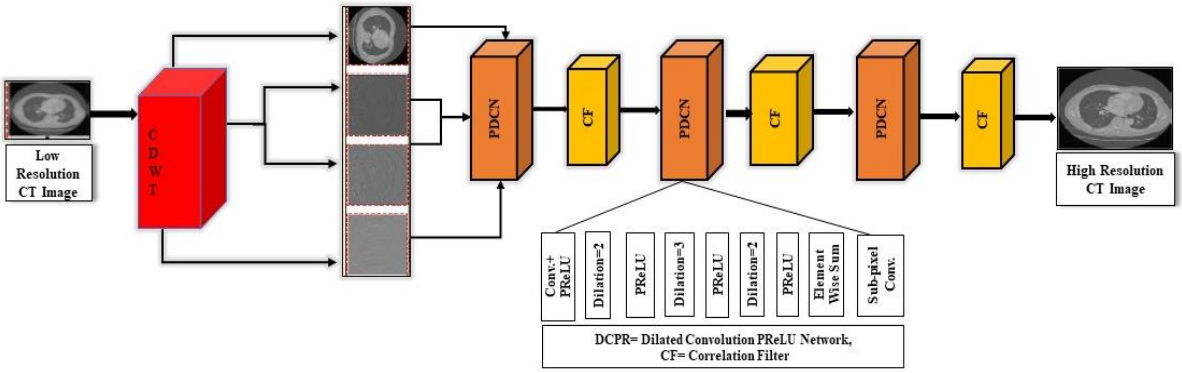


Figure 3.3 Architecture of PDCN

Algorithm 3.3 PDCNCF Training

Input: $\boldsymbol{lr}_{\mathrm{i}}$, low-resolution images;

CWT network + PDCN Network;

Training image dataset,

$$\mathrm{T_{Id}} = \{lr_{\mathrm{i}}^{\mathrm{n}}, hr_{\mathrm{i}}^{\mathrm{n}}\}, \text{where n} \in \text{size}(\mathrm{T_{Id}})$$

KW_{t} is Kernel Weight

$*_{\mathrm{d}}$ is dilation factor

φ = correlation filter factor

Output: $\mathrm{sr}_{\mathrm{i}}^{\mathrm{n}}$, super-resolution images.

1: Initialization: $KW_{\mathrm{t}}, *_{\mathrm{n}}, \varphi$, epoch;

2: While $\mathrm{loss_{PDCN}}$ reaches convergence do

3: $\mathrm{HR}_{\mathrm{i}}^{\mathrm{n}} = [KW_{\mathrm{t}} * \mathrm{lr}_{\mathrm{i}}, \varphi, KW_{\mathrm{t}} *_2 \mathrm{lr}_{\mathrm{i}}, \varphi, KW_{\mathrm{t}} *_3 \mathrm{lr}_{\mathrm{i}}, \varphi, KW_{\mathrm{t}} *_2 \mathrm{lr}_{\mathrm{i}}, \varphi, KW_{\mathrm{t}} * \mathrm{lr}_{\mathrm{i}}]$

4: Minimize ($\mathrm{loss_{PDCNCF}}$)

5: end

6: Return $\mathrm{hr}_{\mathrm{i}}^{\mathrm{n}}$

3.2.4 Experimental Results and Discussion

In this Chapter, we performed experiments on medical image datasets such as DRIVE [26], CHASEB1 [27], MRI [28], and Ultrasound [29]. On Windows 10, the proposed PDCN SR network was built with MATLAB 2020a and the MatConvNet package. To train the PDCN network, 200 training images from the datasets DRIVE [26], CHASEB1 [27], MRI [28], and Ultrasound [29] are chosen at random. The training sample is divided into 48 x 48 image patches in this case. Down-sampling HR images yields the input

images. HR images are considered to be labeled data. The filter size was initially set to 128 and then reduced to 32. The proposed network is set up for Super-resolution images with down sampling at x/2, x/4, and x/8. The input image patch used to train the model is 48 pixels in size. Images are degraded by x/8 during training using bicubic down sampling. The training is finished after 300 epochs. For model testing, DRIVE [76], CHASEB1 [77], MRI [78], Ultrasound [79], and Histopathological dataset [75] are used. The following metrics are used: Structural Similarity Index (SSIM), Peak Signal Noise Ratio (PSNR), Naturalness Image Quality Evaluator, and T-Test Observer. PSNR and SSIM have high values, indicating improved reconstruction, whereas a lower T-Test. The proposed network is performed on five datasets and to evaluate the performance of the image using PSNR, SSIM, and T-Test for the multiskilling factors is shown in Table I. According to the IGAN [74], an image size of 1024x1024 is used for testing the Drive dataset. Concerning SAI [75] 128x128 size image dataset is used. The results show that using the integrated network of CWT + PDCN it is preserved the information of pixels and gives the high perceptual quality of the image.

3.2.5 Comparison with State-of-the-Art

We Analysed and compared our proposed network with the deep learning networks based on medical image super-resolution like Improved Generative Adversarial Networks (IGAN) [24], SAI [25], VDSR [22], and SRGAN [23]. Table II, III, and IV shows the comparative result on different datasets and multi-scaling magnification. The proposed network is giving the promising result in multiple magnifications because we have used the proposed network with different loss functions. In Model 1 PDCN network is trained using perceptual loss then in Model 2 PDCN network is trained using Adversarial loss function and finally in Model 3 it used a Combined loss function to minimize the losses. An ablation study is performed with different loss functions to adapt this Combined loss function (Model 3) which also shows better results as compared with a perceptual loss and adversarial loss which is shown in Table V and Figure 3 with a Box plot performed on 200 high resolution generated image of a driving dataset. The initially generated image is subdivided into 20 sets and the average value of the PSNR is taken to assure the best result by using the K-10 cross-validation approach [30]. Figure 3.4 represents some reconstructed medical super-resolution images of various images. As shown in the Figure, PDCN recovered a more accurate result while preserving edges and textures. As a result,

PDCN achieves higher perceptual quality for all types of medical images. super-resolution.

Table 3.5 Evaluation of Proposed PDCN Parameters

	PSNR			SSIM			T-Test		
	D=2	D=4	D=8	D=2	D=4	D=8	D=2	D=4	D=8
MRI [28]	32.695	29.532	29.107	0.9517	0.9104	0.8614	0.726	0.6437	0.5690
Ultrasound [29]	42.142	40.412	40.130	0.9822	0.9735	0.956	0.6377	0.4279	0.2469
DRIVE [76]	52.89	47.038	42.208	0.9952	0.984	0.9677	0.038	0.095	0.145
CHASEB1[77]	43.20	39.57	37.91	0.97	0.942	0.923	0.575	0.453	0.366

Table 3.6 On the DRIVE dataset, state-of-the-art

Methods	**PSNR/SSIM**	
	4x	**8x**
Bicubic	46.41/0.9992	41.16/0.9975
VDSR [22]	43.04/0.9991	37.92/0.9975
SRGAN [23]	30.13/0.9741	26.43/0.9501
IGAN [24]	47.78/0.9995	41.64/0.9982
OURS	47.138/0.984	42.130/0.9966

Table 3.7 Histopathological dataset comparative state-of-the-art

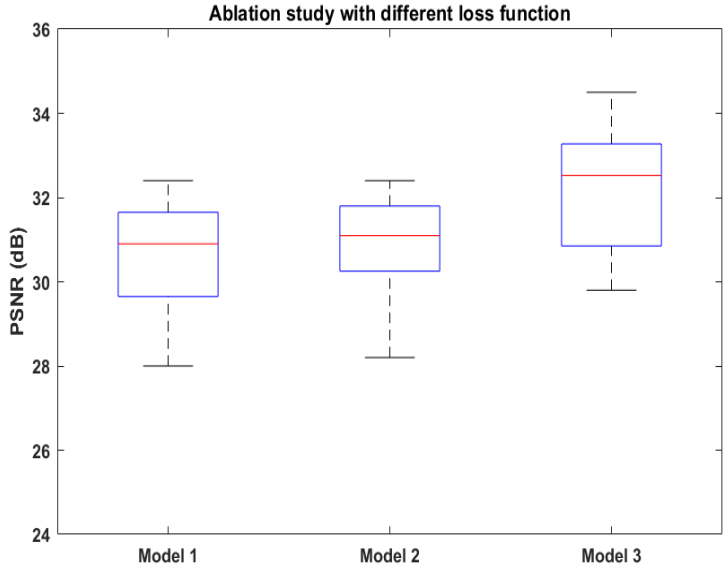
Methods	**PSNR/SSIM**	
	2x	4x
VDSR [22]	30.288/0.9594	-
SAI [25]	32.754/0.9509	30.314/0.8628
OURS	49.74/0.9860	49.67/0.9728

Table 3.8 Ultrasound dataset comparative state-of-the-art

Methods	PSNR/SSIM	
	2x	4x
SRCNN [16]	40.4189/0.980	32.21/0.8858
MSCNN [1]	40.4213/0.982	32.53/0.90210
OURS	**42.082/0.9822**	**40.536/0.935**

Table 3.9 Ablation Study of Proposed Network with Different Loss Functions

Model	**Loss Functions**	**MRI**		**Ultrasound**	
			SSIM	**PSNR**	**SSIM**
Model 1	Perceptual Loss Function	30.684	0.740	41.836	0.728
Model 2	Adversarial Loss	30.940	0.752	41.501	0.749
Model 3	Combined Loss	32.231	0.792	42.081	0.845



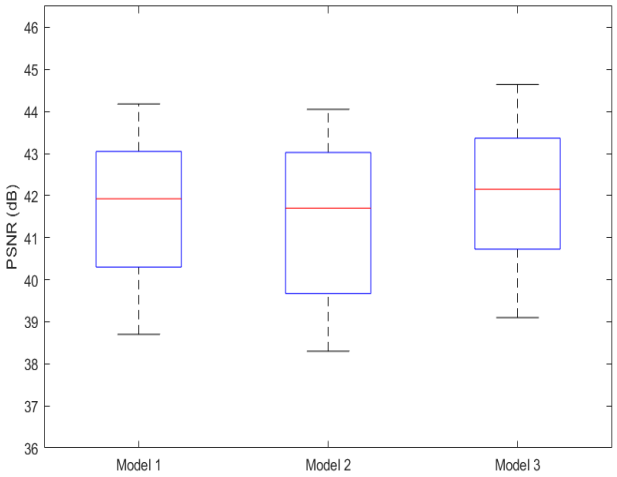


Figure 3.4 Result using K-10 cross validation

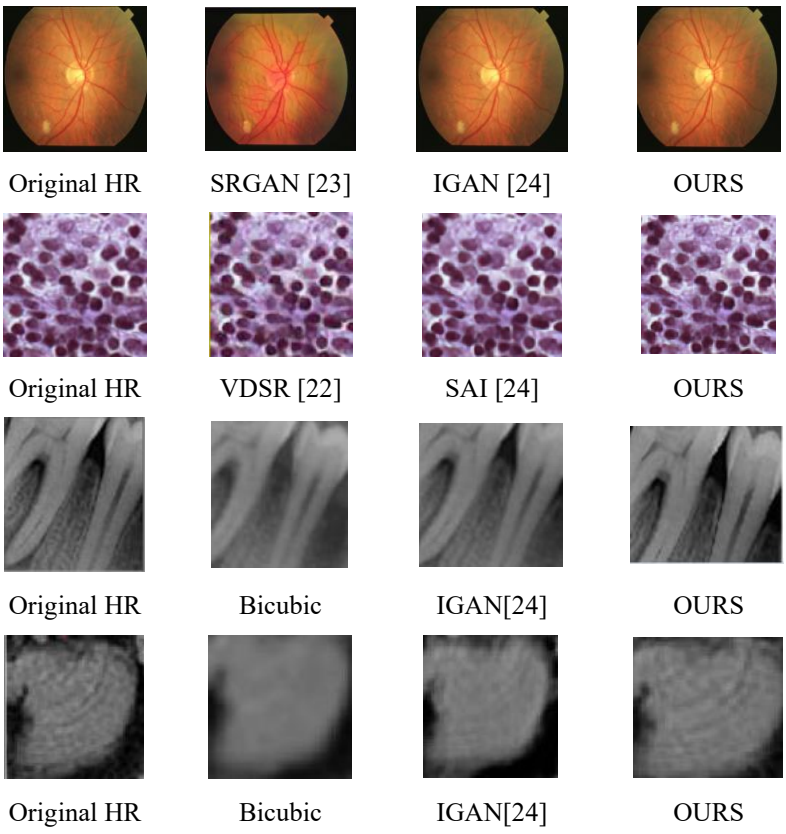


Figure 3.5 Examples of test results on different medical image datasets.

3.2.6 Conclusion

The integration of Complex wavelet transforms with correlation filter and recursive dilated convolution blocks in a proposed network, as well as a new combined loss function, are used to demonstrate the effectiveness of the medical image SR model in this Chapter. The effectiveness of the method is demonstrated by comparing the experimental

results to five different medical input images. The proposed architecture outperforms VDSR [22], SRGAN [23], SAI [25], and IGAN [24] networks on PSNR and SSIM. Furthermore, this Chapter assessed other parameters, such as the T-test, and K-10 Ablation study, that demonstrate the preservation of perceptual quality in reconstructed SR images, and concluded that PDCNCF has overcome the smoothness problem and improved the results.

3.3 Summary

This chapter shows a deep convolutional neural network that uses residual learning to fix images well. The addition of Dilated Convolution, LReLU, and Correlation Filter to the network makes it different from other CNN-based blurring networks in that the output can be processed even when the activation layer inputs are negative. Based on the number of feature maps, the whole network is split into two parts. People have tried to make more feature maps by using principal component analysis to find the best principal component based on eigenvalues and eigenvectors. The image details have become even clearer as the number of feature maps in each layer has grown. In the last step, an image is sharpened and smoothed by a correlation filter that uses brute force to work more quickly. The results of the experiments have shown that Histopathological, Drive, and Ultrasound Datasets have gotten better in both quality and size. Also, the proposed work can be expanded to clean up real-world noisy images, and the same model can be used for other image restoration tasks like image cleaning, deblocking JPEG images, deblurring, etc. And runtime efficiency is evaluated with 150 training image in the network is 96.77%

CHAPTER 4: IMAGE SUPER-RESOLUTION USING PROGRESSIVE NETWORK INTERLEAVED CORRELATION FILTER.

4.1 Introduction

In the field of computer vision, Single Image Super Resolution (SISR) is a very classic and important activity. Its primary objective is to improvise the Low Resolution (LR) image with super-resolution (SR) technologies [81]-[83]. It is generally used in applications such as surveillance that recognizes and identifies images obtained from security cameras in low-resolution images. Medical is another field where Magnetic resonance imaging (MRI) images in high resolution (HR) are challenging due to scan time, and signal-to-noise ratio (SNR). The solution to such a problem is to use super-resolution techniques by recreating HR images from the LR images [84]. This is formalized as:

$$I_{HR} = D(I_{LR}) \tag{4.1}$$

Where, D = degradation function

I_{LR}= LR image

I_{HR}= HR image

In the past few years, patch-based or interpolation methods are being deployed to produce SR images that rely on a smooth transition of adjacent pixels. But the drawback of the interpolation method is that it produces aliasing and ringing effects due to discontinuity in pixels. In recent years, with the development of deep learning, Convolutional Neural Networks (CNN) have been widely used in computer vision applications and analysis. While deploying CNN for Super-Resolution, complex mapping between LR and HR images is possible with better recovery of missing high-frequency information related to the image such as edges, textures, etc. Thus, it produces better performance as compared to other traditional approaches. In this Chapter, various networks have been analyzed to overcome the limitation of hardware that produces low-quality images. One of the very first and most popular SR methods is super-resolution convolution neural network (SRCNN) [10] which resulted in impressive outputs. The architecture of SRCNN consists of three layers i.e., patch extraction layer, non-linear mapping layer, and recreation layer. The patch extraction layer extracts dense patches from the input image and represents with non-linear mapping of convolutional filters. MSE loss function was also used.

The post-up sampling occurs in the high-resolution up-sampling space, which needs more computational space. This was reduced in post-sampling feature extraction in LR space. Another modification of SRCNN [10] was Fast Super-Resolution Convolutional Neural Network (FRCNN) [86] which was designed for post-up sampling by reducing convolution filters and lessening computation and memory utilization. This model was based on the inception model. This FSRCNN is better and faster than SRCNN. Then, residual networks were developed for upsampling of images such as Enhance Deep Super Resolution (EDSR) [87]. EDSR was developed using multiple residual blocks. The convolution layers were separated by batch normalization layer by reducing memory utilization by 40%. So, by seeing the effectiveness of the residual learning we have proposed a model termed Multilevel Progressive Recursive Dilated Networks with Correlation Filter (MPRDNCF) to enhance the adaptability of the model and to improve the objective function.

The key contributions of this Chapter are as follows:

- In this Chapter, state-of-the-art image super-resolution techniques and frameworks are presented along with their challenges.
- It presented a literature review of different super-resolution models designed using a deep learning approach.
- A model based on CNN for upscaling method for performance enhancement of image super-resolution.
- Finally, it gives a comparative state-of-art with existing techniques.

4.2 Related Work

In [88] author proposed a method of super-resolution for Positron emission tomography (PET) which is a 3D medical imaging modality. Super-resolution (SR) is a method of converting Low resolution (LR) images into high-resolution (HR) images. PETs are of great relevance in the medical field. In this Chapter, a network is designed which takes four types of input namely radial location, Axial location, and HR MR, and the main input in all methods is LR PET. These inputs are then passed through various convolutional layers and rectified linear units (ReLU). The following methods have been implemented. All network training is done using the L_1 loss function rather than L_2 because L_1 shows improved performance.

