

Image Super-Resolution by Combining Non-Local Sparse Attention and Residual Channel Attention

by

Bhavsar Manali Hiteshkumar
202011022

A Thesis Submitted in Partial Fulfilment of the Requirements for the Degree of

MASTER OF TECHNOLOGY
in
INFORMATION AND COMMUNICATION TECHNOLOGY
to

DHIRUBHAI AMBANI INSTITUTE OF INFORMATION AND COMMUNICATION TECHNOLOGY

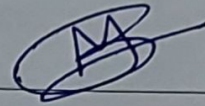


May, 2022

Declaration

I hereby declare that

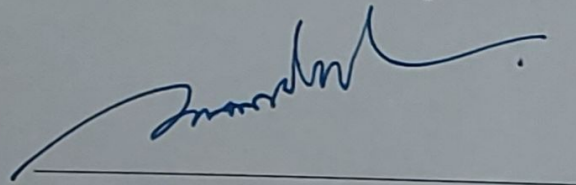
- i) the thesis comprises of my original work towards the degree of Master of Technology in Information and Communication Technology at Dhirubhai Ambani Institute of Information and Communication Technology and has not been submitted elsewhere for a degree,
- ii) due acknowledgment has been made in the text to all the reference material used.



Bhavsar Manali Hiteshkumar

Certificate

This is to certify that the thesis work entitled **Image Super-Resolution by Combining Non-Local Sparse Attention and Residual Channel Attention** has been carried out by BHAVSAR MANALI HITESHKUMAR for the degree of Master of Technology in Information and Communication Technology at *Dhirubhai Ambani Institute of Information and Communication Technology* under my/our supervision.



Prof. Srimanta Mandal
Thesis Supervisor

Acknowledgments

The satisfaction and euphoria that accompany the successful completion of any task would be incomplete without mentioning the people who made it possible.

I wish to express my deep sense of gratitude towards my thesis supervisor **Prof. Srimanta Mandal** for allowing me to work with him and patiently listening to my doubts and addressing them. His support and guidance motivated me to approach the problem with enthusiasm and excitement.

I am also thankful to my friends **Krishna Savaliya** and **Akash Dhedhi** for always supporting me and by my side throughout the research work. With your company, this research work would be more memorable and joyful.

Anything is possible when you have the right people there to support you. My friend **Dhruvi Bhavsar** and **Devanshi Panchal**. I certainly know that you are my forever well-wishers and supporters.

I would also like to thank my classmates, Shivangi, Shradhha, Dhyanil for helping me. I am also gratefully thanking our college staff for providing facilities to complete the thesis.

Finally, but most importantly, I would like to express my gratitude to my parents and family for their unwavering love and support, which did not allow the tough times to dominate me. I acknowledge the help of all those who had given encouragement and associated themselves in one way or the other in completing the thesis work.

Contents

Abstract	v
List of Tables	vi
List of Figures	vii
1 Introduction	1
1.1 Single Image Super-Resolution	1
1.2 Objective	3
1.3 Contribution of Thesis	4
1.4 Organization of Thesis	4
2 Literature Survey and Background	5
2.1 Deep Learning Architectures for SISR task	5
2.1.1 SRCNN	5
2.1.2 RCAN (Residual Channel Attention Network)	7
2.1.3 NLSN (Non Local Sparse Attention Network)	9
2.2 Sparse Representation	10
2.3 Image Quality Assessments	10
2.3.1 Peak Signal-to-Noise Ratio (PSNR)	10
2.3.2 Structural Similarity Index (SSIM)	11
3 Proposed Method	13
3.1 Non Local Sparse Attention	13
3.1.1 Non Local Attention	14
3.1.2 Non-local Attention with sparsity constraint	15
3.1.3 Attention Bin using LSH(Locality Sensitive Hashing)	15
3.2 Residual Groups	17
3.3 Loss Function	17

4	Experiments and Results	18
4.1	Dataset	18
4.2	Training and Implementation details	19
4.3	Results and Comparison	19
4.4	Other Experiments	25
4.4.1	Other Experiments	25
4.4.2	RCAN_Dense	25
4.4.3	RCAPAN(Residual Channel Attention and Pixel Attention Network)	26
4.4.4	RPCSN(Residual Channel Pixel Spatial Network)	27
4.4.5	Results of Experiments	28
4.5	Experiments with other Image Modalities	29
4.5.1	Depth Map	29
4.5.2	X-Ray Images	31
5	Conclusion & Future Scope	33
5.1	Conclusion	33
5.2	Future Work	33
	References	34

Abstract

Single Image Super-Resolution (SISR) is an ill-posed problem that aims to generate a high-resolution (HR) image from a single low-resolution (LR) image. The low resolution image and its associated features are very rich in low frequency information. The main objective of super-resolution is to add relevant high frequency detail to complement the available low frequency information. Classical techniques such as non-local similarity and sparse representations both have shown promising results in SISR task in past decades. Nowadays, deep learning techniques such as convolutional neural networks (CNN) can extract deep features to improve the results of SISR task. However, CNN does not explicitly consider the similar information in the image. Hence, we employ non-local sparse attention (NLSA) module in the CNN framework such that it can explore the non-local similarity within an image. We consider sparsity in the non-local operation by focusing on a particular group named attention bin among many groups of features. Non-local Sparse Attention is intended to retain the long-range of non-local operation modeling capacity while benefiting from the efficiency and robustness of sparse representation. Additionally, we try to rescale the channel-specific features adaptively while taking into account channel interdependence by using residual channel attention. In this thesis work, we try to incorporate and combine the advantages of non-local sparse attention (NLSA) and residual channel attention to produce results similar to state-of-the-art methods.