The dataset was obtained from Alzheimer's disease which belongs to neuroimaging field of medical science. The study shows that CNN vastly outperforms penalized deconvolution and partial volume correction at the SR imaging task. The study concluded that deep networks CNN produces better results and are clinically more promising. This Chapter shows that the drawback is the lack of portability of trained models to input with image characteristics. In [89] author proposes two models using three layers of CNN and four layers of CNN. The first model with 3 layers of CNN uses four optimization algorithms named Stochastic gradient Descent (SGD), Adaptive Gradient (AdaGrad), Root Mean Square prop (RMSprop), and Adaptive moment estimation (ADAM) which helps in improved performance of CNN whereas 4-Layers CNN use only ADAM optimization algorithm because, among four, the performance of ADAM is much better than rest three algorithms. In 3-layer CNN, the first 2 layers of Convolution are followed by Rectified Linear Unit (ReLU). This ReLU is used for mapping inputs to outputs and the generated MSE is used as a loss function.

In 4-layers CNN each layer is designed to perform a specific task the first layer is responsible for extracting features from the input image, the second layer is responsible for the non-linear mapping of feature maps, the third layer is responsible for the fusion of features extracted in previous layers and last that is fourth layer summed up the features to finally reconstruct HR image. A comparison of experimental results of all four optimization algorithms shows that ADAM is the best among them. If 3-layer CNN and 4-Layers CNN are compared then four-layer CNN is optimal as it already uses the best optimization algorithm. Moreover, an extra layer of Convolution extends its capability and improves its performance to reconstruct the corresponding HR image from LR-image. The reconstructed images achieve higher scores on PSNR and SSIM as compared to the previous traditional method of interpolation. But this work is limited to selected images of MRI and CT only. In [90] a multi-supervised super-resolution model called Path SRGAN is constructed to learn the mapping of low-resolution cytopathological images into high resolution so that it can help in the accuracy and early detection of cervical cancer. To improve these super-resolution techniques, this Chapter proposes a Generative Adversarial Network (GAN) based method using CNN. The algorithm is designed as a two-stage generator which makes its performance better than the previous classic U-Net and SRResNet generators. The limitation of this model was that image pairs of different resolutions are required to train the model, and this adds difficulties in terms of obtaining training data. It is quite expensive to generalize our data of diverse scanning

styles which needs the gathering of more image pair. Just like in the medical field, super-resolution is needed in other LR images [91]. Unlike previous methods of SR that aim at increasing the PSNR and SSIM metrics which results in Blur images which is a major issue in the field of the face recognition system. This study proposes a method of adversarial framework that does not only produce HR-image but it also addresses the issue of Blur image. In this study, a novel adversarial network structure is proposed to solve both the issues of deblur and HR image. In [11], a five-layer CNN is proposed to perform the aforesaid task. Firstly, the spatial resolution of the Low-resolution input image is done to form the feature image. Then these feature images are fed into encoders to extract the hidden features. These features are conveyed to a 2-branch encoder which then produces HR images with and without blur. The proposed method consists of 5 layers of CNN, face region prior, and a generator with a modified U-Net and discriminators. The input LR image is firstly mapped into 8 x spatial resolution utilizing five layers CNN which is then passed to two branched encoders. Discriminators are also applied to verify whether the images produce are fake or real facial images. The image generated is sharper and more realistic. The method is designed in such a way that is more advantageous in the face recognition field. In the method, one of the branches is responsible for the reconstruction of facial structure and the second one is responsible for the generation of images with more fine facial details and thus produces two images blur and without blur respectively. To avoid data loss, skip connections are also used. The proposed model is trained using the ADAM optimizer which is trained for 200 epochs first on the CelebA-HQ dataset. It takes a 32 x 32 input image to produce a 256 x 256 output image. The limitation of [11] was that if the blur LR image is too small then it is difficult to generate an HR image. It is also assumed that the camera position is correct or assumed that the image taken is unrealistic uniform. If the image taken has movements while clicking the picture or the image taken as input has motions then the proposed method for overcoming the blur of the image will not be successful. In [92] the proposed model is based on CNN to improve the spatial resolution of infrared (IR) images. It also removes artifacts.IR is important for various fields and therefore it is essential to store good-quality IR images. Generally, IR is of poor quality as compared to visible (VIS) images because of various noises in IR images due to imperfections in the IR imaging system and disturbances in the environment. Keeping in mind the concept that the Low-frequency band affects the objective quality of the image whereas High-frequency sub-

bands affect the perceptual quality significantly. Unlike visible (VIS) images IR images are different and lack various characteristics.

VIS has more clear details as compared to IR images as it is formed by the thermal difference. Also, the VIS image contains more boundary information therefore author took IR images guided by VIS images for the reconstruction purpose. SSIM and the Correlation coefficient are calculated to check the difference between an IR image and a VIS image to avoid an error. To achieve the said task the process goes on with IR/VIS image dataset. For an image Low frequency means color changes slowly whereas High Frequency means changes happen faster or drastically. Considering these two scenarios author proposes a two-branch CNN, the first branch is responsible for the recovery of LF subbands of IR image to suppress the noise and the second one is responsible for the reconstruction of remaining detail information. An algorithm is composed of two functionalities First Image Generation Network (IGN) and Artifacts Remove Network (ARN). Rather than mentioning the exact number of layers in the network Convolution blocks are used at various stages for extracting features, reconstruction, and suppressing noise. To validate the work perceptual similarity is estimated and taken as loss function. For training, CVC14 is used for 50 epochs for the detail restoration by IR images and 100 epochs for the detail restoration of VIS images. The model is trained on TNO. On comparing it has been noticed that PSNR and SSIM are not well matched to show good quality but results are better than ground truth perceptually. For qualitative results, dataset CVC14 is composed of 2 sequences containing 3695 images for the day and 3390 images during the night. To prove, its performance is better than previous methods, the test is also done on real captured images. The limitation of this work was that the size of paired IR/VIS image dataset is small as well as variety of pictures from various cameras is not considered.

In [93] author has considered both spatial and spectral resolution of images for SR. The architectures are named Simultaneous spatial-spectral joint super-resolution (Sim SSJSR) and a separated spatial-spectral joint SR (Sep SSJR). In Sim SSJR several convolution layers are used with a Convolutional kernel in each layer that results in one more extra constant dimension as Dc that represents the no of kernels in each layer. Testing the model with different no of layers shows that a five-layer CNN is best. Each layer is attached with a ReLU activation function to adaptively learn the parameters of rectifiers. Traditional MSE measured is much larger in case of outliers therefore l 2-norm-based loss function is considered and in Sep SSJR separate reconstruction is done. Two different

CNNs are built one for solving the issue of HR-MSI spectral issue and another one for resolving the spatial issue of LR-MSI. All training is done using an ADAM optimizer where a backpropagation strategy is adopted for training the dataset with a learning rate of 0.0002. Overall, two loss functions are considered namely absolute reconstruction error and average means square error are considered. In [94] author focuses on the same task of SR for face images. Generally, these methods are classified into three categories namely interpolation-based, restoration based, and learning-based. Interpolation methods are fast but their reconstruction is unstable. Restoration methods have stable reconstruction but require more calculation time. The last learning-based method is the most important SR as it autonomously learns the mapping function for the reconstruction of HR images. In [94] author termed the model a Guided Cascaded Face Super Resolution Network (GCFSRnet) comprises of two layers. The first layer is a prior recovery network (G_1) and the second one is the structural constraint network (G_2). In the first layer PReLU activation layer with two convolution layers was used. In G_2, three convolution layers and to recover the HR feature map with deconvolution layers. Four loss functions are considered namely reconstruction loss, perceptual loss, landmark loss, and adversarial loss respectively. The dataset used in this study is CASIA WebFace for training because the study requires front-face images. The training is done in two phases. The model performed well with the upscale factor 4x and 8x but was still left to explore upscale factor 16. GAN networks need to be very strong enough to simulate the real face image based on very less information extraction from the LR image. Another drawback of this Chapter was that the differentiation between a reconstructed image and a GT image is too large and the study focuses on front facial expression mainly and therefore does not generate better results with side poses. In [15] gives two networks one is SSNet-M which is considered to be the light one and an enhanced network of the same is SSNet. The proposed method uses a sampling method after convolution layers. In both the architecture sampling block are placed followed by Convolution layers. Two depth-wise separable convolutions are adopted in SNet-M to perform Non-linear mapping. After that supersampling is done and then pixel shifting is done. Finally, 3 x 3 Convolution layers are set for down-sampling to produce the output image. In SSNet an enhanced version is made with multi-recursive learning. To get the non–linear mapping, Features are extracted from the input image and fed into Depth wise separable convolution and then these non-linearly mapped features are fed again into Depth wise separable convolution using 5 x 5 layers. The Pointwise convolutional used is 5 x 5 and 1x1 respectively. To validate the

work, the mean square error (MSE) loss function is used. THE DIV2K dataset is used for training and testing and training is done by using interpolation. Also, ADAM is applied for better results. The model is trained for 100 epochs. In this work, the sampling is done in both ways up sampling and down sampling in a way to achieve the desired output which increases the computational complexity.

4.3 Problem Identification

These are the issues that degraded the image quality at high resolution which is presented in the prior network [85][95] [103-105].

- The structure of Resnet uses 40% of the memory and it just normalizes the feature so it decreased the speed. The recovery of fine texture is required at larger upscaling.
- There is a problem is the restoration of images of high frequency and can't be suited for blurred images so denoising is required.
- They did not use geometric self-ensemble for the unknown downscaling category.
- The prior approach is less suitable for medical images and surveillance.
- Reconstructing a text or structured scene at high frequency is a challenging task.

In this section, a super-resolution problem is formulated to regenerate the HR image from the input LR image. The degradation of the LR image is evaluated as in equation (4.2):

$$Y = D_s X \tag{4.2}$$

Where, X= HR image

Y= LR image

D_s= Down-sampling degradation scale

D_s maps the input image X of size P x Q into P/s and Q/s images.

In this Chapter, we have performed D_s using bicubic down-sampling having scaling factor 's'.

$$X = U_S D_S X \tag{4.3}$$

Where U_s = upscaling with scaling factor 's'.

In is noted that if we apply up-sampling on down-sampled images then some information is lost during down-sampling as well as the up-sampling process. So, while designing the SR technique, it is required to minimize these losses. This loss or error can be reduced by designing optimal layers and selection of optimal parameters. These are discussed in the next section in detail.

4.4 Methodology

In this section, we will systematically describe the proposed Multilevel Progressive Recursive Dilated Networks with Correlation Filter (MPRDNCF) for image super-resolution and analyze the techniques. The proposed MPRDNCF is an advance version of SRNet which is formulated and described in detail in the following subsections.

4.4.1 Overview of MPRDNCF Network

MPRDNCF network is cascading of recursive CNN models which up-samples input image in steps of 2x magnification up to 8x magnification. As shown in Figure 4.1, it consists of multiple cascaded layers that consist of two main stages i.e., feature extraction and image reconstruction.

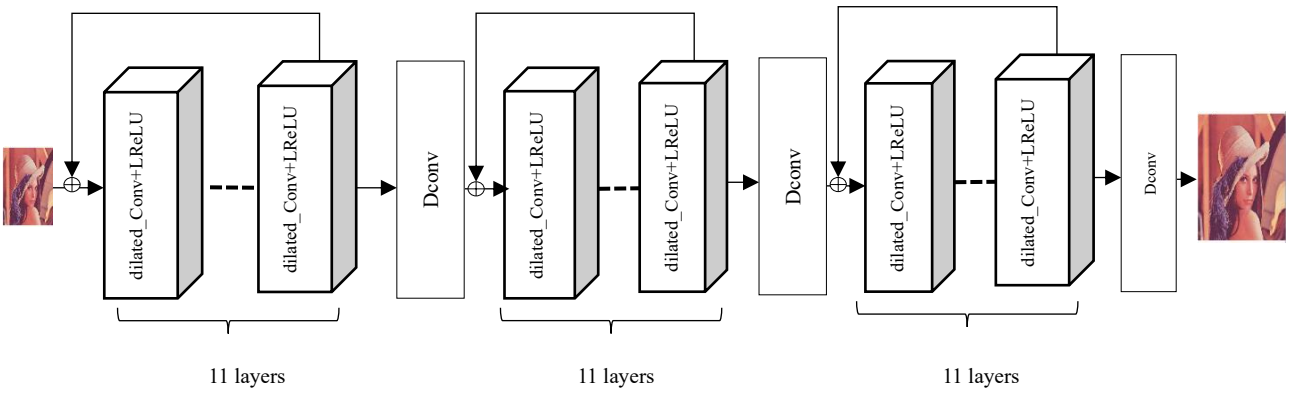


Figure 4.1 Basic Architecture of Proposed Model

In this model we have proposed the recursive layers of dilated convolution layers with leaky ReLU activation function and the difference between the convolution and dilated convolution is shown in Figure 4.2. The benefit of using dilated convolutions [97] is that it expands the reception convolution field and increases the effectiveness of the pixel-level resolution. It is formulated as in Equation (4.4):

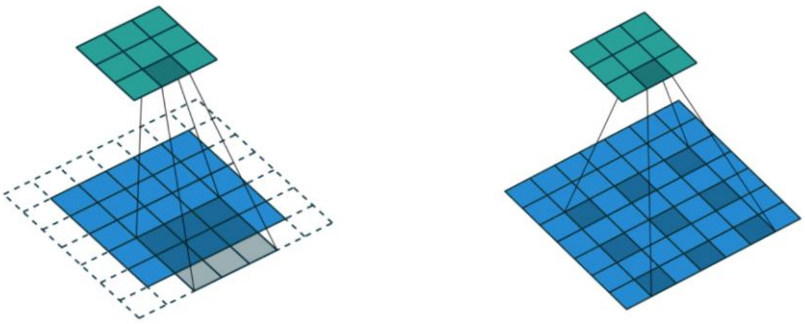
$$Standard\ conv = I * k_t$$
$$Dialated\ conv = I *_d k_t \tag{4.4}$$

Where, I = discreate signal

k_t = kernel

$*$ = convolution

d = dilation factor



(a) Standard Convolution (b) Dilated Convolution (d=2)

Figure 4.2 Difference between standard and dilation convolution [18]

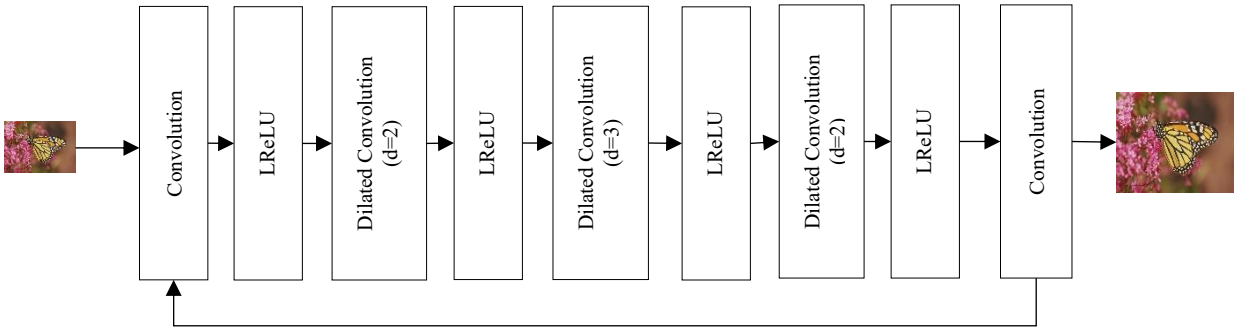


Figure 4.3 Basic Architecture of dilated_Conv+LReLU Layer

In this Chapter dilated convolution operators are used with the same filter size but different dilation factor. The kernel weight of the convolution layer is increased by increasing the dilation factor. This leads to an increase in the receptive field without any loss in resolution. The output receptive field of a pixel p represents a set of pixels. The receptive field grows linearly when stride is 1 whereas it grows exponentially when s>1. The drawback of dilation convolution is that it may cause some blind spots if it is used directly in SR applications [96]. So, in this Chapter, we have slightly increased the dilation factor (Figure 4.3) from standard convolution to dilation convolution. This results in a more accurate correlation between LR and HR images. This recursive dilated_Conv+LReLU layer is formulated as equation (4.5):

$$[k_s * x, \alpha, k_s *_2 x, \alpha, k_s *_3 x, \alpha, k_s *_2 x, \alpha, k_s * x] \quad (4.5)$$

Each recursive dilated_Conv+LReLU layer in cascading architecture consists of many iterations. The output of each cascaded network is feedbacked to minimize the losses. The loss function used here is described in the next section. The minimization of loss is performed by an ADAM optimizer. The model is in the recursive stage until the minimum loss is not achieved. After reaching minimum loss, the output features are de-convoluted and passed on to the next recursive CNN layer. In this way, a low-resolution image is magnified up to 8x. As the network is lightweight because it reduces network size which increased the speed of the training. Along with speed the model also uses the shared parameters between a cascaded recursive network that ultimately achieves high-quality images due to the application of recursive iterations.

4.4.2 Architecture of MPRDNCF

The MPRDNCF model used recursive cascading of dilated_Conv+LReLU module as a generator of higher magnification image, the network architecture consists of 33 layers that take a low-resolution image as input and super-resolution at a magnification of 8x. The progressive architecture upscales the images with recursive cascading of the dilated_Conv+LReLU module and each module upscales the image only with a magnification factor of 2x which is shown in Figure 4.4. The recursive module minimizes the losses at each cascading layer. The minimized parameters are shared among recursive modules and optimal parameters are passed from one cascaded layer to another. This increased the capacity of the super-resolution network and balance the computational complexity.

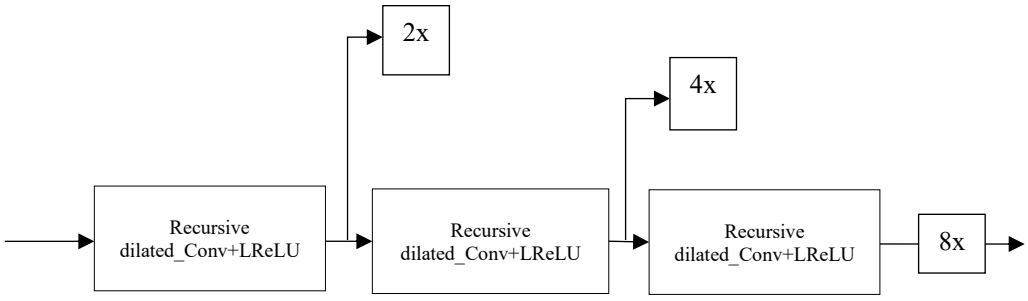


Figure 4.4 Architecture of MPRDNCF

In the proposed residual learning process three dilated_Conv+LReLU blocks are used. The model is based on state-of-art of ensuring convolution of LR sub-pixels to HR sub-pixels. After processing the last block, the sup-pixel convolution layer is applied on HR

feature maps to one single HR image. The memory usage complexity was reduced by decreasing the number of filters from 128 to 64 with increasing network capability. For all CNNs, the kernel size is set to be 3×3 and stride is taken to be 1. Deconvolution layer is applied to upscale the feature maps at the output of each MPRDNCF module. For training the proposed model, residual images are constructed that are trained. The Gaussian kernel is used to smooth the input image and these smooth images are subtracted from the original image to generate a residual image with pixel information. All the network is trained by using the loss function as stated in equation (4.6):

$$loss_{\text{MPRDNCF}} = loss_{mse} + loss_{ssim} \tag{4.6}$$

The loss is calculated between HR images of different resolutions, i.e. 2x, 4x, and 8x, with upscaled LR images.

MSE loss is generally used for regression tasks. This is evaluated by finding the mean of the squared difference between the target and reconstructed value which is shown in equation (4.7):

$$loss_{mse} = \frac{\sum_{i=1}^{N}((x_t - x_r)^2}{N} \tag{4.7}$$

Where, x_t=target value, X_r=reconstructed value

N= Number of samples. Similarly, SSIM is used for the comparisons between two images' output HR image and the original image and to determine the difference precisely.

$$SSIM_{loss} = \frac{1}{2}(1 - E_x[SSIM(I_{LR}, \hat{I}_{HR})]) \tag{4.8}$$

Where, SSIM= Structural Similarity

Ex [·] = Expectation operator over the spatial coordinates, I_{LR} , $\hat{I}_{HR}$ = LR image & reconstructed image respectively.

4.4.3 Training Details

The training process is performed in two steps: in the first step the model learns the parameters for each module in the second step, these parameters are used for learning. In this learning process, LR and HR images are required as a reference as these reference images are used for effective learning of the model. For training, the input 'N' LR images, $[I_{LR}]_n$, and HR image$[I_{HR}]_n$, are taken and simple bicubic downsampling is performed and a degraded image is formed, $[I_D]_i$. After downsampling, the image set, $\{[I_{HR}]_n, [I_{LR}]_i, [I_D]_i\}_{i=1}^{n}$, are fed into a model for training. In this Chapter, we have used

residual learning for super-resolution. To train the MPRDNCF network, 200 training images are taken from BSD100 [99] and CelebA [94] datasets. The training sample taken here is converted into image patches of size 41×41. The input image is taken by applying down sampling on HR images. The HR images are considered to be labeled data. Here, we have applied the recursive residual learning despite prediction. The filter size was initially set to 128 and slightly decreased to 32. For testing of MPRDNCF network, 'Set5' [100], 'Set14' [101] 'Urban100' [102], and 'BSD100' [99], CelebA [94] are considered. A comparative analysis of the training parameters of SR techniques is illustrated in Table 4.1.

Table 4.1 Comparative Analysis of Different CNN Architectures Used for Image Super Resolution

Ref	Base Model	No. of Layers	Activation Function	Learning Rate	Batch size	Epoch	Data Set	Image Size	Loss Function
[88]	VDSR	20	ReLU	0.0003	10	400	Neuroimaging dataset	256×256	L1 loss
[89]	SRNet	20	ReLU	-	-	100	-	33×33	MSE
[90]	GAN	20	ReLU	1e-5	-	-	Cytopathological	256×256	MSE, GAN and Perceptual Loss
[91]	GAN	12	-	0.0002	64	200	CelebA	256×256	Adversarial, Perceptual and Pixel Loss
[92]	VGG-16	5	ReLU	1e-3	64	100	CVC14	128×128	Low frequency loss
[93]	SRNet	3-5	ReLU	0.0002	-	-	ROSIS AVIRIS Cuprite	64×64	L1-norm L2-norm
[104]	SRNet	10	PReLU	0.0001	16	100	CASIA Web Face and CelebA	256×256	Reconstruction, Perceptual, Landmark and

									Adversarial Loss
[105]	Recursive CNN	50	-	0.001	32	100	Set5, Set14, B100, Urban100 and Manga109.	256×256	MSE
[106]	Residual	20	ReLU	0.02	64	-	Set5, Set14, BSD100, Urban100	256×256	MSE
MPRDNCF	Recursive and Residual CNN	33	LReLU	0.0003	32	100	Set5, Set14, BSD100,	256×256	MSE, SSIM loss

4.5 Results and Discussions

This section deals with the validation of model simulation experiments. A brief description of a database and performance measurement is given in this section. The experimental setup which was used for evaluation is also discussed. Eventually, empirical findings are discussed to illustrate the effect of conceptual model contributions as well as to investigate the overall system's robustness.

4.5.1 Performance Metrics

4.5.1.1 Peak Signal-to-Noise Ratio (PSNR)

The very important tool to enhance image quality by denoising the image is known as PSNR. The greater the value of PSNR better the denoising is and the formula of the PSNR is given in Equation (4.9).

$$PSNR = 20log_{10}\left(\frac{Max_f}{\sqrt{MSE}}\right) \tag{4.9}$$

Where, $Max_{f=}$ maximum pixel value, MSE = Mean Square Error.

4.5.2. Structural Similarity Index Measurement (SSIM)

One of the quality assessment methodologies in digital image processing is similarity index measurement. it is used to determine the comparison between images as well as videos. This index is completely reference-based and determines the noisy or compressed or distorted images with some reference images. Therefore, to improve the performance

of traditional methods such as peak signal-to-noise ratio (PSNR) and mean squared error (MSE), SSIM is designed. The calculation of the structural similarity formula with different windows for measuring x_1 and x_2 of the same size N x N is shown in Equation (4.10):

$$SSIM\ (x1, x2) = \frac{(2\mu_1\mu_2 + k_1)(2\sigma_{12} + k_2)}{(\mu_1^2 + \mu_2^2 + k_1)(\sigma_1^2 + \sigma_2^2 + k_2)} \tag{4.10}$$

Where, μ_1= mean of x, μ_2 = mean of, σ_{x1}^2 = variance of x1, σ_{x2}^2= variance of x2, σ_{x1x2}= co-variance of x1 and x2, k_1 and k_2 are variables to stabilize the separation with a weak denominator.

4.5.3 Implementation Details

The proposed network is designed for image super-resolution with downsampling x/2, x/4, and x/8. The input image patch size is 41 which is used to train the model. The images are degraded using a bicubic down sampler by x/8 during training. The depth of the network is 23 to extract spatial information extracted out of images. For residual learning, the loss function is used to train the model, termed $loss_{MPRDNCF}$. Xavier initialization is used to initialize weights and biases for the deep networks. The minimum batch size of the training module was 32. The training is done for the 300 epochs. For simulation modeling, the MatConvNet package is used to train the MATLAB (R2020a) environment to train the model. The system configuration is Intel(R) Core (TM) i5 with CPU 1.60 and 2.11GHz with 2GB Graphics.

4.5.4 Dilated Convolution Analysis

In this section, we have analyzed the impact of using recursive dilated CNN architecture. The model is trained with 200 images with the same configuration, The Figure. 4.4 shows the correlation analysis of input LR and output HR varying with many epochs. The LR image patch is represented as L_i, as well as HR image patch is represented as H_i. Their correlation is represented in Equation (4.11):

$$C(P) = \frac{\sum_{p=1}^{P} covariance(L_p, H_p)}{\sum_{p=1}^{P} \sqrt{covariance(L_p, L_p) covariance(H_p, H_p)}} \tag{4.11}$$

Where, L_p= LR image patch and H_p= HR image patch, C = correlation between LR and HR image patches.

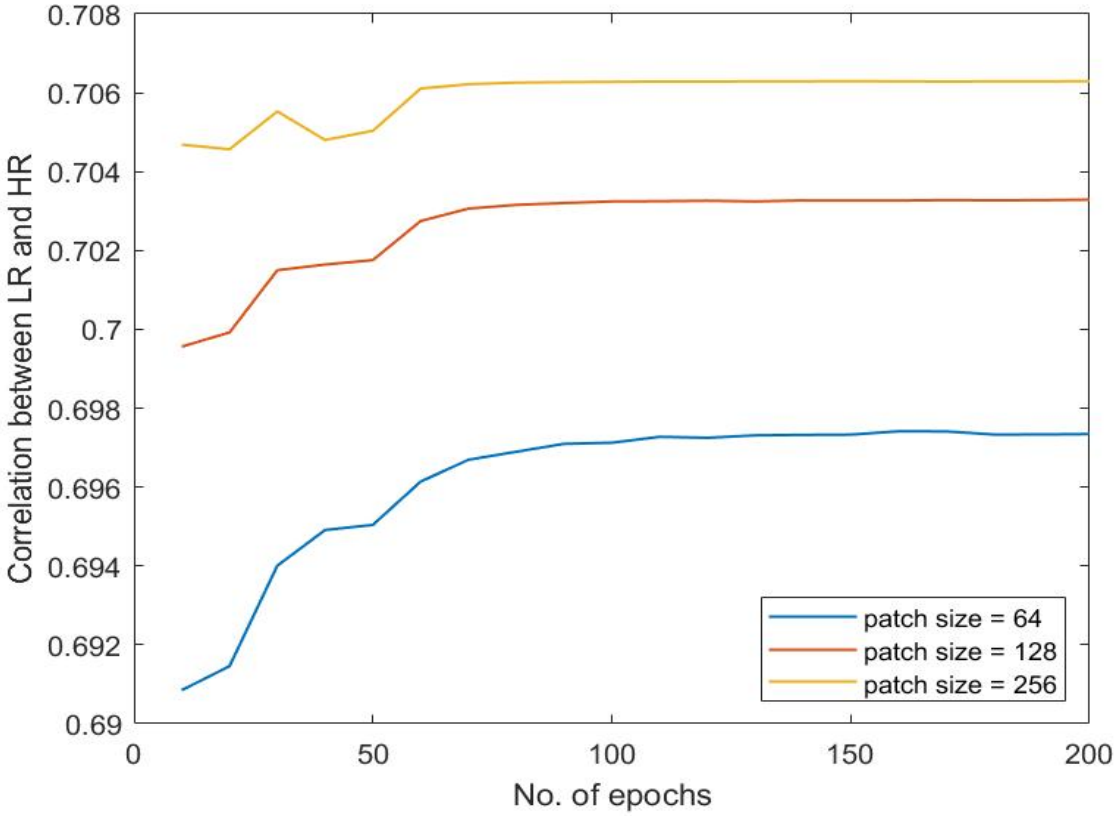


Figure 4.4 Correlation between LR and HR images with varying epochs

This represents the correlation maps between reconstructed HR and LR images with different patch sizes. From Figure, it has been noted that as the patch size is increased, its correlation is increased.

4.6 Performance Analysis

In this section, we have presented the performance analysis of the proposed MPRDNCF model. The model is tested on 5 datasets as discussed above. In Table 4.2, we have presented the PSNR and SSIM evaluation on input image for scale factors 2x, 4x, and 8x respectively. As the model is designed with recursive nature and internal parameters are preserved while cascading which shows the improvement in the result of the MPRDNCF model. It can be said that by introducing the dilated recursive model, pixel information is preserved which can be proved by achieving improved PSNR and SSIM values.

Table 4.2 PSNR and SSIM Evaluation of MPRDNCF

	PSNR				**SSIM**		
	s=2	s=4	s=8		s=2	s=4	s=8
SET5	43.68	35.98	30.42	SET5	0.9864	0.9349	0.8487
SET14	36.08	29.6	25.33	SET14	0.9425	0.8192	0.663
BSD100	36.22	29.6	26.04	BSD100	0.9531	0.8021	0.6523
URBAN100	35.47	27.92	23.16	URBAN100	0.9603	0.823	0.616
CELEBA	41.46	40.51	33.96	CELEBA	0.9975	0.9682	0.9077

4.7 Comparison with State-of-the-Arts

We compare our model with different state of art methods for image super-resolution. We compare our model with other SR models, i.e., Fast Super-Resolution Convolutional Neural Networks FSRCNN [85], Super-Sampling Networks (SSNet) [95], Dilated Convolutions for Single Image Super-Resolution (DCSR) [96], Very Deep Super Resolution network (VDSR) [93], Deeply Recursive Convolutional Network (DRCN) [94] and Deep Recursive Residual Network (DRRN) [95]. For comparison, we have trained our dilated_Conv + LReLU module (Figure. 4.3) individually for scale factors 2, 3, and 4. The input image size of 540 × 360 is taken as a testing image as stated in [15]. The proposed model is a lightweight and recursive progressive network structure that can be applied for a broad range of applications in the real world. Table 3 illustrates the comparative state-of-art. As compared to others, we have adopted the deep learning methodology for multiple training strategies. Among the above-mentioned state of art techniques DCSR [96], VDSR [88] and DRRN [95] had proved their achievements in recent years. In this Chapter, we have used two performance metrics such as PSNR and SSIM as these metrics show the similarity structure between ground truth and reconstructed image. As shown in Table 4.3, MPRDNCF had showed up its improvement over other existing techniques. Figure. 4.5 show some reconstructed SR images of different images. The Figure showed that MPRDNCF had recovered more accurate results while preserving edges and textures.

Table 4.3 Comparative State-of-Arts

	Scale	**Set5**	**Set14**	**URBAN100**	**BSD100**
Bicubic	2	33.65/0.93	30.34/0.87	26.88/0.841	29.56/0.844
FSRCNN [5]		37.00/0.956	32.63/0.908	29.88/0.902	31.58/0.890
SSNet [15]		37.51/0.958	33.19/0.912	30.88/0.914	31.79.0.894
DCSR [16]		37.54/0.958	33.14/0.9141	30.76/0.9142	31.90/0/8959
VDSR [23]		37.53/0.958	33.03/0.912	30.76/0.914	31.90/0.896
DRCN [24]		37.63/0.957	33.04/0.912	30.75/0.913	31.85/0.894
DRRN [25]		37.74/0.959	33.23/0.914	31.23/0.919	32.05/0.897
MPRDNCF		42.06/ 0.9820	35.69/ 0.9501	34.25/ 0.9552	34.80/ 0.9384
	Scale	**Set5**	**Set14**	**URBAN100**	**BSD100**
Bicubic	3	30.39/0.8682	27.55/0.7742	24.46/0.7349	27.21/0.7385
FSRCNN [5]		33.16/0.914	29.43/0.824	26.43/0.808	28.53/0.791
SSNet [15]		33.89/0.922	29.83/0.830	27.09/0.826	28.69/0.795
DCSR [16]		33.94/0.9234	30.28/0.8354	27.24/0.8308	28.86/0.7985
VDSR [23]		33.66/0.9213	29.77/0.8314	27.14/0.8279	28.82/0.7976
DRCN [24]		33.82/0.9226	29.76/0.8311	27.15/0.8276	28.80/0.7963
DRRN [25]		34.03/0.924	29.96/0.835	27.53/0.838	28.95/0.8
MPRDNCF		37.25/0.9520	31.45/0.8839	29.45/0.8784	30.95/0.8543
	Scale	**Set5**	**Set14**	**URBAN100**	**BSD100**
Bicubic	4	28.42/0.810	26.00/0.702	23.14/0.657	25.96/0.667
FSRCNN [5]		30.71/0.865	27.59/0.756	24.61/0.727	26.98/0.715
SSNet [15]		31.54/0.884	28.01/0.764	25.08/0.746	27.08/0.738
DCSR [16]		31.58/0.887	28.21/0.771	25.30/0.757	27.32/0.726
VDSR [23]		31.35/0.883	28.01/0.767	25.18/0.752	27.29/0.725
DRCN [24]		31.53/0.885	28.02/0.767	25.14/0.751	27.23/0.723
DRRN [25]		31.68/0.889	28.21/0.772	25.44/0.764	27.38/0.728
MPRDNCF		34.91/0.923	29.05/0.818	26.88/ 0.796	28.97/ 0.782