Keywords: Deep Learning Techniques, Single Image Super Resolution, Channel Attention, Non-local Sparse Attention.

List of Tables

4.1	List of Datasets used for SISR task	18
4.2	Quantitative analysis of different Architectures (Scale 2 & 3)	21
4.3	Quantitative analysis of different Architectures(Scale 4)	22
4.4	Other Experiments' Quantitave results	29
4.5	Quantitative results of Depthmap SR on scale 4 upsampling on Middlebury dataset in terms of RMSE values	31
4.6	Quantitative results of this on scale 2 upsampling in terms of PSNR	32

List of Figures

1.1	Image Super Resolution: The objective	1
1.2	Some Applications of Image Super Resolution	2
2.1	Architecture of Super-Resolution Convolutional Neural Networks [3]	6
2.2	Architecture of Residual Channel Attention Network [28]	7
2.3	Architecture of Residual Channel Attention Block	8
2.4	Architecture of Non Local Sparse Attention Network[18]	9
3.1	Proposed Method Architecture	13
3.2	Non Local Sparse Attention[18]	14
3.3	Example of attention bins[18].	16
3.4	Residual Channel Attention Block	17
4.1	Qualitative Results on Scale 2 (Set5: Butterfly.png)	22
4.2	Qualitative Results on Scale 3 (Set14: comic.png)	23
4.3	Qualitative Results on Scale 4 (Urban100: img_002.png)	23
4.4	Qualitative Results on Scale 4 (Urban100: img_093.png)	24
4.5	(a) CA: Channel Attention; (b) SA: Spatial Attention; (c) PA: Pixel Attention. [31]	25
4.6	RCAN_dense Architecture	26
4.7	RCAPAN Architecture	27
4.8	RPCS Architecture	28
4.9	Qualitative results of our network on scale 4 on Middlebury 2005 Art and Laundry image[21]	30
4.10	Qualitative results of our network on scale 2 upsampling	32

CHAPTER 1

Introduction

1.1 Single Image Super-Resolution

Single Image Super-Resolution (SISR) aims to generate a high resolution (HR) image from a given low resolution (LR) image.



Figure 1.1: Image Super Resolution: The objective

It has various applications in medical imaging(e.g., Figure 1.2(a)), surveillance (e.g., Figure 1.2(b)), remote sensing, etc. In medical imaging such as magnetic resonance imaging(MRI) the resolution of the image is very crucial to detect any damaged tissue/ tumor. For LR MRI slice, such detection will be a difficult task. Here, SR can play a significant role by providing a HR MR slice. Further, in surveillance applications such as distant face recognition by a CCTV footage is not a trivial task as the captured face image could have resolution which is very low. It has been found that exiting face recognition algorithms do not produce optimal results on face images with resolution lower than 16×16 [32]. Furthermore, in remote sensing, images captured by a satellite may not have sufficient resolution for classifying different objects/events. Likewise, every industry needs HR images for different applications.

Most of the CNN-based methods do not consider the similar information across images. Further, the significance of sparse representation is seldom considered in the SR task. In this proposed method, our goal is to combine the advantages of classical methods such as non-local and sparse representation with deep learning techniques. To retain the global modeling ability of the non local similarity, with the efficiency and robustness we try to impose sparse representation with non-local attention. Further, we suggest a channel attention (CA) mechanism for adaptable rescaling of each channel-wise feature. This CA mechanism enables our proposed network to focus on more relevant channel-wise features while improving discriminative learning performance. Indicatively, we try to embed the non-local sparse attention and residual channel attention in a CNN-framework to super-resolve a single image.

1.2 Objective

Our work aims to implement a deep learning model for generating high-resolution image from its low-resolution counterpart. The focus is significantly on residual connections and attention networks. Our objectives are

- To combine the advantages of classical techniques such as non-local similarity and sparse representation with deep learning.
- To enable CNN based networks to explore non-local information similarity in an image with sparsity.
- To allow a smoother flow of information from input to output in a residual channel attention framework.
- To investigate different attention mechanisms like channel attention, pixel attention, spatial attention, and non-local sparse attention.
- To reduce computational cost by utilizing sparse representation in non-local attention.

1.3 Contribution of Thesis

In the field of Image Super-Resolution, we provide a comprehensive overview and discussion of the recent techniques. In order to fulfil the mentioned objectives, the following key contributions are made:

- Non-local sparse attention block is employed in CNN-framework such that the network can explore the non-local similarity in an image using sparse representation.
- The sparsity can reduce the time requirement of non-local operations.
- We induce a channel attention mechanism to assign weights to different channels according to their importance.
- We analyze different attention mechanism and combination of them to solve the SR task.
- The proposed model architectures are evaluated qualitatively and quantitatively on the five benchmark test datasets for different scale factors.

1.4 Organization of Thesis

In the remaining thesis, Chapter 2 highlights a few existing deep learning architectures for SISR (Single Image Super-Resolution) task. Further it briefly summarizes the sparse representation framework along with a few image quality assessment metrics. Chapter 3 discusses the proposed method based on non-local sparse attention and channel attention. In Chapter 4, we discuss the experiments and results of the existing method and proposed method. This chapter also includes the other experiments done to solve the SISR task. Chapter 5 concludes the paper with some key points that can be addressed in future.

CHAPTER 2

Literature Survey and Background

2.1 Deep Learning Architectures for SISR task

A technical overview of super resolution is provided by [20]. Many recent papers have used deep CNNs to address the SISR issue due to their excellent feature representation capabilities. The Image Super Resolution using Convolutional Neural Network(SRCNN)[3]'s innovative, deep application of single image super-resolution is a CNN - based technology to address SISR, a three-layer super-resolution convolution network. After SRCNN effectively implemented the Deep Learning Network for a super-resolution task, the researchers proposed a variety of deeper and more effective models. Very Deep SR (VDSR)[9] enlarges the depth of the network by assembling more convoluted layers with the residual practice. They tried to collect hierarchical features using dense residual connections with the residual learning for better results. Attention processes in the Deep Neural Network help focus on key data while ignoring unwanted data to the network. Over the past few years, deep CNN-based image-enhancing techniques have been effectively implemented with attention mechanisms. The Residual Channel Attention Network ([28]RCAN) allows the network to concentrate on more informative channels using the Residual Channel Attention Block (RCAB). Channel Attention considers every convolution layer as a separate function, ignoring their correlation.

2.1.1 SRCNN

The Image Super Resolution using Convolutional Neural Network(SRCNN)[3] is the first deep neural network architecture in this SR task. It is mainly divided in three parts as given in its architecture described in the Figure 2.1:

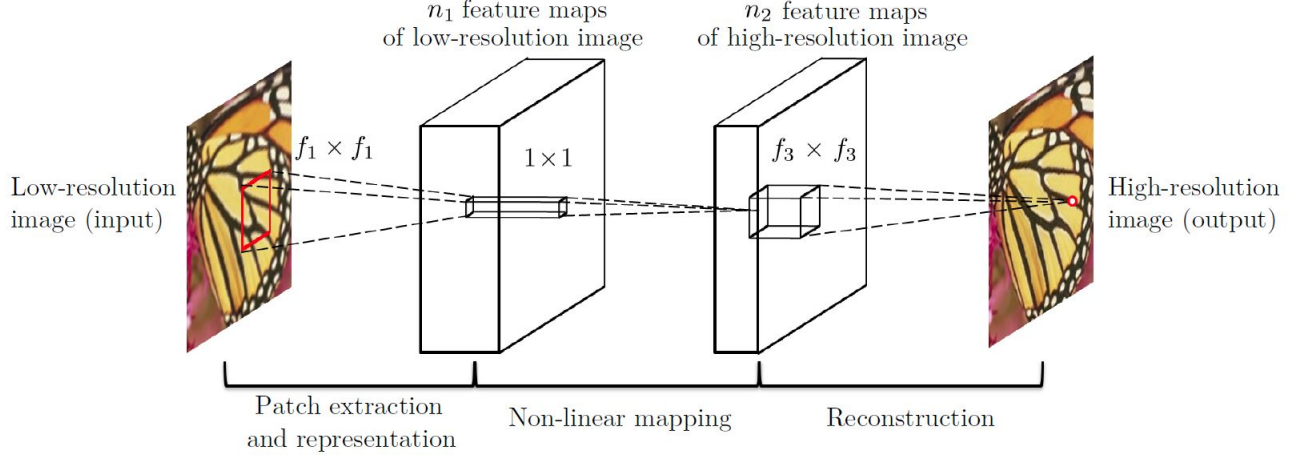


Figure 2.1: Architecture of Super-Resolution Convolutional Neural Networks [3]

1. **Patch Extraction and Representation:** SRCNN is the preupsampling method of SR task. So here the upscaling of the low-resolution image is done first. Then we extract features $F_1(Y)$ using a convolution with ReLU.

$$F_1(Y) = \max(0, W_1 * Y + B_1) \quad (2.1)$$

In this case, X is a high-resolution ground truth image, and Y is bicubic up-sampled image low-resolution image. The size of W_1 is $c \times f_1 \times f_1 \times n_1$, where c is the number of channels, f_1 and n_1 are the size of the filters and the number of filters, respectively, and B_1 is an n_1 dimensional bias vector.

2. **Non-linear mapping:** Now, we have features $F_1(Y)$ and we need to perform non-linear mapping on this $F_1(Y)$ by doing the given computation:

$$F_2(Y) = \max(0, W_2 * F_1(Y) + B_2) \quad (2.2)$$

It is a mapping of an n_1 -dimensional vector to an n_2 -dimension vector. When $n_1 > n_2$, we can imagine it as dimensionality reduction like PCA with non-linearity. 1×1 convolutions are performed to introduce more non-linearity to improve accuracy. A similar approach is also applied in GoogLeNet to introduce non-linearity and reduce the number of convolutions. It is here to map low-resolution vectors to high-resolution vectors.