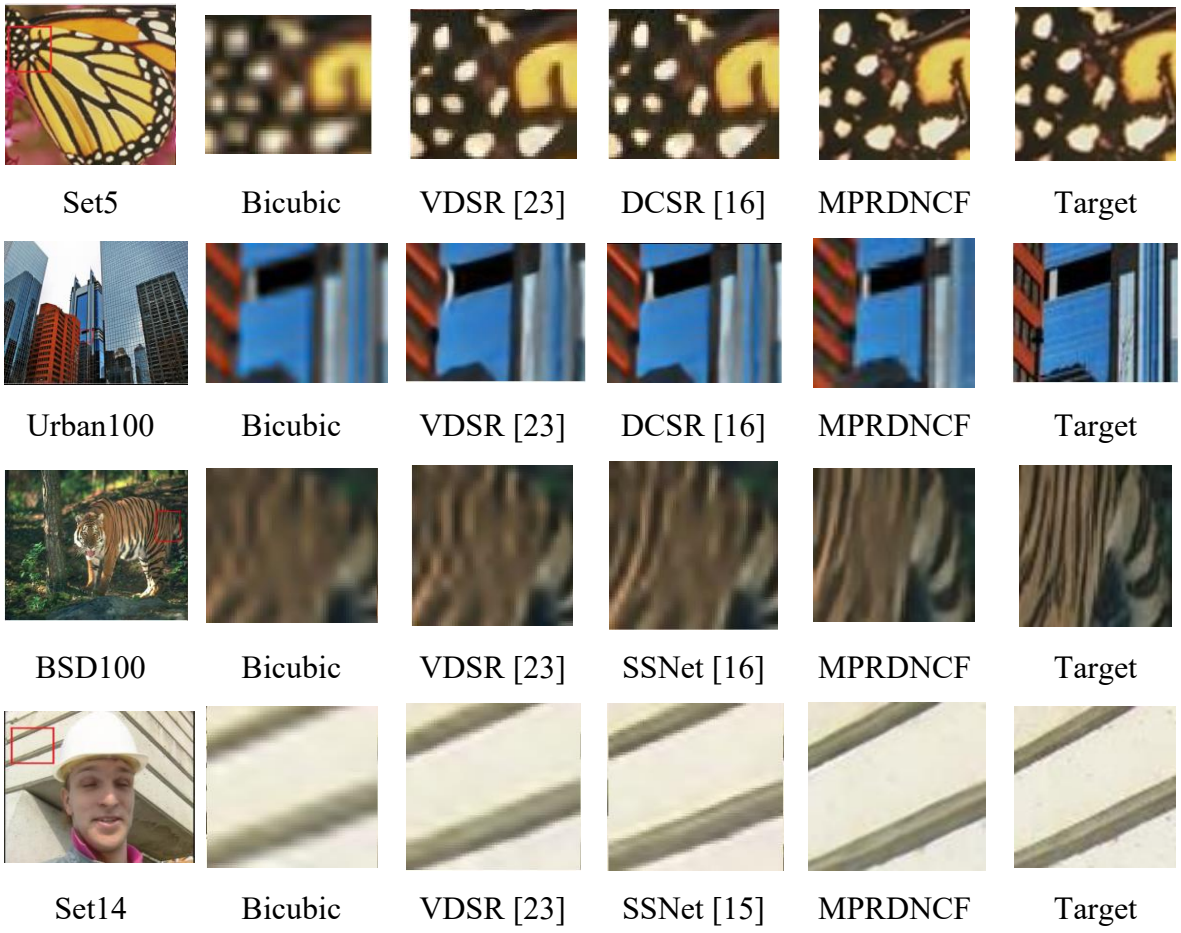


Figure 4.5 Super Resolution Results on Set5, Set14, BSD100 and Urban100 images with scale factor 4

4.8 Run Time Efficiency

It is defined as the ration of total useful work (input image) trained w.r.to the total time which is required for the training. Here, the input image used for the training is 200 images and the total time is 205 minutes for the completion of the training of image.

Run Time Effieciney = (Useful Work / Total Time) × 100%

$$= (200/205) \times 100\ \% \ = 97.50\%$$

The Run time efficiency of Multilevel Progressive Recursive Dilated Networks with Correlation Filter (MPRDNCF) is 97.50 %

4.9 Summary

In this chapter, we use dilated convolution, which makes the receptive field bigger without losing any pixel information. The dilated convolution is set up as a recursive residual network, which keeps the parameters. So, the model is called Multilevel

Progressive Recursive Dilated Networks with Correlation Filter (MPRDNCF), and it uses a progressive approach with cascading recursive residual networks to increase the size of an image. This module scales it up using different scaling factors and uses a deconvolution layer to make it bigger. During the training of the network, the MPRDNCF model used a method called "recursive residual learning," which shares information between the convolution layer and the identity layer. MPRDNCF is built out of 33 layers of CNN. We have shown the analysis of the results on Set5, Set14, Urban100, and BSD100 and having the run time efficiency of 97.50%

CHAPTER 5: FACIAL IMAGE SUPER-RESOLUTION USING PROGRESSIVE NETWORK INTERLEAVED CORRELATION FILTER.

5.1 Introduction

Face image super-resolution is a specific domain of image super-resolution and it is used to convert low-resolution face images into high-resolution face images. Face image super-resolution has many applications related to face vision tasks such as face detection, surveillance, and face editing, where the quality of images is degraded by the input of low-resolution images [107]. The solution for the face image problems is extremely large and most of the techniques try to produce better SR results by learning high-frequency information using relationships between HR and LR images from a training dataset. According to empirical studies [108], present algorithms require a facial image resolution of at least 32 * 32 to 64 * 64. Under the VLR problem, recognition performance will significantly deteriorate [107]. This is because, as can be seen in Figure 5.1, the VLR face image has very little information and many image details have been lost. Recovering the face image's lost features could be a solution to the VLR issue. Super-resolution (SR) methods are therefore used [109–111]. Learning-based SR techniques are always utilized to improve the resolution of the face images due to the self-similarity of the face images. Significant progress has been made by the different techniques in the area of learning-based face image super-resolution in recent years [112]-[116]. The learning of facial image features is still an unsolved problem and degradation in reconstructed high-resolution face images is mainly due to the problem of solution space growing exponentially for facial images, total variation, smoothness, and perceptual quality. The limitation of the prior techniques in the recent research is as follow -Perceptual quality is required for a tiny image (24x24), Occlusion in the image creates problem during reconstruction, High-frequency image like hair reconstruction is a challenging task, and Significant pose variation [115][116]. Improving the problem that has been discussed above gives a promising result that will improve the quality of the face image. For this image super-resolution concept was adopted by researchers [117]-[131] that relies on a single model for large upscaling. However, these approaches do show any effective results on face hallucination problems.

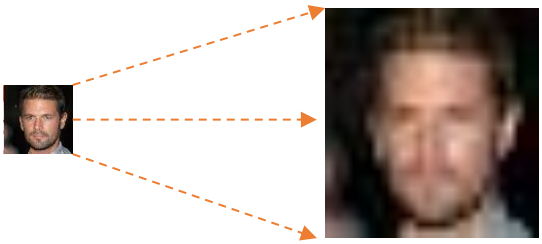


Figure 5.1 Example of VLR in Face Super-Resolution

In this Chapter, we tried to resolve the problem of tiny image super-resolution of 24*24 to achieve the highest magnification factor of 8x by improving its perceptual quality using a deep convolution neural network (CNN). One of the most effective network models for super-resolution models is residual learning [127][128]. Reconstruction parameter optimization is easy in residual learning that reduces the training errors with increased depth whereas computational complexity is also decreased. Progressive learning was adopted with the residual network in [107]. This approach reduces the training loss as previous layer parameters are fed forward to the next layer. However, it was observed in [128] that performance degrades with pose variations. Then, patch-based residual learning [112] was introduced in Collage Net. This increases the performance as patch similarity was evaluated. However high storage memory was required in this model. Then GAN network was introduced in [34] for face super-resolution but it works well with large scaling factors of single image super-resolution. To reduce these issues [114] designed collaborative learning by cascading upscaling modules. Therefore, motivated by these existing models' advantages, the designed model FHSR model is based on a progressive residual learning loss-aware cascaded upscaling model.

Our Model is known as Progressive Upscaling Face Hallucination Super-resolution (FHSR). The upscaling module is designed as a loss-aware model. The FHSR model is designed on a feed-forward convolution neural network (CNN) using progressive residual learning. The model consists of three CNN loss-aware upscaling network (UN) blocks.

The UN modules are designed based on convolution architecture that is based on residual learning. The output of CNN in residual learning is $\hat{y}$ rather than upscaling factor 'k'. This residual image $\hat{y}$ is subtracted from the LR image y to upscale LR to the HR image. It was observed that with increasing CNN depth there is a degradation in the performance of CNN. So, to resolve these issues, residual learning is used. It is observed that residual learning can be learned more easily than the original unreferenced mapping. In each UN

module first of all the input image is degraded using a bicubic down sampler. In image super-resolution, down-sampling is performed over pixels that reduces the spatial resolution preserving the pixel representation. This helps in the reduction of the storage requirement of images. Mathematically, it is represented as in Equation (5.1):

$$y = x \downarrow k \tag{5.1}$$

Where, x = input image

y= LR image

↓= downsampling

k= down-sampling factor

Then LR image (y) is decomposed into feature maps concatenated together and fed into CNN modules. By doing such, CNN can handle the three-dimensional input image. However, many of the existing techniques do not recover the complete information of low-resolution images. So, they are not being able to generate an image nearer to the real image. The local face hallucination is obtained after each block that maps the LR features to the HR feature. Upscaling these local feature maps generates the global super-resolution image. The loss-aware module will pass the losses of the previous layer to the next layer this will reduce the reconstruction loss. Given input facial low-resolution image, $I_{F\text{-}LR}$ generates output facial high-resolution image, $I_{F\text{-}HR}$. For this FHSR algorithm generates the corresponding HR images out of HR feature maps generated iteratively. This can be formulated as Equation (5.2):

$$I_{\mathrm{F-HR}} = F(\alpha, I_{\mathrm{F-LR}}) \tag{5.2}$$

Where, F = Face hallucination function (FHSR), α = learning parameters.

The FHSR used the joint loss function, mean square error (MSE) loss, and structural similarity (SSIM) loss function, termed (MSE-SSIM loss) for the training procedure. The network model is designed and termed as upscaling network modules that are based on residual learning CNN models. The UN modules consist of 33 CNN layers that take input facial images of size 24x24. The modules take these LR images and perform super-resolution with an 8x magnification factor. The UN modules are taken to progressively generate an upscaled image. After every module loss function is fed-forward to the next UN module to handle the losses that occurred during the training process. The results of these loss-aware UN modules are cascaded together to generate conditions for the bottom-

up learning process (from weak learner to strong learner) that gives better convergence towards results as compared to single learning.

5.1.1 Motivation

As the face is one of the most frequently utilized biometric cues, it is highly desirable to obtain face images of high quality to provide essential information for identity verification. However, in real-world situations, the faces that are caught can not only have a low resolution because of the great distance between them. Because of the poor resolution, it is difficult, if not hard, to extract useful knowledge to enable downstream applications like facial attribute categorization and face verification. These applications require a higher level of detail than the low resolution provides. As a result, the development of more sophisticated techniques to imagine LR images has become an absolute necessity [26]-[35]. Recently, deep learning models [36]-[42] have shown their effectiveness for image super-resolution. However, creating images with a high resolution still results in images with a low level of identity similarity. Recently, as a result of the fast development of the CNN (convolutional neural network) in image processing, an increasing number of researchers are utilizing CNN to train an end-to-end model for face SR. CNN demonstrates a high level of performance in a variety of image-processing tasks, including image classification and segmentation. Similar to image SR, several efforts can increase the quality of SR by making use of a large number of datasets. However, simply utilizing a larger amount of data for the training of the model is not enough. There is certain background information associated with facial images that prevent them from being used for general image SR. The difficulties associated with face super-resolution stem from the fact that the previous facial information is not available, and it is impossible to directly learn from the LR facial image. Hence, our model introduces progressive upscaling face hallucination super-resolution (FHSR) which is a progressive reconstructive network. Generally, most of the state of art analysis gives the output of a high-resolution image in a single shot but in the case of higher magnification like 8x, it is a tough task to generate appropriate results. To overcome the problem this network is designed to increase the perceptual quality of the image during reconstruction and give clear information about the identity of the image nearer to the ground truth image. The resolution is performed in steps of 2x magnification and three loss-aware upscaling modules are used to achieve a better magnification of 8x as compared to other state of art techniques.

5.1.2 Key Contribution and Chapter Organizations

The key contributions of this Chapter are as follows:

- In this Chapter, state-of-the-art about-face hallucination techniques and face super-resolution frameworks are presented along with their challenges.
- This Chapter discusses the upscaling face hallucination architecture which super resolves the very tiny image of 24*24.
- The Chapter designed novel loss-aware upscaling layers that are cascaded together and progressive. The loss-aware layers will feed forward the losses to the next layers. This will minimize the reconstruction loss. The layers used the combined loss function.
- The loss-aware upscaling module introduced a correlation filter at the end of the residual network. This will help in removing the blind spot generation.
- Therefore, the designed FHSR model resolves the issue of blind spots, checkboard, identity information loss, and fine texture detail, reduces the network complexity, and also increases the receptive field.
- The Chapter also presented the ablation study of the baseline model with designed loss functions.
- In last this Chapter gives a comparative state-of-art with existing techniques.

5.2 Related Work

The combination of a layer with residual learning and skip connection are used from techniques [22] that use shared parameters for the constant information of the trained images which helps in an ensemble of the image at the reconstruction end of the image. Then, further problems of face hallucination were addressed by [107] using the progressive upscaling model. The cascaded convolution neural network model was adopted for face hallucination that used identity priors while training. In [142]-[145] SR models are cascaded and upscaling was performed accordingly. The result obtained shows with the noisy image there are reconstruction losses. Along with that pose, variation degrades the performance. The model compared the result with the SRGAN [34] model. Collage Net [112][146], a face super-resolution network is designed that learns the multi-scale feature information hierarchically.

After that compute the patch-wise similarities between the input and reference feature pyramid then select the K most similar reference patches to each input patch. The limitation of this model is that it is less efficient as the patch grouping demands high

memory and computational complexities [147]. A residual generative adversarial network [34] is proposed to restore the HR face image with proper identification from the LR face image. After training the network, the LR image and its semantic information are used which is trained on the same-identity face pair of the LR face image and HR face image [148]. One limitation is model has a high training time. In [34] author adopted GAN as a base model. This model works on the loss function of the GAN model and reduces the reconstruction loss but for large upscaling factors (4x). From these models, it was concluded that single-image super-resolution methods do not apply to real-world applications due to their high computational complexity. Then to resolve such issues, the author in [133] presented a cascaded architecture of the residual network. In [134], the author stated face super-resolution is a generative task and facial features contain identity information. It was noticed that previous face hallucination approaches largely ignored facial identity recovery. This model introduced the identity information of faces but was also a single model. In this model, the learning rate was high which resulted in increased training losses. In [135] author introduced the identity loss function for face super-resolution.

In [137] author proposed a GAN network for image super-resolution and proposed perceptual loss for improvement of network performance. But this model outperforms better when the input patch image size is greater than 96x96. In [138] author used GAN architecture for face hallucination and worked on loss function but in this model, image quality was not preserved. In [139] used the VGG network for image super-resolution. In this work, they worked for 4x scaling factors only and didn't work on tiny image patches. Another contribution was presented as LapSRN [143]. This model was designed as a pyramid architecture with a ReLU activation function. The drawback of the ReLU network is that it doesn't perform for a negative slope. The model was also not adopted for tiny image patches. This model was also not adopted for noise artifacts. These drawbacks are removed in our proposed model. Table 5.1 represents the comparative features of existing models and our model.

Table 5.1 Comparative Analysis of Different CNN Architectures Used for Image Super-Resolution

Ref	BM	L	AF	LR	BS	E	II	LF	SISR
[107]	SRNet	53	LReLU	0.0003	8	80	24 × 24	SSIM	No

[106]	ResNet	-	ReLU	0.0001	9	100	-	Perceptual+ adversarial	No
[34]	GAN		ReLU	0.0001	16	-	-	L2 loss	Yes
[131]	GAN	-	ReLU	0.0001	-	100	-	Perceptual	Yes
[134]	ResNet	-	-	0.0001	16	-	-	L1 loss	Yes
[135]	ResNet	-	PReLU	0.02	128	-	24×28	Identity loss	Yes
[137]	GAN	32	ReLU	0.0002	16	300	96 × 96	perceptual	Yes
[143]	Recursive Block	-	ReLU	0.001	64	80	-	L2 loss	Yes

BM= Base Model, L= Number of layers, AF= Activation function, LR= Learning rate, BS= Batch Size, E= Epochs, II=Input Image size, LF=Loss Function, NH= Noise Handling, SISR= Single image super-resolution

5.3 Methodology

5.3.1 Novelty of Work

In this section, the novelty of this work is the proposed progressive network with residual learning, novel loss-aware upscaling network with correlation filters, and combined loss function. The Chapter has presented ablation study models with and without UNCF layers with different loss functions. The Chapter also presented the performance of the proposed progressive FHSR with the UNCF layer and combined loss function with image augmentation (rotation).

5.3.2 Overview of Progressive FHSR Network

FHSR network consists of the progressive network which up-samples face image in steps of 2x magnification, to achieve 8x magnification. As shown in Figure 5.2, it consists of multiple loss-aware upscaling networks with a correlation filter module (UNCF). The UNCF module is used as a feature extraction block which is used for mapping and though it extracts the feature of the image in low-resolution space after that, it feed-forward to the up-scaling layer which is done by a deconvolution layer and passed to the next stage of UNCF module which is used for the refining of image and upscale to the next level. With features, losses of this module are also passed to the next layer. Therefore, different parts of the progressive FHSR network are defined in the below sub-sections.

5.3.2.1 Upscaling Network Layer Architecture

The loss-aware UNCF layer is cascaded to progressively perform super-resolution from 2x to 8x upscaling factors. Figure 5.3 presents the architectural diagram of UNCF. The UNCF layer is cascaded together with convolution and dilated convolution layers with Leaky ReLU (LReLU) activation function. The dilated convolution layer is interleaved with the convolution layer to increase the receptive area from the input image. This will result in the creation of some blind spots. Therefore, to resolve this issue, the work designed a model, which slightly increases the kernel weight of the convolution layer by increasing the dilation factor. This leads to a linear increase in the receptive field without any loss in resolution. Let x_i be the input to UNCF denoted as H × W × C channel. The non-linear mapping between UNCF is represented as $F(.)$ Consists of 3×3 conv and dilated convolution layers. The UNCF layer is mathematically represented as:

$$x_{i+1} = [kw_t *_n A_{R_i}, CF_i] \tag{5.3}$$

Where, kw_t = kernel weight

$*_n$ = dilation factor

A_{R_i} = receptive area

CF_i= Correlation filters

x_{i+1}= output of UNCF layer

UNCF performs channel-wise concatenation that non-linearly maps inputs, x_i, to output, x_{i+1}, as normal residual blocks.

The output from the UN layers UN1, UN2, and UN3 is up-sampled together to 8x of 24 × 24 input images. The output after each UN layer is compared with the ground truth. This process improves the feature at each layer and results in better reconstruction. Different weights of loss functions are passed to the reconstruction layer according to their importance. These weights contribute to determining the overall losses of cascaded UN layers. The loss function adopted in the FHSR model is termed a combined loss formed by adding MSE loss and SSIM loss. Mathematically, combined loss ($loss_{CL}$) is evaluated as:

$$loss_{CL} = loss_{mse} + loss_{ssim} \tag{5.4}$$

The $loss_{CL}$ was calculated between HR images of different resolutions, i.e. 2x, 4x, and 8x, with upscaled LR images.MSE loss is generally used for regression tasks. This is evaluated by finding the mean of the squared difference between the target and

reconstructed value. The MSE loss is great for ensuring that our trained model has no outlier predictions with huge errors since the MSE puts a larger weight on these errors due to the squaring part of the function. Mathematically it is evaluated as:

$$loss_{mse} = \frac{\sum_{i=1}^{N}((x_t - x_r)^2}{N} \tag{5.5}$$

Where, x_t =Target value, x_r = Reconstructed value, N= Number of samples.

Similarly, SSIM stands for Structural Similarity Index which is used to measure perceptual similarity between two images. SSIM is used for the comparisons between two images (generated image and original image) and to determine the difference precisely. As compared to loss functions such as L2 loss, SSIM loss gives a better correlation among pixels. Mathematically, it is represented as:

$$loss_{SSIM} = \frac{1}{2}(1 - E_x[SSIM(I_{LR}, \hat{I}_{HR})]) \tag{5.6}$$

Where, SSIM= Structural Similarity

Ex [·] = Expectation operator over the spatial coordinates, I_{LR} , $\hat{I}_{HR}$ = LR image & reconstructed image respectively.

For each UN layer, loss weight is determined as:

$$loss_{CL-UN_i} = w_i * loss_{mse_i} + u_i * loss_{SSIM_i} \tag{5.7}$$

Where, w_i = ith UN layer weight for MSE loss which is calculated as:

$$w_i = \frac{loss_{mse}}{(loss_{mse} + loss_{ssim})} \tag{5.8}$$

Where, w_i = ith UN layer weight for MSE loss which is calculated as:

$$u_i = \frac{loss_{SSIM}}{(loss_{mse} + loss_{ssim})} \tag{5.9}$$

1) Progressive UNCF Architecture

In this network, a low-resolution image is refined by three cascaded UNCF layers for the higher magnification of 8x and also shared the parameters to decrease the memory footprints of the network and shared recursively. The architectural details of progressive

UNCF are presented in Table 5.2. FHSR network reduces the size of the architecture which increases the speed of the training. The FHSR model used the UN module as a generator of higher magnification images in the form of cascading, the network architecture consists of 33 layers that take a low-resolution face image as input and super-resolution at a magnification of 8x. This progressive network applies a loss function at an intermediate stage to ensure better control of the training process. We designed the UN module with a mixed convolution network that heavily relies on the sharing of parameters and the sub-pixel convolution layer used for image upscaling. Our network is motivated by the success of FNN models in various problems. [134] [135] and the state of the art is ensured by deconvolution in prior work [136]. Similarly, the residual block of the upscaling network module consists of two convolutions -batch normalization- and activation with the element-wise sum for ensuring the constant memory footprints for all the UN modules and decreasing the number of filters in the convolutional layers by a factor of 2x with each step. This increased the capacity of the super-resolution network and balanced the computational complexity. To upscale, the feature maps at the output for the better reconstruction of an image we rely on the deconvolution layer proposed in [137]. it increases the spatial dimensions of the feature maps and aggregate pixels to form multiple low-resolution feature maps and therefore it reduces the number of feature maps by a 4x factor. After reaching the final stage of the network the target resolution is obtained by passing through the feature maps with the final deconvolution layer with three output channels that resolves 8x output RGB image. In the proposed residual learning process three UN blocks are used. The model is based on state-of-art ensuring convolution of LR sub-pixels to HR sub-pixels. The first UN block consists of a sequence of convolution and LeakyRELU activation layers are used for non-linear mapping. The second and third UN module consists of sequence convolution batch normalization and LeakyReLU [131]. After processing the last block, the sub-pixel convolution layer is applied by the cascading of the HR feature maps to one single HR image. The memory usage complexity was reduced by decreasing the number of filters with increasing network capability. The deconvolution layer is applied to upscale the feature of the face images at the output of each UN module. In Algorithm 5.1, the training strategy is illustrated. The training process is performed in two steps: in the first step the model learns the parameters for each UN module in the second step, these parameters are used for learning. In this learning process, LR and HR images are required as a reference as these reference images are used for the effective learning of the model. For training, the

input 'N' LR images, $[I_{LR}]_n$, and HR image$[I_{HR}]_n$, are taken simple bicubic downsampling is performed and a degraded image is formed, $[I_D]_i$. After downsampling, the image set, $\{[I_{HR}]_n, [I_{LR}]_i, [I_D]_i\}_{i=1}^n$, are fed into the model for training. For training the FHSR model, residual facial images are constructed that are trained with UN modules. The Gaussian kernel is used to smooth the input image and these smooth images are subtracted from the original image to generate a residual image with facial information.

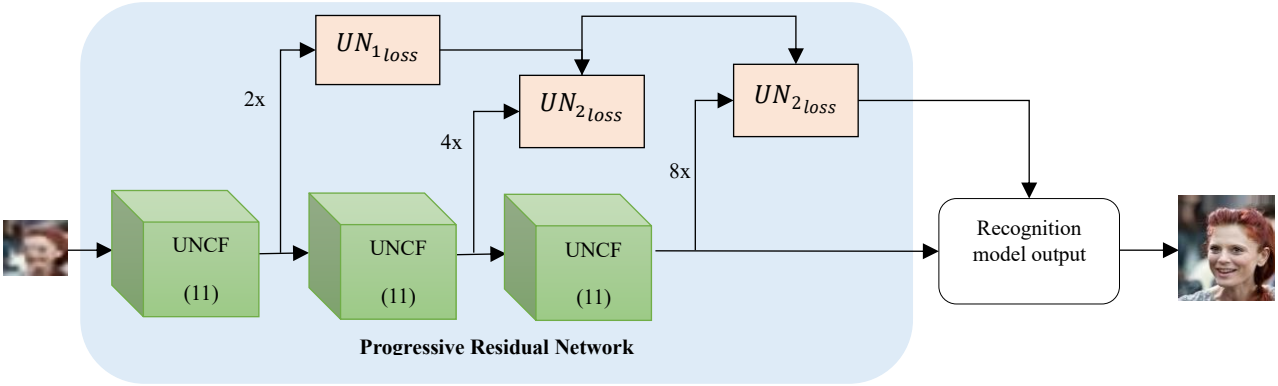


Figure 5.2 Architecture of FHSR

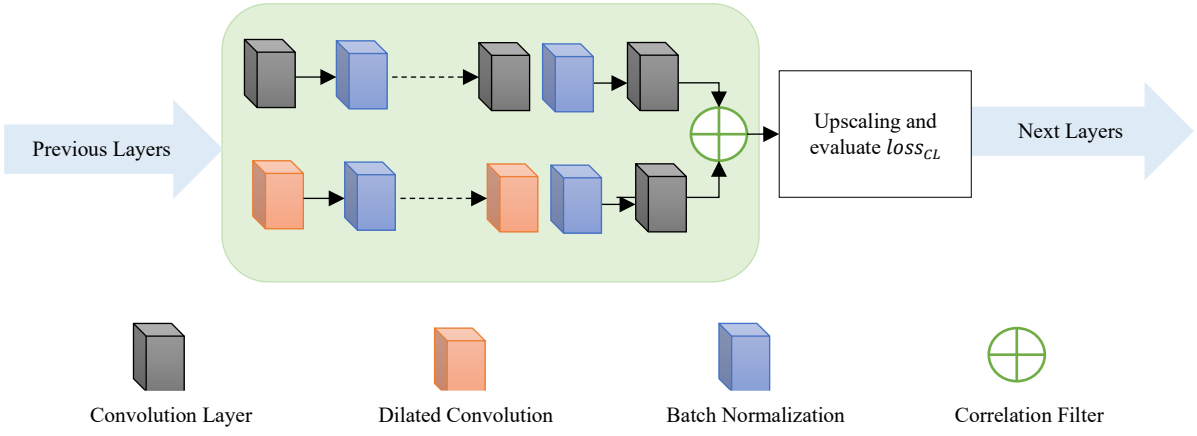


Figure 5.3 Architecture of UNCF Layer

Table 5.2 Description Of network architecture of UN modules

CNN Layer	Learning Modules	Number of Filters	Stride	Output size
1	LR image	-	-	-
2-11	$UNCF$	512	1 1	24×24 24×24
12	Upscaling by Subpixel network (Generate 2x image)	Convolution of feature maps	1	48×48
13-22	$UNCF$	256	1 1	48×48 48×48
23	Upscaling by Subpixel network (Generate 4x image)	Convolution of feature maps	1	96×96
24-32	$UNCF$	128	1 1	96×96 96×96
33	Upscaling by Subpixel network (Generate 8x image)	Convolution of feature maps	1	192×192

Algorithm 5.1: Training of Progressive FHSR

Input: LR_i, LR images;

UNCF Network;

Training dataset, $T_d = \{I_{\mathrm{F-HR}}, I_{\mathrm{F-LR}}\}$,

$$where\ n \in size(T_d)$$

kw_t is kernel weight

$*_n$ is dilation factor

CF_i= correlation filter factor

Output: $I_{\mathrm{F-SR}}$, Super-resolution images.

1: Initialization: $kw_t, *_n, CF_i, epoch$;

2: While $loss_{\mathrm{CL}}$ reaches convergence do

3: $I_{\mathrm{F-HR}} = [kw_t * LR_i, CF_1, kw_t *_2 LR_i, CF_2, kw_t *_3 LR_i, CF_3, kw_t *_2 LR_i, \alpha, k_t * LR_i]$

4: Minimize ($loss_{CL-UN_i}$)

5: end

Return $I_{\mathrm{F-HR}}$

5.3.2.2 Training Details

The FHSR network is designed for face hallucination with down-sampling x/2, x/4, and x/8. The input image patch size is 24×24 which is used to train the model. In this Chapter, images of aligned CelebA datasets are used for training purposes. For testing the trained model, aligned CelebA, Helen, and LFW datasets are used. The images are degraded using a bicubic down sampler by x/8 during training. The depth of the network is 33 to extract spatial information extracted out of images. For residual learning, the loss function is used to train the model, termed $loss_{CL}$. Xavier initialization is used to initialize weights and biases for the deep networks. The minimum batch size of the training module was 64. Adam optimizer for training the models with a learning rate of 10^{-4}. The training is done for the 100 epochs. For simulation modeling, the MatConvNet package is used to train the MATLAB (R2020a) environment to train the model. The system configuration is Intel(R) Core (TM) i5 with CPU 1.60 and 2.11GHz with 2GB Graphics.

5.4 Experiments

This section is dedicated to presenting the validation of simulation experiments for the proposed model. This section gives a brief description of database description and performance metrics. The result's ablation study is also presented as compared to baseline models and elaborates on the performance of the designed FHSR model. Finally, a comparative state-of-art of the proposed progressive FHSR with existing state-of-art models is performed.

5.4.1 Dataset Description

For the testing of images, we have selected four benchmark datasets, i.e., Labeled Faces in the Wild (LFW) [138], CASIA Web Face [139], CelebA [107], and Helen [136]. We use the CASIA web face dataset to get comfortable with the limits of FHSR. The dataset holds a total of 494k images of 10k characters and addresses a medium-estimated dataset proper for learning CNNs for various face-related vision endeavors. Since the dataset images of size, 250x250 pixels are by and large unclearly altered from one place to another on the face, we simply consider the central patches of 192x192 pixels of the images and use these pixels as the purpose behind the examinations. Lastly, we level the images using the Gaussian part (σ = 0.20 x down-inspecting component) and sub-model the images utilizing a bicubic arrangement. At the time of testing, we used the absolute labeled faces in the wild dataset with 13k facial images and 5k subjects similar to images from CelebA, Helen. We have considered the LFW dataset as the first phase since it

features images of variable quality surrounded in unrestricted conditions and likewise addresses a massive test for super-resolution models. Even more fundamentally, it comprises no cover with Casia web face in regards to character, which is vital to ensure a sensible and reasonable evaluation of the FHSR Model. The CelebA & Helen datasets, of course, are picked to test the execution of FHSR on images of unforeseen characteristics in contrast with LFW and, thus, overview the theory capacities of our face hallucination model. We see that the CelebA and Helen datasets contain images of significant standards, yet also critical proportions of JPEG-pressure things. Henceforth, we take the going with actions to preprocess the datasets. Our Model first yields the facial regions using the provided achievement bearings to achieve equivalent produces to the images present in the Casia Web face datasets and LFW dataset. By then, we consider taking the most raised target images from the datasets and down-model them to 192*192 pixels utilizing Gaussian and bicubic development. Our Model at that point treats the subsequent square 192*192 pixel images as the objective high-goal images. With this methodology, we measure a measure of 320 images from Helen and 1k images from CelebA that structure the test set for our examinations.

5.4.2 Ablation Study

To ensure the capabilities of this proposed model series the experiment is done for upscaling the image. The training of the model is described in section III and evaluates the performance of the network. The trained image is examined through each network and step by step it increases the pixel value for the regeneration of the image at the reconstruction end which achieves a better perceptual quality of the image during magnification.

The proposed FHSR (progressive loss-aware upscaling network with correlation filter) increases the pixel information at the reconstruction end. In this section, the following models are designed for ablation study (Table 5.3).

Baseline (Model I): The progressive residual learning is used initially having the combination of Convolution layer, batch normalization, and LeakyReLU activation function. The loss function adopted for the baseline model is $loss_{MSE}$ only at the final layer of the model.

Baseline+SSIM (Model II): In this model, a baseline network is used with $loss_{SSIM}$ only at the final layer.

Progressive FHSR (Model III): The progressive loss-aware UNCF network is used with $loss_{MSE}$ function.

Progressive FHSR (Model IV): The progressive FHSR with a loss-aware UNCF network is used with combined loss ($loss_{CL}$) function.

Table 5.3 Model Characteristics for Ablation Study

Models	Baseline	Loss	CF
Model I	√	MSE	×
Model II	√	SSIM	√
Model III	√	MSE	√
Model IV	√	MSE+SSIM	√

Table 5.4 shows the performance evaluation of the ablation study of UNCF compared with the Baseline Models presented in Table 5.3. The performance was calculated in terms of PSNR and SSIM for the models I to IV. For LFW, the PSNR and SSIM for Model I are 26.021 and 0.756.

For Model II, PSNR and SSIM is 26.414 and 0.770. For model II, PSNR and SSIM is 26.425 and 0.780. For model IV (Progressive FHSR loss-aware UNCF network), the PSNR and SSIM are 27.120 and 0.791. For HELEN, the PSNR and SSIM for model I are 25.950 and 0.761. For model II, the PSNR and SSIM are 26.507 and 0.762.

For Model III, the PSNR and SSIM are 26.660 and 0.764. For model IV (Progressive FHSR with loss-aware UNCF network), the PSNR and SSIM are 27.037 and 0.779. For CELEBA, the PSNR and SSIM for model I are 25.920 and 0.751. For Model II, the PSNR and SSIM are 26.525 and 0.771. For model III, the PSNR and SSIM are 26.313 and 0.775.

For Model IV (Progressive FHSR with loss-aware UNCF network), the PSNR and SSIM are 27.023 and 0.785. It is clear from all the values that in Model IV (progressive FHSR with loss-aware UNCF network), the PSNR and SSIM are maximum.