3. **Reconstruction Process:** After the non-linear mapping, reconstruction is re-

quired. For that, also we will do convolution:

$$F(Y) = W_3 * F_2(Y) + B_3 \quad (2.3)$$

In this network, we are using the conventional loss function average of mean squared error for training, which can be computed as follow:

$$L(\Theta) = \frac{1}{n} \sum_{i=1}^n \|F(Y_i; \Theta) - X_i\|^2 \quad (2.4)$$

2.1.2 RCAN (Residual Channel Attention Network)

RCAN[28] uses a very deep learning model and inter-dependencies among channels to improve SISR task results. A residual in residual (RIR) module comprises several residual groups connected by long skip connections to construct a very deep network. In addition, each residual group contains residual blocks with short skip connections. A channel attention approach is suggested to rescale channel-wise characteristics by adapting channel inter-dependencies. RCAN is made up of four parts, as seen in architecture Figure 2.2:

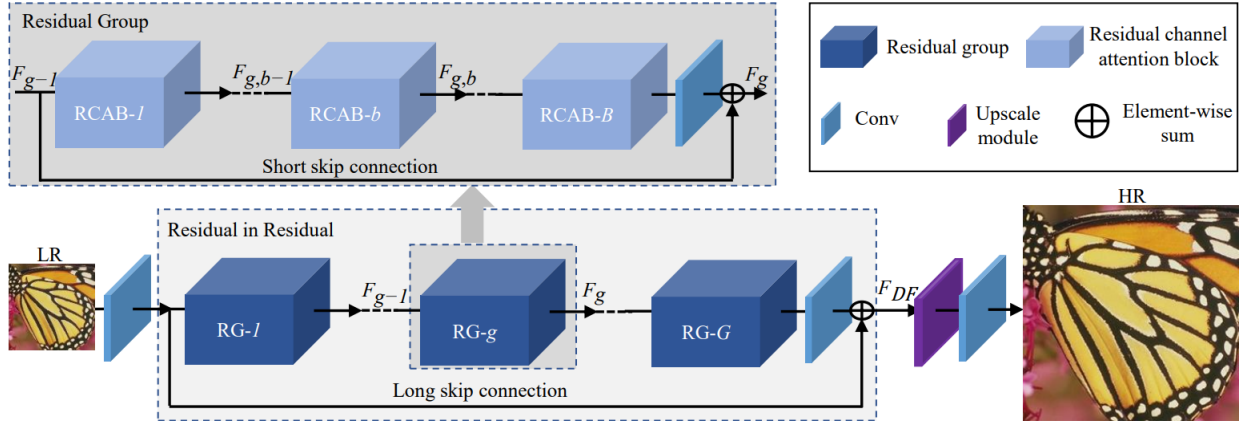


Figure 2.2: Architecture of Residual Channel Attention Network [28]

1. **Shallow Feature Extraction:** Here, we use one convolutional layer for obtaining the shallow features F_0 from the given LR input as given below, where H_{SF} is shallow feature extraction function, i.e., convolution function.

$$F_0 = H_{SF}(I_{LR}) \quad (2.5)$$

2. **Residual in Residual (RIR):** RIR is also known as Deep Feature Extraction.

The RIR module as shown in Figure 2.2 uses the previous feature for deep feature extraction in this step using the long skip connection. H_{RIR} stands for the proposed RIR configuration containing G residual groups (RG).

$$F_{DF} = H_{RIR}(F_0) \quad (2.6)$$

This presented RIR for deep feature extraction could achieve significant depth to date while also providing a large receptive field using B Residual Channel Attention Blocks as described in the Figure 2.2 which is using short skip connections. The architecture of the Residual Channel Attention Block (RCAB), which consists of channel attention mechanisms, is shown in Figure 2.3.

- (a) Channel Attention: In this experiment main two concerns are: First off, there are a lot of low-frequency and valuable high-frequency components in the LR space. Low-frequency components appear to be more complex. Usually, regions with their many edges, textures, and other details make up the high-frequency components. However, each filter in the Convolution layer uses a local receptive field to operate. As a result, the output of convolution is limited in its ability to use context outside of the immediate region. From this analysis we use global average pooling to transform the channel-wise global spatial information into a channel descriptor.

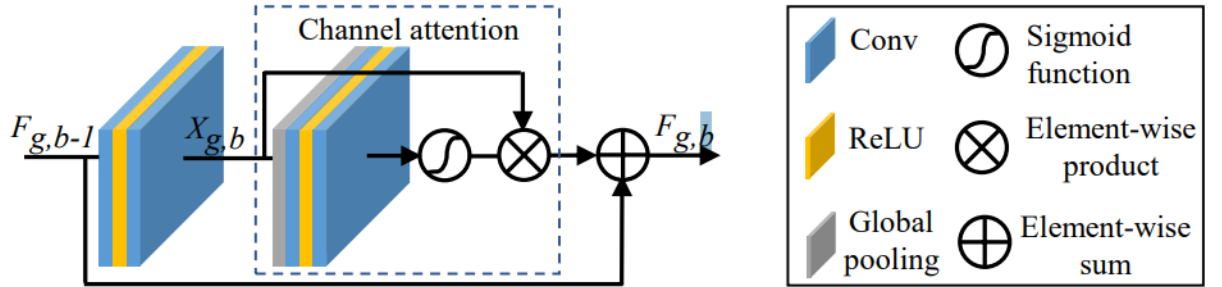


Figure 2.3: Architecture of Residual Channel Attention Block

3. **Upscale module and reconstruction part:** We upscale the deep features generated from RIR module and these upscaled features are then reconstructed by one convolution layer,

$$F_{UP} = H_{UP}(F_{DF}) \quad (2.7)$$

$$I_{SR} = H_{REC}(F_{UP}) = H_{RCAN}(I_{LR}) \quad (2.8)$$

In above computation, H_{REC} and H_{RCAN} stands for the reconstruction layer and the function of the RCAN, respectively.