Table 5 shows the ablation study of Model IV (progressive FHSR with loss-aware UNCF network) with different loss functions. The performances are calculated in terms of PSNR and SSIM for the model having different losses. For LFW, the PSNR and SSIM for perceptual Loss are 25.684 and 0.730. For Function Super, PSNR and SSIM are 26.940 and 0.782. For Identity Loss Adversarial, PSNR and SSIM are 26.162 and 0.752 For Combined Loss, the PSNR and SSIM are 27.231 and 0.802. For HELEN, the PSNR and

SSIM for perceptual Loss are 25.601 and 0.740. For Function Super, the PSNR and SSIM are 26.721 and 0.790. For Identity Loss Adversarial, the PSNR and SSIM are 25.943 and 0.778. For Combined Loss, the PSNR and SSIM are 26.125 and 0.798. For CELEBA, the PSNR and SSIM for perceptual Loss are 25.736 and 0.748. For Function Super, the PSNR and SSIM are 26.501 and 0.769. For Identity Loss Adversarial, the PSNR and SSIM are 25.409 and 0.756. For Combined Loss, the PSNR and SSIM are 26.81 and 0.815.

Table 5.4 Ablation Study of Baseline Models

Model	LFW		HELEN		CELEBA	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Model I	26.021	0.756	25.950	0.761	25.920	0.751
Model II	26.414	0.770	26.507	0.762	26.525	0.771
Model III	26.425	0.780	26.660	0.764	26.313	0.775
Model IV (Progressive FHSR)	27.120	0.791	27.037	0.779	27.023	0.785

Table 5.5 Ablation Study of Model IV with Different Loss Functions

Model	LFW		HELEN		CELEBA	
	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Perceptual Loss Function	25.684	0.730	25.601	0.740	25.736	0.748
Super Identity Loss	26.940	0.782	26.721	0.790	26.501	0.769
Adversarial Loss	26.162	0.752	25.943	0.778	25.409	0.756
Combined Loss	27.231	0.802	26.125	0.798	26.081	0.815

Table 5.6 shows the ablation study of Model IV (progressive FHSR with loss-aware UNCF network) with Rotation Degrees varied from 0° to 90°. The PSNR and SSIM for 0° are 27.231 and 0.814 respectively. For 15°, the PSNR and SSIM are 25.546 and 0.801. For 30°, the PSNR and SSIM are 25.553 and 0.803. For 45°, the PSNR and SSIM are

25.452 and 0.791. For 60°, the PSNR and SSIM are 25.221 and 0.778. For 75°, the PSNR and SSIM are 24.784 and 0.762. For 90°, the PSNR and SSIM are 24.786 and 0.771. The PSNR is maximum for a 30° Rotation angle and SSIM for 0°.

Table 5.6 Ablation Study of Model IV with Rotation Degree

Rotation	PSNR	SSIM
0°	27.231	0.814
15°	25.546	0.801
30°	25.553	0.803
45°	25.452	0.791
60°	25.221	0.778
75°	24.784	0.762
90°	24.786	0.771

Figure 5.4 represents the result of the ablation study, The increase in the pixel value of the image is shown stepwise firstly it is generated by baseline architecture then introducing the SSIM loss function when it is generated using cascaded with MSE and SSIM loss function it super-resolved the image with good perceptual quality multi-scaling is done and gives better results as compared to the targeted image. Whereas Figure 5.4 shows the robustness of the proposed progressive FHSR.

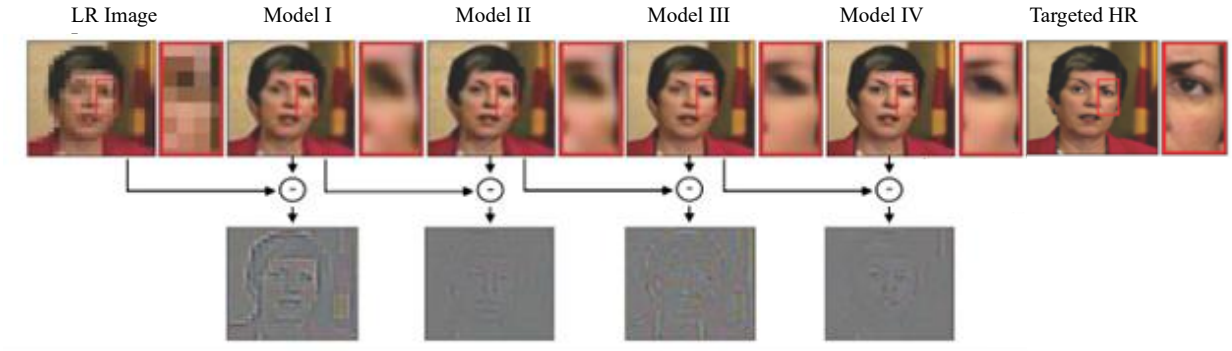


Figure 5.4 Ablation Study on Baseline Models

Figure 5.5 Robustness Visualization of Progressive FHSR (UNCF). The first row represents the generation of the HR image from the 24*24 LR face image. The bottom row represents the original image.

5.4.3 Result Analysis

To analyze the network performance of the proposed face hallucination method, progressive FHSR with UNCF layers with other state-of-the-art methods is presented in this section. LFW, HELEN, and Celeb A dataset images are used for testing the network performances. Table 7 shows the PSNR Analysis of model progressive FHSR with UNCF layers. PSNR is evaluated for LFW, HELEN, and CELEBA. Three upscaling factors are used i.e. 2x,4x, and 8x. In LFW, for 2x PSNR is 31.174. For 4x PSNR is 30.059 and for 8x PSNR is 27.756. In HELEN, 2x PSNR is 31.312 For 4x PSNR is 30.168 and for, 8x 27.741. In CELEBA, 2x PSNR is 31.70. For 4x PSNR is 30.55 and for 8x PSNR is 27.910. Table 8 shows the SSIM analysis of progressive FHSR with UNCF layers. SSIM is evaluated for LFW, HELEN, and CELEBA. In LFW, for 2x SSIM is 0.945 For 4x SSIM is 0.901, and for 8x SSIM is 0.856. In HELEN, 2x SSIM is 0.944 For 4x SSIM is 0.899 and for 8x, SSIM is 0.855. In CELEBA, 2x SSIM is 0.945. For 4x SSIM is 0.903 and for 8x SSIM is 0.861.

5.4.4 Comparisons with State of Art techniques.

Table 7 and Table 8 show PSNR and SSIM comparison with state-of-art models. We compare our progressive FHSR with UNCF layers model with 10 state-of-the-art super-resolution and face hallucination models, i.e., the Super-identity convolutional neural network (SICNN)[135], Very Deep Super Resolution Network (VDSR) [139], the perceptual loss based SR model (Lp) [141], the Cascading Residual Network (CARN) [123], the Deep Laplacian Pyramid Super-Resolution Network (LapSRN) [124], the Super- Resolution Generative Adversarial Network (SRGAN) [26], the Enhanced Deep Residual Network (EDSR) [134] and the Ultra Resolving Discriminative Generative Network (URDGN) [138]. A comparative analysis of performance parameters such as PSNR and SSIM values for scaling factors 2x, 4x, and 8x are presented and illustrated in

Tables 5.7 and 5.8 respectively. It has been observed that the proposed progressive FHSR with the UNCF layers model outperforms better as compared to other methods for face hallucination. Since some of these models were introduced for general super-resolution problems, we re-train all the models on the Casia Web-Face dataset and use open-source codes for their implementations of the authors for a fair comparison. Some of the techniques are used for general image super-resolution issues. However, we tried to overcome all the possible losses while generating high-resolution images. In this experiment, we have reconstructed 98.8% of image pixels from low-resolution images. Initially, the VDSR [140] & SRCNN [141], models have low receptive fields as compared to the proposed progressive FHSR with the UNCF layers model. LapSRN [124] & CARN [125], and C-SRIP [107] have a similar topology to cascading models but still fail to resolve at 8x magnification of facial images.

Table 5.7 PSNR Evaluation on LFW, HELEN, and CELEB A Datasets

Models	**LFW**			**HELEN**			**CELEBA**		
	2x	**4x**	**8x**	**2x**	**4x**	**8x**	**2x**	**4x**	**8x**
Bicubic	27.275	24.824	24.256	27.511	24.951	24.413	26.684	23.951	23.214
VDSR [139]	28.717	25.875	25.285	28.993	25.573	25.176	28.832	25.813	24.241
URDGN [138]	28.663	25.982	25.761	28.732	25.686	26.892	28.772	25.964	25.145
EDSR [134]	29.225	27.514	25.653	29.571	27.962	25.326	30.185	27.296	25.843
SICNN [135]	29.105	27.626	25.862	29.352	27.921	26.867	30.078	26.127	25.926
Lp [141]	29.502	27.731	25.532	29.541	27.984	27.954	30.236	27.125	25.564
SRGAN [131]	30.254	27.831	25.769	31.412	28.055	26.049	30.572	27.463	25.834
ESRGAN [137]	30.265	27.834	26.120	31.250	28.623	26.845	30.672	27.520	26.530
CSRIP [107]	30.831	27.995	27.164	31.271	28.221	27.076	30.891	27.63	26.022
Progressive FHSR	31.174	30.059	27.756	31.312	30.168	27.741	31.70	30.55	27.910

Table 5.8 SSIM Evaluation on LFW, HELEN, and CELEB A Datasets

Models	LFW			HELEN			CELEBA		
	2x	4x	8x	2x	4x	8x	2x	4x	8x
Bicubic	0.877	0.76	0.706	0.883	0.766	0.707	0.870	0.744	0.307
VDSR [139]	0.92	0.81	0.730	0.915	0.781	0.74	0.923	0.813	0.434
URDGN [138]	0.93	0.84	0.750	0.916	0.824	0.792	0.910	0.822	0.742
EDSR [134]	0.91	0.79	0.756	0.912	0.861	0.752	0.932	0.843	0.756
SICNN [135]	0.92	0.89	0.762	0.922	0.833	0.798	0.923	0.846	0.665
Lp [141]	0.93	0.90	0.772	0.924	0.841	0.745	0.926	0.851	0.668
SRGAN [131]	0.94	0.85	0.693	0.941	0.865	0.722	0.941	0.851	0.561
ESRGAN [137]	0.95	0.86	0.72	0.952	0.872	0.745	0.932	0.865	0.625
CSRIP [107]	0.945	0.87	0.812	0.95	0.882	0.821	0.952	0.877	0.636
Progressive FHSR	0.945	0.901	0.856	0.944	0.899	0.855	0.945	0.903	0.861

Figure 5.6 Reconstruction Comparative State-of-Art Performance

CSRIP [1] model uses squeeze net architecture for the training but progressive FHSR with UNCF layers has used its architecture of 33 layers which is shown in Table 5.2. EDSR [134] network is a very complex structured network. As compared with our model we are having 33-layer architecture with less complexity of the network. After that, we compared our model with the Generative Adversarial network GAN [26] and Enhanced

Super Resolution GAN [31], URDGN [138], SICNN [135], and Lp [141] model which provided improved results but they were slightly behind our model progressive FHSR with UNCF layers. We tried to overcome all the issues like, in the GAN network, they can't distinguish between the real/fake image at x = 0.5 and having memory issues during training because of using the batch normalization layer.

It is being overcome by ESRGAN [137] by replacing the batch normalization layer with the RRDB network which is termed a recursive residual dense block. URDGN [138] uses the combination of the deconvolution layer with the summation of loss functions like MSE and adversarial but is still slightly behind our model in comparison with the parametric evaluation of average PSNR and SSIM. DnCNN [131] model tries to overcome the issue of the lower receptive field but does not determine the 8x magnification general and the facial image accurately. It introduces the mixed convolution approach but due to the use of the dilated convolution it increased the receptive field but at the reconstruction end of the image, the deconvolution layer generates the blind spots. As compared to our network we have used the Dilated convolution layer to increase the receptive field, the Leaky Relu Layer as an activation function which helps to activate the value at 0.1x at (x<o), and the sub-pixel convolution layer is used for removing the check board issue generated by the deconvolution layer at the reconstruction end and to avoid blind spots generation we have tried to use correlation filter and also proposed sum of two-loss function MSE and SSIM, which is not being used in the previous studies. The whole network is newly designed with 33 layers and successfully provides better perceptual quality in comparison with the average PSNR and SSIM among all the models. The architecture of the progressive FHSR with the UNCF layers model allows us to super-resolve the image at several magnification factors. To determine the quality of the image we tried to compare our model with 10 different architectures. We trained 10 models with all three datasets and compared them with our architecture in terms of the average PSNR and SSIM. To improve the quality of the image at 24*24 low resolution we have used some new layers which help in improving the quality discussion of the layer in the previous section. Tables 5.6 and 5.7 show the comparative study of the networks with our network and Gan, Esrgan networks are the second runner which is giving very closer results as compared with our model. All the other techniques are performing well in 2x and 4x magnification but fail to generate better visual quality at an 8x magnification factor. We also observed that VDSR [34], SRCNN, Lp [141], and SICNN [135] are producing competitive performance but still, they have

low PSNR and SSIM. As compared to the CSRIP Model they have a 52-layer architecture and we reduce the layer size and use 33 layers which reduces the training time of the network A convolution Filter is used with the deconvolution layer for reducing the blind spot generation at the reconstruction end which makes our model much better as compared to all the other models. Therefore, our model has having simple architecture cascading network which is good for the super-resolution of a very tiny image like 24*24. Some of the comparisons of generated images are shown in Figure 5.6.

5.5 Run Time Efficiency

It is defined as the ratio of total useful work (input image) trained w.r.to the total time which is required for the training. Here, the input images used for the training is 5k images and the total time is 5730 minutes for the completion of the training of image.

Run Time Effieciney = (Useful Work / Total Time) × 100%

$$= (5000/5730) \times 100\ \% = 97.46\%$$

FHSR Network run time efficiency evaluated as 97.46%.

5.6 Summary

This chapter proposes a novel FHSR model with a loss-aware upscaling network layer. Integrating correlation filters and a combined loss function into the upscaling layer. The model is designed as a cascade of progressive upscaling modules with dense skip connection layers. Each upscaling module's loss functions are evaluated, cascaded, and fed to the subsequent layer to minimize reconstruction losses. The architecture upscaled the images with three cascading modules that magnify twofold, fourfold, and eightfold. In addition, the chapter presented the ablation study, which confirmed that the designed FHSR model is superior to baseline models. This architecture solves the problem of extracting more features with minimal loss and builds the high magnification at the end of reconstruction. The comparative study demonstrates superior performance to that of contemporary models and has a run time efficiency of 97.46%.

CHAPTER 6. CONCLUSION AND FUTURE SCOPE

In this work, many existing works are presented on image super-resolution using CNN, and their performance analysis is also presented that can be extensively used in to study. In recent research applications in the image processing field, it is highly recommended to develop new algorithms or models for improving existing image super-resolution techniques with the application of deep learning. The existing works mainly focus on CNN architecture and influencing parameters such as PSNR and SSIM for determining the efficacy of the model. There are very few works that have provided a deep description of the impact of the loss function and learning parameters for image super-resolution applications. The introduction and Recent work explore the technological innovations that are done by optimizing the loss function used in different CNN models designed for image super-resolution. The super-resolution aspects that are focused in this review are:

- Network Designs: Good networks are those that have fewer complex designs with fast training speed. This Chapter helps in getting information regarding the CNN network.
- Learning Strategies: Different Loss functions like perceptual loss, super identity loss, Perceptual loss, and L1 and L2 loss functions have been discussed with their performances and a comparison is also given.
- The upscaling techniques discussed help to magnify the image.

Performance Metric is presented which helps in getting information by which we can evaluate the performance of the super-resolution image.

In Chapter 3 By embedding correlation filter and recursive dilated convolution blocks in the network of CFIPC and by using a new hybrid loss function, $loss_{CFIPC}$, the effectiveness of the medical image SR model is presented. The experimental results are evaluated on five different medical input images and the effective result. The best result on 2x upscaling, the DRIVE dataset gives a PSNR value of 52.89 and at 4x and 8x, the Histopathological dataset achieved 48.27 and 46.61 respectively. Similarly, the best SSIM value was observed on the histopathological dataset at 2x/4x/8x upscaling of 0.9986, 0.9972, and 0.9972 respectively. The state-of-art shows that CFIPC outperforms better than (MSCNN), VDSR, SRGAN, (IGAN), and SAI networks on parameters such as PSNR and SSIM. Moreover, this letter also evaluated other parameters, such as the T-test that show the preservation of perceptual quality of reconstructed SR images. CFIPC shows the significance that it overcomes the smoothness problem and improves the results

with a run-time efficiency of 96.77%. In Chapter 4 we have proposed a Super-Resolution method by combining the advantages of cascaded deep residual learning and dilated convolutions layers. We have observed that the recursive nature of the proposed model preserves the network parameters while cascading. The model is modeled as Multilevel Progressive Recursive Dilated Networks with Correlation Filter (MPRDNCF). The MPRDNCF also showed that it has overcome the problem of smoothness and improved the results by minimizing the loss function while training. This Chapter also presented the correlation analysis of LR-HR image patches and observed that more accurate and better results are achieved. We have trained the network for different scale factors and observed the results on different test sets. The Chapter also presented the comparative state-of-arts and proved its efficiency over other techniques and having run-time efficiency of 97.50%. In future work, our model can be used for real-time video streaming problems. In Chapter 5 we have designed and implemented a novel progressive model with a loss-aware function for image super-resolution in medical images, general images, and facial images. Initially in Facial image super-resolution UNCF layers, and a CNN model for face hallucination super-resolution. The model is progressive and all layers pass their loss information to the next layer. This helps the model for minimizing the overall prediction errors. The result analysis was presented for 2x, 4x, and 8x magnification factors. The Work also presents the ablation study in comparison to baseline models. As the model is designed as loss-aware layers, therefore, the ablation study was presented with a variation of loss functions and tested over LFW, HELEN, and CELEBA datasets. The progressive FHSR model with loss-aware UNCF layers with combined loss function shows the highest PSNR and SSIM values. The designed progressive FHSR model with loss-aware UNCF layers shows the highest PSNR (i.e., 31.70, 30.55, and 27.910 at 2x, 4x, and 8x respectively) on the CELEBA dataset. Whereas the obtained SSIM values were 0.945, 0.903, and 0.861 for 2x, 4x, and 8x on the CELEBA dataset. The result analysis also presented better PSNR and SSIM values over state-of-art models. Therefore, it is concluded that our model overcomes the problem of smoothness and improves the super-resolution result by reducing the loss at the training end and reducing the size of the network which improves the training speed of the images and achieves better performance as compared to the previous architecture. Apart from this, dilated convolution is used in this model that overcomes the issue of receptive field and blind-spot generation at the reconstruction end with high magnification and having a run-time efficiency of 97.46%. But still, there exist some limitations in the model. The model

is designed basically for 2D images, therefore in the future, the model performance will be evaluated on 3D images and videos. Our model can be used for other images such as Remote Sensing images and video super-resolution in future work.