In this network L_1 loss function is used as given below:

$$L(\Theta) = \frac{1}{N} \sum_{i=1}^N \|H_{RCAN}(I_{LR}^i) - I_{HR}^i\|_1 \quad (2.9)$$

2.1.3 NLSN (Non Local Sparse Attention Network)

The non-local prior is yet another extensively used image prior. Prior to minor patterns recurring inside the same image, SISR employing Non-Local Attention(NLA) could be a noticeably more common approach for using image self-similarity. The non-local operation looks globally for comparable patterns and aggregate those connected features selectively to improve the representation. Although NLA is perceptive and appealing in fusing characteristics, using it in the SISR job will raise several overlooked issues i.e., i) The receptive field of features in deeper layers tends to be global; therefore, mutual correlation calculation across deep features is not accurate. ii) Calculating feature similarity across every pixel locations is essential for global non-local attention. As a result, we get quadratic computational cost to image size. One option for addressing the above mentioned issues is to confine searching range of non-local operation inside a local neighborhood. However, it lowers commuting costs by missing out on a lot of global data. The architecture of NLSN network is given in figure 2.4.

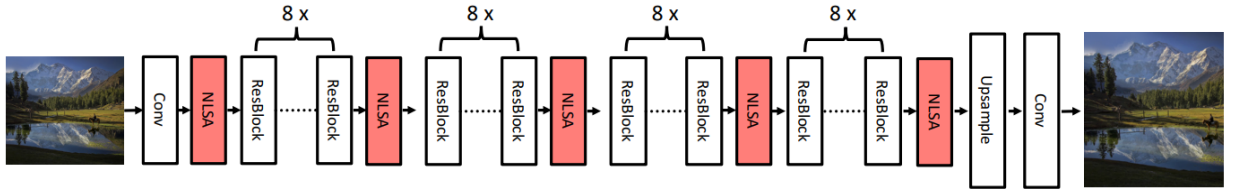


Figure 2.4: Architecture of Non Local Sparse Attention Network[18]

The presented NLSA will allow the computational cost of non-local to be reduced from quadratic to asymptotic linear in terms of spatial dimensions. Searching for similar cues inside a narrower content-correlated bin will also direct the module's attention to more informative and related locations. As a result, NLSA maintain the conventional non-local operation's global modelling capability while benefiting from the resilience and efficiency of its sparse representation.

2.2 Sparse Representation

Suppose we have an overcomplete dictionary $D^{d \times n}$ ($d < n$) of $x_1, x_2, \dots, x_n \in R^d$ i.e., n known examples. So, a query $y \in R^d$, is represented as a weighted sum of a elements in D:

$$y = \alpha_1 x_1 + \alpha_2 x_2 + \dots + \alpha_n x_n \quad (2.10)$$

here, α_i is a coefficient with reference to x_i . Above equation is rewritten as:

$$y = D\alpha \quad (2.11)$$

where $\alpha = [\alpha_1, \alpha_2, \dots, \alpha_n]$. In this equation finding α is an ill-posed problem. To solve this problem, in sparse representation we assume that y need to be sparsely represented which implies that α is going to be sparse.

$$y_i = D \alpha \text{ s.t. } \|\alpha\|_0 \leq k \quad (2.12)$$

Here k and $\|\cdot\|_0$ bounds and counts the number of non-zero elements of α , respectively.

2.3 Image Quality Assessments

Image quality measures an image's visual features, including its perceptual quality. Image quality assessment categorizes into objective and subjective computational techniques. Subjective techniques are concerned with human perception (i.e., realistic images). Although they are not very consistent, objective techniques are currently in use, as they frequently fail to reflect human perception accurately. We will discuss some of the most often used iqa methods and techniques here.

2.3.1 Peak Signal-to-Noise Ratio (PSNR)

The peak signal-to-noise ratio (PSNR) is an image quality statistic that describes the ratio between an image's or signal's maximum achievable strength and the power of corrupting noise that influences the quality and representation of an image. We can estimate a picture's PSNR by comparing it to an ideal clean image with the highest potential power. The ideal image is the ground truth image in the SR problem, and the predicted image is the SR reconstructed image. One of the most widely used image quality measurements focused on image reconstruction

quality. We can give the PSNR of the ground truth picture I with N pixels and the reconstructed image S as:

$$PSNR = 10 \cdot \log_{10} \left(\frac{L^2}{\frac{1}{N} \sum_{i=1}^N (I(i) - \hat{I}(i))^2} \right) \quad (2.13)$$

For an 8-bit representation of an image, L is 255, which is the most typical case. One disadvantage of utilizing PSNR is that it only pertains to the pixel-level difference using MSE. It disregards the visual perception of an image, which frequently results in poor performance in determining the reconstruction quality of an image or real scene, where we are more concerned with the human perception and perceptual quality of an image. Despite this, the most generally used image quality metric is still the PSNR, partly due to a lack of entirely accurate perceptual metrics.