REFERENCE

[1] Grm, K., Scheirer, W. J., & Štruc, V. (2019). Face hallucination using cascaded super-resolution and identity priors. *IEEE Transactions on Image Processing*, *29*, 2150-2165.

[2] Xue, S., Qiu, W., Liu, F., & Jin, X. (2020). Faster image super-resolution by improved frequency-domain neural networks. *Signal, Image and Video Processing*, *14*, 257-265.

[3] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 212-220).

[4] Xia, R., Chen, Y., & Ren, B. (2022). Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), 6008-6018.

[5] Glasner D., Bagon S., & Irani M., "Super-Resolution from a Single Image." *Available: www.eecs.berkeley.edu/Research/Projects/CS/vision/grouping/segbench.*

[6] Shoeiby, M., Robles-Kelly, A., Timofte, R., Zhou, R., Lahoud, F., Susstrunk, S., ... & Ganapathysubramanian, B. (2018). PIRM2018 challenge on spectral image super-resolution: methods and results. In *Proceedings of the European Conference on Computer Vision (ECCV) Workshops*.

[7] Timofte, R., Agustsson, E., Van Gool, L., Yang, M. H., & Zhang, L. (2017). Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 114-125).

[8] Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5197-5206).

[9] Song, X., Dai, Y., & Qin, X. (2017). Deep depth super-resolution: Learning depth super-resolution using deep convolutional neural network. In *Computer Vision–*

ACCV 2016: 13th Asian Conference on Computer Vision, Taipei, Taiwan, November 20-24, 2016, Revised Selected Papers, Part IV 13 (pp. 360-376).

[10] Dong, C., Loy, C. C., He, K., & Tang, X. (2014). Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13* (pp. 184-199).

[11] Bovik, A. C. (2009). Basic gray-level image processing. In *The Essential Guide to Image Processing* (pp. 43-68). Academic Press.

[12] Li, C., & Wand, M. (2016). Combining Markov random fields and convolutional neural networks for image synthesis. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2479-2486).

[13] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 694-711).

[14] Lai, W. S., Huang, J. B., Ahuja, N., & Yang, M. H. (2017). Deep laplacian pyramid networks for fast and accurate super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 624-632).

[15] Ding, D., Kong, L., Wang, W., & Zhu, F. (2021). A progressive CNN in-loop filtering approach for inter-frame coding. *Signal Processing: Image Communication*, *94*, 116201.

[16] Miao, Y., Zhang, B., Lin, J., Zhao, M., Liu, H., Liu, Z., & Li, H. (2022). A review on the application of blind deconvolution in machinery fault diagnosis. *Mechanical Systems and Signal Processing*, *163*, 108202.

[17] Sharma, A., Srivastava, B. P., & Shankar, P. N. (2023). Facial Image Super-Resolution with CNN, "A Review". In *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)* (pp. 1-6). IEEE.

[18] Kaloev, M., & Krastev, G. (2021, June). Comparative analysis of activation functions used in the hidden layers of deep neural networks. In *2021 3rd*

International Congress on Human-Computer Interaction, Optimization and Robotic Applications (HORA) (pp. 1-5). IEEE.

[19] Hore, A., & Ziou, D. (2010, August). Image quality metrics: PSNR vs. SSIM. In *2010 20th International Conference on Pattern Recognition* (pp. 2366-2369). IEEE..

[20] Sai, Y., Jinxia, R., & Zhongxia, L. (2009, December). Learning of neural networks based on weighted mean squares error function. In *2009 Second International Symposium on Computational Intelligence and Design* (Vol. 1, pp. 241-244). IEEE.

[21] Ji, Y., Dai, Y., Ma, J., Zhao, K., & Cheng, Y. (2020, August). Generative adversarial network for image deblurring using content constraint loss. In *2020 Chinese Control And Decision Conference (CCDC)* (pp. 1985-1990). IEEE.

[22] Pan, Q., Pan, Y., Zhou, J., Wang, H., Tao, L., & Kwan, H. K. (2022, November). CycleGAN with the dual adversarial loss for bone-conducted speech enhancement. In *TENCON 2022-2022 IEEE Region 10 Conference (TENCON)* (pp. 1-4). IEEE.

[23] Sharma, A., & Shrivastava, B. P. (2022). Different Techniques of Image SR Using Deep Learning: A Review. *IEEE Sensors Journal*, *23*(3), (pp. 1724-1733).

[24] Hui, T. W., Loy, C. C., & Tang, X. (2016). Depth map super-resolution by deep multi-scale guidance. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14* (pp. 353-369).

[25] Mac Aodha, O., Campbell, N. D., Nair, A., & Brostow, G. J. (2012). Patch-based synthesis for single-depth image super-resolution. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12* (pp. 71-84).

[26] Ahn, N., Kang, B., & Sohn, K. A. (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 252-268).

[27] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).

[28] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[29] Gu, F. (2022, April). Research on Residual Learning of Deep CNN for Image Denoising. In *2022 7th International Conference on Intelligent Computing and Signal Processing (ICSP)* (pp. 1458-1461). IEEE.

[30] Ayyoubzadeh, S. M., & Wu, X. (2021). High-frequency detail accentuation in CNN image restoration. *IEEE Transactions on Image Processing*, *30*, 8836-8846.

[31] Riegler, G., Rüther, M., & Bischof, H. (2016). Atgv-net: Accurate depth super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part III 14* (pp. 268-284).

[32] Fujimoto, A., Ogawa, T., Yamamoto, K., Matsui, Y., Yamasaki, T., & Aizawa, K. (2016, December). Manga109 dataset and creation of metadata. In *Proceedings of the 1st international workshop on comics analysis, processing and understanding* (pp. 1-5).

[33] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 136-144).

[34] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).

[35] Anwar, S., Khan, S., & Barnes, N. (2020). A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, *53*(3), (pp. 1-34).

[36] Kim, J. H., & Lee, J. S. (2018). Deep residual network with enhanced upscaling module for super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 800-808).

[37] Xu, S., Zhang, J., Sun, K., Zhao, Z., Huang, L., Liu, J., & Zhang, C. (2021, July). Deep convolutional sparse coding network for pansharpening with the guidance of side information. In *2021 IEEE International Conference on Multimedia and Expo (ICME)* (pp. 1-6). IEEE.

[38] Kim, J., Lee, J. K., & Lee, K. M. (2016). Deeply recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637-1645).

[39] Tai, Y., Yang, J., Liu, X., & Xu, C. (2017). Memnet: A persistent memory network for image restoration. In *Proceedings of the IEEE International Conference on Computer Vision* (pp. 4539-4547).

[40] Chen, Y., Tai, Y., Liu, X., Shen, C., & Yang, J. (2018). Fsrnet: End-to-end learning face super-resolution with facial priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2492-2501).

[41] Chen, Y., Tai, Y., Liu, X., Shen, C., & Yang, J. (2018). Fsrnet: End-to-end learning face super-resolution with facial priors. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2492-2501).

[42] Li, Z., Yang, J., Liu, Z., Yang, X., Jeon, G., & Wu, W. (2019). Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition* (pp. 3867-3876).

[43] Jiang, L., Zhong, M., & Qiu, F. (2020, October). Single-image super-resolution based on a self-attention deep neural network. In *2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI)* (pp. 387-391). IEEE.

[44] Fan, Z., Dan, T., Yu, H., Liu, B., & Cai, H. (2020, July). Single fundus image super-resolution via cascaded channel-wise attention network. In *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)* (pp. 1984-1987). IEEE.

[45] Zhang, Z., Liu, K., Xia, L., & Li, Y. (2022, May). Multi-channel Network with Dense Attention for Image Super-Resolution. In *2022 3rd International Conference on Computer Vision, Image and Deep Learning & International Conference on Computer Engineering and Applications (CVIDL & ICCEA)* (pp. 665-670). IEEE.

[46] Xue, Z., Wang, L., & Guo, P. (2019, June). Slot-based Image Captioning with WGAN. In *2019 IEEE/ACIS 18th International Conference on Computer and Information Science (ICIS)* (pp. 241-246). IEEE.

[47] Ma, Y., Liang, Y., Zhang, W., & Yan, S. (2019, November). SAR target recognition based on transfer learning and data augmentation with LSGANs. In *2019 Chinese Automation Congress (CAC)* (pp. 2334-2337). IEEE.

[48] Yao, X., Wu, Q., Zhang, P., & Bao, F. (2019). Adaptive rational fractal interpolation function for image super-resolution via local fractal analysis. *Image and Vision Computing*, *82*, (pp. 39-49).

[49] Basak, J. (2009, February). Detection of neural activities in FMRI using Jensen-Shannon Divergence. In *2009 Seventh International Conference on Advances in Pattern Recognition* (pp. 39-42). IEEE.

[50] Yu, X., & Porikli, F. (2016, September). Ultra-resolving face images by discriminative generative networks. In *European conference on computer vision* (pp. 318-333).

[51] Bhawna, G., Chyophel, L. D., Ayush, D., & Shui-Hua, W. (2021). Correction to: A weighted least squares optimization strategy for medical image super-resolution via multiscale convolutional neural networks for healthcare applications. *Complex & Intelligent Systems*, *7*(6), (pp.. 3325-3325).

[52] Chen, Y., Liu, L., Phonevilay, V., Gu, K., Xia, R., Xie, J., ... & Yang, K. (2021). Image super-resolution reconstruction based on feature map attention mechanism. *Applied Intelligence*, *51*, (pp. 4367-4380).

[53] Wan, L., Sun, Y., Sun, L., Ning, Z., & Rodrigues, J. J. (2020). Deep learning-based autonomous vehicle super-resolution DOA estimation for safety driving. *IEEE Transactions on Intelligent Transportation Systems*, *22*(7), (pp. 4301-4315).

[54] J. Huang, L. Wang, J. Qin, Y. Chen, X. Cheng and Y. Zhu, "Super-Resolution of Intravoxel Incoherent Motion Imaging Based on Multisimilarity," *in IEEE Sensors Journal*, vol. 20, no. 18, (pp. 10963-10973).

[55] Xia, R., Chen, Y., & Ren, B. (2022). Improved anti-occlusion object tracking algorithm using Unscented Rauch-Tung-Striebel smoother and kernel correlation filter. *Journal of King Saud University-Computer and Information Sciences*, *34*(8), (pp. 6008-6018).

[56] Chen, Y., Liu, L., Tao, J., Xia, R., Zhang, Q., Yang, K., & Chen, X. (2021). The improved image inpainting algorithm via an encoder and similarity constraint. *The Visual Computer*, *37*, (pp.1691-1705).

[57] Brown, K. G., Ghosh, D., & Hoyt, K. (2020). Deep learning of spatiotemporal filtering for fast super-resolution ultrasound imaging. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, *67*(9), (pp.1820-1829).

[58] Han, J., & Wang, C. (2019). TSR-TVD: Temporal super-resolution for time-varying data analysis and visualization. *IEEE transactions on visualization and computer graphics*, *26*(1), (pp.205-215).

[59] Chen, G., Wang, H., Chen, K., Li, Z., Song, Z., Liu, Y.,& Knoll, A. (2020). A survey of the four pillars for small object detection: Multiscale representation, contextual information, super-resolution, and region proposal. *IEEE Transactions on systems, man, and cybernetics: systems*, *52*(2), (pp.936-953).

[60] Yamashita, K., & Markov, K. (2020). Medical image enhancement using super-resolution methods. In *Computational Science–ICCS 2020: 20th International Conference, Amsterdam, The Netherlands, June 3–5, 2020, Proceedings, Part V 20* (pp. 496-508).

[61] Tang, S., Song, P., Trzasko, J. D., Lowerison, M., Huang, C., Gong, P., & Chen, S. (2020). Kalman filter-based microbubble tracking for robust super-resolution ultrasound microvessel imaging. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, *67*(9), (pp.1738-1751).

[62] Kanakaraj, S., Nair, M. S., & Kalady, S. (2020). Adaptive importance sampling unscented Kalman filter with kernel regression for SAR image super-resolution. *IEEE Geoscience and Remote Sensing Letters*, *19*, (pp.1-5).

[63] Deeba, F., Kun, S., Dharejo, F. A., & Zhou, Y. (2020). Wavelet-based enhanced medical image super-resolution. *IEEE Access*, *8*, (pp.37035-37044).

[64] Xue, S., Qiu, W., Liu, F., & Jin, X. (2020). Wavelet-based residual attention network for image super-resolution. *Neurocomputing*, *382*, (pp.116-126).

[65] Xin, J., Li, J., Jiang, X., Wang, N., Huang, H., & Gao, X. (2020). Wavelet-based dual recursive network for image super-resolution. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), (pp.707-720).

[66] Tian, C., Zhuge, R., Wu, Z., Xu, Y., Zuo, W., Chen, C., & Lin, C. W. (2020). Lightweight image super-resolution with enhanced CNN. *Knowledge-Based Systems*, *205*, 106235.

[67] Hu, Z., Wang, Y., Zhang, X., Zhang, M., Yang, Y., Liu, X., & Liang, D. (2019). Super-resolution of PET image based on dictionary learning and random forests. *Nuclear Instruments and Methods in Physics Research Section A: Accelerators, Spectrometers, Detectors and Associated Equipment*, *927*, (pp.320-329).

[68] Tanno, R., Worrall, D. E., Ghosh, A., Kaden, E., Sotiropoulos, S. N., Criminisi, A., & Alexander, D. C. (2017). Bayesian image quality transfer with CNNs: exploring uncertainty in dMRI super-resolution. In *Medical Image Computing and Computer Assisted Intervention− MICCAI 2017: 20th International Conference, Quebec City, QC, Canada, September 11-13, 2017, Proceedings, Part I 20* (pp. 611-619).

[69] Irani, M., & Peleg, S. (1990, June). Super-resolution from image sequences. In *[1990] Proceedings. 10th International Conference on Pattern Recognition* (Vol. 2, pp. 115-120).

[70] Tsai, R. Y., & Huang, T. S. (1984). Multiframe image restoration and registration. *Multiframe image restoration and registration*, *1*, (pp.317-339).

[71] Zhang, Z., Wang, X., & Jung, C. (2018). DCSR: Dilated convolutions for single image super-resolution. *IEEE Transactions on Image Processing*, *28*(4), (pp.1625-1635).

[72] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).

[73] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).

[74] Bing, X., Zhang, W., Zheng, L., & Zhang, Y. (2019). Medical image super-resolution using improved generative adversarial networks. *IEEE Access*, *7*, (pp.145030-145038).

[75] Chen, Z., Guo, X., Woo, P. Y., & Yuan, Y. (2021). Super-resolution enhanced medical image diagnosis with sample affinity interaction. *IEEE Transactions on Medical Imaging*, *40*(5), (pp.1377-1389).

[76] Al-Rawi, M., Qutaishat, M., & Arrar, M. (2007). An improved matched filter for blood vessel detection of digital retinal images. *Computers in biology and medicine*, *37*(2), (pp.262-267).

[77] Fraz, M. M., Remagnino, P., Hoppe, A., Uyyanonvara, B., Rudnicka, A. R., Owen, C. G., & Barman, S. A. (2012). An ensemble classification-based approach applied to retinal blood vessel segmentation. *IEEE Transactions on Biomedical Engineering*, *59*(9), (pp.2538-2548).

[78] Rueda, A., Malpica, N., & Romero, E. (2013). Single-image super-resolution of brain MR images using overcomplete dictionaries. *Medical image analysis*, *17*(1), (pp.113-132).

[79] Gagan, J. H., Shirsat, H. S., Mathias, G. P., Mallya, B. V., Andrade, J., Rajagopal, K. V., & Kumar, J. H. (2022). Automated Segmentation of Common Carotid Artery in Ultrasound Images. *IEEE Access*, *10*, (pp.58419-58430).

[80] Chen, Y., Zhang, H., Liu, L., Tao, J., Zhang, Q., Yang, K., ... & Xie, J. (2021). Research on image inpainting algorithm of improved total variation minimization method. *Journal of Ambient Intelligence and Humanized Computing*, (pp.1-10).

[81] Choi, J. S., & Kim, M. (2017). Single image super-resolution using global regression based on multiple local linear mappings. *IEEE transactions on image processing*, *26*(3), (pp.1300-1314).

[82] Yang, J., Wang, Z., Lin, Z., Cohen, S., & Huang, T. (2012). Coupled dictionary training for image super-resolution. *IEEE transactions on image processing*, *21*(8), (pp.3467-3478).

[83] Zhu, X., Wang, X., Wang, J., Jin, P., Liu, L., & Mei, D. (2017). Image super-resolution based on sparse representation via direction and edge dictionaries. *Mathematical Problems in Engineering*, *2017*.

[84] Zhang, F. L., Wang, J., Shechtman, E., Zhou, Z. Y., Shi, J. X., & Hu, S. M. (2016). Plenopatch: Patch-based plenoptic image manipulation. *IEEE transactions on visualization and computer graphics*, *23*(5), (pp.1561-1573).

[85] Dong, C., Loy, C. C., He, K., & Tang, X. (2015). Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, *38*(2), (pp.295-307).

[86] Dong, C., Loy, C. C., & Tang, X. (2016). Accelerating the super-resolution convolutional neural network. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 391-407).

[89] Lim, B., Son, S., Kim, H., Nah, S., & Mu Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 136-144).

[90] Song, T. A., Chowdhury, S. R., Yang, F., & Dutta, J. (2020). Super-resolution PET imaging using convolutional neural networks. *IEEE transactions on computational imaging*, *6*, (pp.518-528).

[91] Sun, Y., Zhang, W., Gu, H., Liu, C., Hong, S., Xu, W., ... & Gui, G. (2019). Convolutional neural network-based models for improving super-resolution imaging. *Ieee Access*, *7*, (pp.43042-43051).

[92] Ma, J., Yu, J., Liu, S., Chen, L., Li, X., Feng, J., ... & Cheng, S. (2020). PathSRGAN: multi-supervised super-resolution for cytopathological images using the generative adversarial network. *IEEE transactions on medical imaging*, *39*(9), (pp.2920-2930).

[93] Yun, J. U., Jo, B., & Park, I. K. (2020). Joint face super-resolution and deblurring using the generative adversarial network. *IEEE Access*, *8*, (pp.159661-159671).

[94] Yang, Y., Li, Q., Yang, C., Fu, Y., Feng, H., Xu, Z., & Chen, Y. (2020). Deep networks with detail enhancement for infrared image super-resolution. *IEEE Access*, *8*, (pp.158690-158701).

[95] Mei, S., Jiang, R., Li, X., & Du, Q. (2020). Spatial and spectral joint super-resolution using convolutional neural network. *IEEE Transactions on Geoscience and Remote Sensing*, *58*(7), (pp.4590-4603).

[96] Cao, L., Liu, J., Du, K., Guo, Y., & Wang, T. (2020). Guided cascaded super-resolution network for face image. *IEEE Access*, *8*, (pp.173387-173400).

[97] Hung, K. W., Zhang, Z., & Jiang, J. (2019). Real-time image super-resolution using recursive depthwise separable convolution network. *IEEE Access*, *7*, (pp. 99804-99816).

[98] Zhang, Z., Wang, X., & Jung, C.: "DCSR: Dilated convolutions for single image super-resolution", in *IEEE Trans. Image Process.*, 2019, 28, (4), (pp. 1625–1635).

[99] Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., & Wang, Z. (2016). 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* (pp. 1874-1883).

[100] Dumoulin, V., & Visin, F. (2016). A guide to convolution arithmetic for deep learning. *arXiv preprint arXiv:1603.07285*.

[101] Timofte, R., De Smet, V., & Van Gool, L. (2015). A+: Adjusted anchored neighborhood regression for fast super-resolution. In *Computer Vision--ACCV 2014: 12th Asian Conference on Computer Vision, Singapore, Singapore, November 1-5, 2014, Revised Selected Papers, Part IV 12* (pp. 111-126).

[102] Bevilacqua, Marco and Roumy, Aline and Guillemot, Christine and Alberi-Morel, Marie Line *Low-complexity single-image super-resolution based on nonnegative neighbor embedding.* In: Proceedings of the 23rd British Machine Vision Conference (BMVC). BMVA Press, 135.1-135.10.

[103] Zeyde, R., Elad, M., & Protter, M. (2012). On single image scale-up using sparse representations. In *Curves and Surfaces: 7th International Conference, Avignon, France, June 24-30, 2010, Revised Selected Papers 7* (pp. 711-730).

[104] Huang, J. B., Singh, A., & Ahuja, N. (2015). Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 5197-5206).

[105] Kim, J., Lee, J. K., & Lee, K. M. (2016). Deeply recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1637-1645).

[106] Zhang, Y., Tian, Y., Kong, Y., Zhong, B., & Fu, Y. (2018). Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 2472-2481).

[107] Grm, K., Scheirer, W. J., & Štruc, V. (2019). Face hallucination using cascaded super-resolution and identity priors. *IEEE Transactions on Image Processing*, *29*, 2150-2165.

[108] T. Kanade, J. F. Cohn and Yingli Tian, "Comprehensive database for facial expression analysis," *Proceedings Fourth IEEE International Conference on Automatic Face and Gesture Recognition (Cat. No. PR00580)*, Grenoble, France, 2000, (pp. 46-53).

[109] Liu, W., Wen, Y., Yu, Z., Li, M., Raj, B., & Song, L. (2017). Sphereface: Deep hypersphere embedding for face recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 212-220).

[110] Haris, M., Shakhnarovich, G., & Ukita, N. (2021). Task-driven super-resolution: Object detection in low-resolution images. In *Neural Information Processing: 28th International Conference, ICONIP 2021, Sanur, Bali, Indonesia, December 8–12, 2021, Proceedings, Part V 28* (pp. 387-395).

[111] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).

[112] Kim, J. S., Ko, K., & Kim, C. S. (2021). Gluing reference patches together for face super-resolution. *IEEE Access*, *9*, (pp.169321-169334).

[113] Luo, S., & Lu, J. (2022). GFNet: a gradient information compensation-based face super-resolution network. *IEEE Access*, *10*, (pp.8073-8080).

[114] Zhang, Q., Yang, G., & Zhang, G. (2021). Collaborative network for super-resolution and semantic segmentation of remote sensing images. *IEEE Transactions on Geoscience and Remote Sensing*, *60*, (pp.1-12).

[115] Yu, X., Fernando, B., Ghanem, B., Porikli, F., & Hartley, R. (2018). Face super-resolution guided by facial component heatmaps. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 217-233).

[116] Nazeri, K., Thasarathan, H., & Ebrahimi, M. (2019). Edge-informed single image super-resolution. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*.

[117] Seif, G., & Androutsos, D. (2018, April). The edge-based loss function for single image super-resolution. In *2018 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)* (pp. 1468-1472).

[118] Mac Aodha, O., Campbell, N. D., Nair, A., & Brostow, G. J. (2012). Patch-based synthesis for single-depth image super-resolution. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12* (pp. 71-84).

[119] Zhu, X., Wang, X., Wang, J., Jin, P., Liu, L., & Mei, D. (2017). Image super-resolution based on sparse representation via direction and edge dictionaries. *Mathematical Problems in Engineering*, *2017*.

[120] Anari, V., Razzazi, F., & Amirfattahi, R. (2019). A sparse analysis-based single image super-resolution. *Computers*, *8*(2), (pp.41).

[121] Choi, J. H., Kim, J. H., Cheon, M., & Lee, J. S. (2020). Deep learning-based image super-resolution considering quantitative and perceptual quality. *Neurocomputing*, *398*, (pp.347-359).

[122] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770-778).

[123] Z. Xu, Q. Ma and F. Yuan, "Single color image super-resolution using sparse representation and color constraint," in *Journal of Systems Engineering and Electronics*, vol. 31, no. 2, (pp. 266-271).

[124] Tabrizchi, H., Mosavi, A., Vamossy, Z., & Varkonyi-Koczy, A. R. (2021, June). Densely connected convolutional networks (DenseNet) for diagnosing coronavirus disease (COVID-19) from chest X-ray imaging. In *2021 IEEE International Symposium on Medical Measurements and Applications (MeMeA)* (pp. 1-5).

[125] Wang, Z., Chen, J., & Hoi, S. C. (2020). Deep learning for image super-resolution: A survey. *IEEE transactions on pattern analysis and machine intelligence*, *43*(10), (pp.3365-3387).

[126] Anwar, S., Khan, S., & Barnes, N. (2020). A deep journey into super-resolution: A survey. *ACM Computing Surveys (CSUR)*, *53*(3), (pp.1-34).

[127] Liu, C., Shum, H. Y., & Freeman, W. T. (2007). Face hallucination: Theory and practice. *International Journal of Computer Vision*, *75*, (pp.115-134).

[128] Le, V., Brandt, J., Lin, Z., Bourdev, L., & Huang, T. S. (2012). Interactive facial feature localization. In *Computer Vision–ECCV 2012: 12th European Conference on Computer Vision, Florence, Italy, October 7-13, 2012, Proceedings, Part III 12* (pp. 679-692).

[129] Wang, Z., Bovik, A. C., Sheikh, H. R., & Simoncelli, E. P. (2004). Image quality assessment: from error visibility to structural similarity. *IEEE transactions on image processing*, *13*(4), (pp. 600-612).

[130] Wang, Y., Xie, Z., Xu, K., Dou, Y., & Lei, Y. (2016). An efficient and effective convolutional auto-encoder extreme learning machine network for 3D feature learning. *Neurocomputing*, *174*, (pp.988-998).

[132] Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., & Shi, W. (2017). Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 4681-4690).

[133] Ahn, N., Kang, B., & Sohn, K. A. (2018). Fast, accurate, and lightweight super-resolution with cascading residual network. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 252-268).

[134] Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. (2017). Enhanced deep residual networks for single image super-resolution. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition Workshops* (pp. 136-144).

[135] Zhang, K., Zhang, Z., Cheng, C. W., Hsu, W. H., Qiao, Y., Liu, W., & Zhang, T. (2018). Super-identity convolutional neural network for face hallucination. In *Proceedings of the European Conference on Computer Vision (ECCV)* (pp. 183-198).

[136] Yu, X., & Porikli, F. (2018). Imagining the unimaginable faces by deconvolutional networks. *IEEE Transactions on Image Processing*, *27*(6), (pp. 2747-2761).

[137] Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., Dong, C., ... & Change Loy, C. (2018). Esrgan: Enhanced super-resolution generative adversarial networks. In *Proceedings of the European Conference on Computer Vision (ECCV) workshops*.

[138] Yu, X., & Porikli, F. (2016, September). Ultra-resolving face images by discriminative generative networks. In *European conference on computer vision* (pp. 318-333).

[139] Kim, J., Lee, J. K., & Lee, K. M. (2016). Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 1646-1654).

[140] C. Dong, C. C. Loy, K. He, and X. Tang, “Learning a deep convolutional network for image super-resolution,” in Proc. Eur. Conf. Comput. Vis. (ECCV), 2014, (pp. 184–199).

[141] Johnson, J., Alahi, A., & Fei-Fei, L. (2016). Perceptual losses for real-time style transfer and super-resolution. In *Computer Vision–ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11-14, 2016, Proceedings, Part II 14* (pp. 694-711).

[142] Xue, S., Qiu, W., Liu, F., & Jin, X. (2020). Faster image super-resolution by improved frequency-domain neural networks. *Signal, Image and Video Processing*, *14*, (pp.257-265).

[143] Lai, W. S., Huang, J. B., Ahuja, N., & Yang, M. H. (2018). Fast and accurate image super-resolution with deep laplacian pyramid networks. *IEEE transactions on pattern analysis and machine intelligence*, *41*(11), (pp.2599-2613).

[144] Gao, H., Xiao, J., Yin, Y., Liu, T., & Shi, J. (2022). A mutually supervised graph attention network for few-shot segmentation: the perspective of fully utilizing limited samples. *IEEE Transactions on neural networks and learning systems*.

[145] Gao, H., Qiu, B., Barroso, R. J. D., Hussain, W., Xu, Y., & Wang, X. (2022). Tsmae: a novel anomaly detection approach for the Internet of Things time series data using memory-augmented autoencoder. *IEEE Transactions on network science and engineering*.

[146] Gao, H., Huang, J., Tao, Y., Hussain, W., & Huang, Y. (2022). The joint method of triple attention and novel loss function for entity relation extraction in small data-driven computational social systems. *IEEE Transactions on Computational Social Systems*, *9*(6), (pp.1725-1735).

[147] Gao, H., Xu, K., Cao, M., Xiao, J., Xu, Q., & Yin, Y. (2021). The deep features and attention mechanism-based method to dish healthcare under social IoT systems: an empirical study with a hand-deep local–global net. *IEEE Transactions on Computational Social Systems*, *9*(1), (pp.336-347).

[148] Gao, H., Dai, B., Miao, H., Yang, X., Barroso, R. J. D., & Walayat, H. (2023). A novel gap approach to automatic property generation for formal verification: The gan perspective. *ACM Transactions on Multimedia Computing, Communications and Applications*, *19*(1), (pp.1-22).

LIST OF PUBLICATION

Journals:

[1] Sharma, A., & Shrivastava, B. P. (2022), "Different Techniques of Image SR Using Deep Learning: A Review", *IEEE Sensors Journal*, 23(3), 1724-1733. DOI: 10.1109/JSEN.2022.3229046.

[2] Sharma, A., Shrivastava, B.P., "Facial image super-resolution using progressive network interleaved correlation filter". *Multimed Tools Appl 82*, 29587–29606 (2023).. DOI:https://doi.org/10.1007/s11042-023-14765-8.

[3] Sharma, A., & Shrivastava, B. P. (2022), "Medical image super-resolution using correlation filter interleaved progressive convolution network (CFIPC)", *Electronics Letters*, 58(9), 360-362. DOI: https://doi.org/10.1049/ell2.12467.

[4] Sharma, A., Shrivastava, B.P. & Priya, A. Multilevel progressive recursive dilated networks with correlation filter (MPRDNCF) for image super-resolution. *Multimedia Systems* 29, 2455–2467 (2023). https://doi.org/10.1007/s00530-023-01126-6

[5] Sharma, A., & Shrivastava, B. P., "Complex Wavelet Transform with Progressive Network for Medical Imaging Super Resolution", *Multimedia Tools Applications.* Springer (2023). **SCIE. (Under Revision).**

Conference/ Book Chapter

[1] Sharma, A., Srivastava, B. P., & Shankar, P. N., "Facial Image Super-Resolution with CNN, "A Review"," *2023 IEEE International Students' Conference on Electrical, Electronics and Computer Science (SCEECS)*, pp. 1-6, DOI: 10.1109/SCEECS57921.2023.10063110

[2] Sharma, A, Priya Ayushi, and Srivastava, B. P., "Quantum Blockchain: A Persuasive Approach for Future Security", *Proceedings of IETE National Conference on "VLSI, Communication and Signal Processing", (NCVCS 2021),* 27th-28th November 2021.

[3] Sharma, A., Srivastava, B. P., and Gautam Swati, "A Review on Image Super-Resolution Using GAN", *IGI Globus*, DOI: 10.4018/978-1-6684-7659-8.

AUTHORS CONTRIBUTION

- **Conceptualization**

  Ideas; formulation or evolution of overarching research goals and aims.

- **Investigation**

  Conducting a research and investigation process, specifically performing the experiments, or data/evidence collection.

- **Methodology**

  Development or design of methodology; creation of models.

- **Software**

  Programming, software development; designing computer programs; implementation of the computer code and supporting algorithms; testing of existing code components.

- **Supervision**

  Oversight and leadership responsibility for the research activity planning and execution, including mentorship external to the core team.

- **Validation**

  Verification, whether as a part of the activity or separate, of the overall replication/reproducibility of results/experiments and other research outputs.

- **Visualization**

  Preparation, creation, and/or presentation of the published work, specifically visualization/data presentation.

- **Writing – Original Draft Preparation**

  Creation and/or presentation of the published work, specifically writing the initial draft (including substantive translation).

- **Writing – Review & Editing**

Authors' Contribution statement using Credit with degree of contribution:

Mr. Ajay Sharma: Investigation, Conceptualization (lead), Visualisation, Validation, Writing – original draft (lead); formal analysis (lead); writing – review and editing (equal), Software (lead), Methodology (lead), Writing – original draft (supporting).

Dr. Bhavana P. Shrivastava: Supervision, Conceptualization (supporting), review, and editing (equal).

Image Super Resolution (193114013)

by Ajay Sharma

Submission date: 05-Jun-2023 01:51PM (UTC+0530)
Submission ID: 2109326251
File name: Ajay_thesis.pdf (2.61M)
Word count: 42633
Character count: 245075

Image Super Resolution (193114013)

ORIGINALITY REPORT

10%	5%	7%	2%
SIMILARITY INDEX	INTERNET SOURCES	PUBLICATIONS	STUDENT PAPERS

PRIMARY SOURCES

1	Submitted to iGroup Student Paper	<1%
2	cds.iisc.ac.in Internet Source	<1%
3	S. B. Goyal, Pradeep Bedi, Jugnesh Kumar, Vijaykumar Varadarajan. "Deep learning application for sensing available spectrum for cognitive radio: An ECRNN approach", Peer-to-Peer Networking and Applications, 2021 Publication	<1%
4	www.politesi.polimi.it Internet Source	<1%
5	"Computer Vision – ECCV 2012", Springer Nature, 2012 Publication	<1%
6	Shaohui Mei, Ruituo Jiang, Xu Li, Qian Du. "Spatial and Spectral Joint Super-Resolution Using Convolutional Neural Network", IEEE Transactions on Geoscience and Remote Sensing, 2020 Publication	<1%