2.3.2 Structural Similarity Index (SSIM)

The human vision system is more sensitive to structural and abstract image details[43]. The structural similarity index (SSIM) [23] is proposed to account for the structural similarity between images. The structural similarity of an image is determined via separate comparisons of its three components, which are known as structures, contrast, and luminance. If an image is represented as I with a certain number of pixels (N), the luminance μ_I and contrast σ_I can be calculated as the mean and standard deviation of the image intensity, i.e.,

$$\mu_I = \frac{1}{N} \sum_{i=1}^N I(i)$$

$$\sigma_I = \left(\frac{1}{N-1} \sum_{i=1}^N (I(i) - \mu_I)^2 \right)^{\frac{1}{2}}$$

The i^{th} pixel intensity of an image is represented by I_i . On the other side, we can see the luminance and contrast comparisons, which are denoted by C_l and C_c .

$$C_l(I, \hat{I}) = \frac{2\mu_I\mu_{\hat{I}} + C_1}{\mu_I^2 + \mu_{\hat{I}}^2 + C_1}$$

$$C_c(I, \hat{I}) = \frac{2\sigma_I\sigma_{\hat{I}} + C_2}{\sigma_I^2 + \sigma_{\hat{I}}^2 + C_2}$$

Furthermore, the structural component of an image is represented by the normal-

ized value of pixels, and their correlation values are inner products that assess structural similarity, which is equal to the correlation coefficient between the reconstructed images and ground truth. The structural comparison function C_s is defined as:

$$\sigma_{I\hat{I}} = \frac{1}{N-1} \sum_{i=1}^N (I(i) - \mu_I) (\hat{I}(i) - \mu_{\hat{I}})$$

$$C_s(I, \hat{I}) = \frac{\sigma_{I\hat{I}} + C_3}{\sigma_I \sigma_{\hat{I}} + C_3}$$

The final form of SSIM is gives as:

$$SSIM(I, \hat{I}) = [C_l(I, \hat{I})]^\alpha [C_c(I, \hat{I})]^\beta [C_s(I, \hat{I})]^\gamma$$

Here the α , β and γ are the parameters which are variable.

CHAPTER 3

Proposed Method

As shown in the figure 3.1 in our proposed method's architecture there are two main components

1. Non Local Sparse Attention(NLSA)
2. Residual Group(RG)

The first component is used to embrace the long-range features with minimizing the complexity. And the second component contains residual channel attention blocks to extract high frequency details of an image and improve discriminative capability of the network:

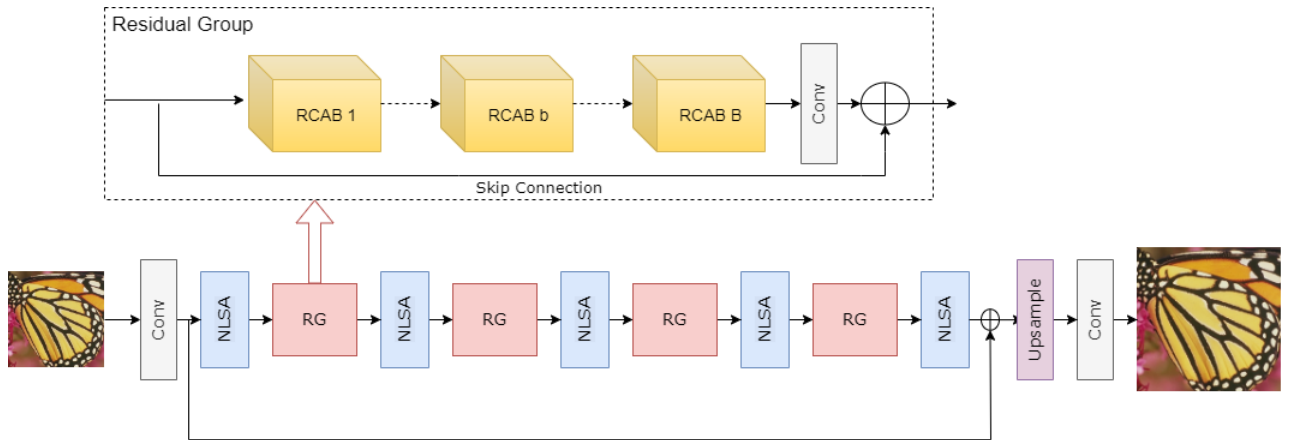


Figure 3.1: Proposed Method Architecture

3.1 Non Local Sparse Attention

To understand non-local sparse attention, we will first discuss about non-local attention and sparsity on non-local attention.

3.1.1 Non Local Attention

The main purpose of non-local attention is to enhance the image by summarizing all the features of an image. For example, let's take an input feature $X \in R^{h \times w \times c}$, now reshape it into an one dimensional feature $X \in R^{n \times c}$ where $n = hw$. The output $y_i \in R^c$ generated as:

$$y_i = \sum_{j=1}^n \frac{f(x_i, x_j)}{\sum_{\hat{j}=1}^n f(x_i, x_{\hat{j}})} g(x_j) \quad (3.1)$$

The above equation is of non-local operations, where $x_i, x_j, x_{\hat{j}}$ are pixel-wise feature at the respective location that is i, j, \hat{j} on X . Here f is for mutual-similarity and g is the function for feature transformation which is computed as follow.

$$f(x_i, x_j) = e^{\theta(x_i)^T \phi(x_j)} = e^{(W_\theta x_i)^T W_\phi x_j}$$

$$g(x_j) = W_g x_j$$

Here, W_θ, W_ϕ, W_g are weight matrices, which means θ and ϕ are learned linear projections.

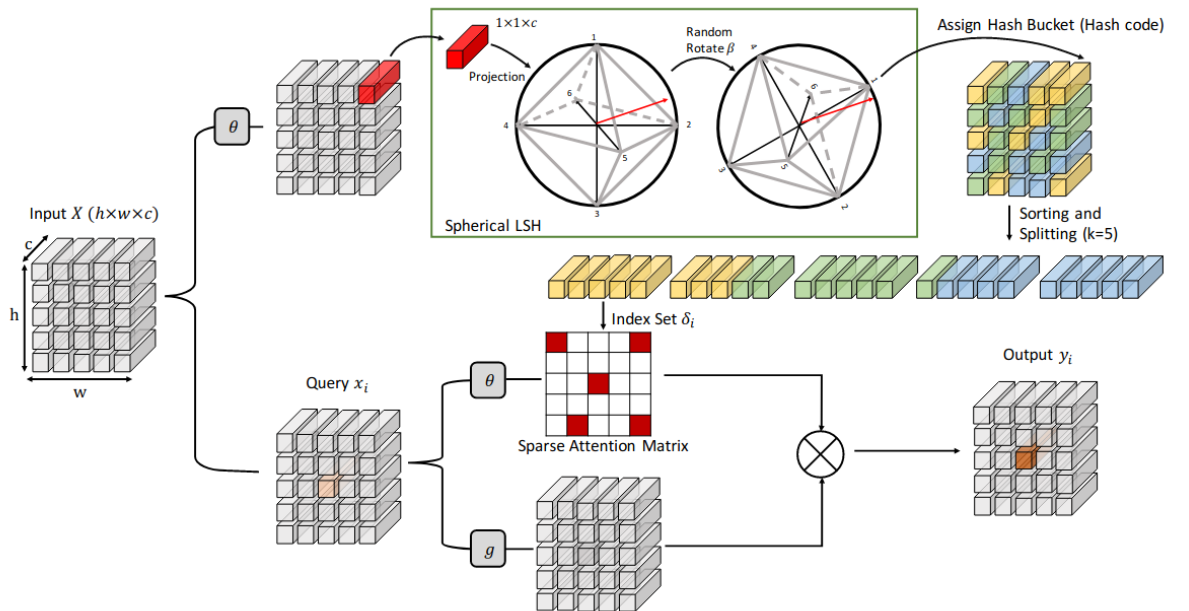


Figure 3.2: Non Local Sparse Attention[18]

3.1.2 Non-local Attention with sparsity constraint

When we try to use non-local attention, it comes with the limitations of the searching range so to concur this limitation we applied sparsity constraint on the non-local attention. The equation 3.1 can also be seen as a sparse representation from equation 2.11 with the substitution as $D = [g(x_1), \dots, g(x_n)] \in R^{c \times n}$ and $\alpha_i = [f(x_i, x_1), \dots, f(x_i, x_n)] \in R^n$, which is, $y_i = D\alpha_i$. According to Equation 3.1, sparsity constraint on non-local attention can be applied by reducing the number of non-zero values of α to a constant k . As a result, the general version of non-local attention with sparsity constraint may emerge as follows:

$$y_i = \sum_{j \in \delta_i} \frac{f(x_i, x_j)}{\sum_{\hat{j} \in \delta_i} f(x_i, x_{\hat{j}})} g(x_j) \quad (3.2)$$

$$y_i = D\alpha_i \text{ s.t. } \|\alpha_i\|_0 \leq k \quad (3.3)$$

Here δ_i indices non-zero elements of the α_i , i.e., $\delta_i = \{j | \alpha_i[j] \neq 0\}$, $\alpha_i[j]$ is j^{th} element in α_i . This δ_i indicates pixel location's group where the query should attend. This δ_i contains the identified locations from which we can calculate non-local attention, these group can be known as Attention Bin.

3.1.3 Attention Bin using LSH(Locality Sensitive Hashing)

A target attention should be sparse as well as include the most significant elements. We can use Locality Sensitive Hashing(LSH) to create the desirable attention bin, which includes global and correlating components as well as the query element. The hashing scheme is locality sensitive if nearby elements are at high possibility to fall into the same hash bin (hash code) whereas distant ones are not. The spherical LSH is an instance of LSH designed for angular distance. We can intuitively think it as randomly rotating a cross-polytope inscribed into a hyper-sphere, as shown in the top branch of Figure 3.2. The hash function projects a tensor onto the hyper-sphere and the closest polytope vertex is selected as its hash code. Thus, if two vectors have a small angular distance, they are likely to fall in the same hash bin, which is also the defined attention bin.

To obtain h hash bins, we must first take projection of targeted tensor onto one hyper-sphere(because we will use spherical LSH) and then randomly rotate that with a matrix $M \in R^{c \times h}$, a sample random rotation matrix with independent and

identically distributed Gaussian entries, that is,

$$\hat{x} = M\left(\frac{x}{\|x\|_2}\right) \quad (3.4)$$

The hash bin is determined as $hb(x) = \operatorname{argmax}_i(\hat{x})$. After hashing every elements, we are able to splitting up the space into bins of correlated elements, and the index set $\delta_i = \{j | hb(x_j) = hb(x_i)\}$ can identify the attention bin of x_i . In reality, the spherical LSH is computed for all elements at the same time using batch matrix multiplication, which adds just a minor computing overhead. By disregarding other noisy as well as less-correlated partitions and determining which bin to attend in advanced, the model can attain high resilience and efficiency.

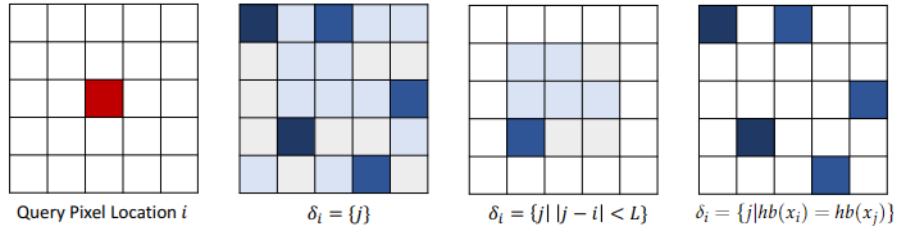


Figure 3.3: Example of attention bins[18].

The attention bin is formed by the darker blue portions in the illustration given in Figure 3.3. In this figure, For query location at i and an index set δ_i the attention bin is formed by the darker blue portions in the illustration. Given a query at i location and an index set δ_i . Here the δ_i decides the group of locations to compute the non-locally fused features. As given in the figure darker blue regions form the attention bin. In the figure, $\delta_i = \{j\}$ denotes the use of full-range pixels, as in typical non-local attention and $\delta_i = \{j | |j - i| < L\}$ denotes a limited attention span in the nearby neighbourhood. While the proposed hash-based attention bin is $\delta_i = \{j | hb(x_i) = hb(x_j)\}$.

After determining the attention bin index set δ_i for the query location i the suggested NLSA can be simply obtained from Equation 3.2. Furthermore, as illustrated in Figure 3.2, NLSA assigns each one of the pixel-wise feature in X to a bin with the similar hash code depending its content relevance, and only the elements of the related bin contribute to the output.

3.2 Residual Groups

As shown in the architecture diagram the residual group consists of multiple Residual Channel Attention Blocks(RCAB). The architecture of RCAB is in given in Figure 3.4. The channel attention directly faces the input data to produce a feature with weights highlighting channel-wise important features. Here, using global average pooling, we convert the channel-wise global spatial features into a channel descriptor. After the success of Residual blocks we try to impose channel attention with the residual block. Further, short skip connections in Residual group enable a smoother information flow from the input towards output. Additionally, it assists in addressing the over-fitting issue, which is often encountered by deep-learning models due to lesser data.

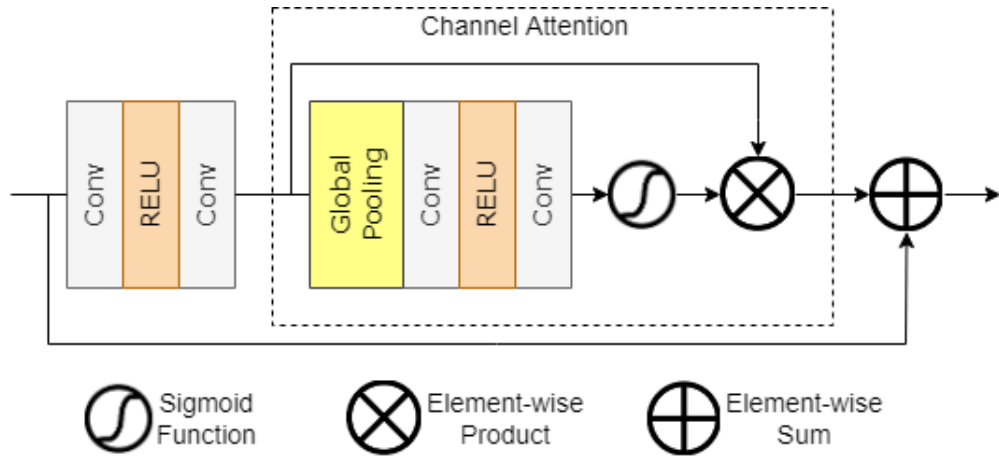


Figure 3.4: Residual Channel Attention Block

3.3 Loss Function

In the proposed method we used L_1 reconstruction loss to train the network. Given a training set $\{I_{LR}^i, I_{HR}^i\}_{i=1}^N$, which contains N LR inputs and their HR counterparts. The goal of training the model is to minimize the L_1 loss function.

$$L(\Theta) = \frac{1}{N} \sum_{i=1}^N \|H_{model}(I_{LR}^i) - I_{HR}^i\|_1 \quad (3.5)$$

Here H_{model} represents the output from the proposed method architecture.

CHAPTER 4

Experiments and Results

4.1 Dataset

Recently, numerous new datasets in Image super-resolution have been introduced, each with a specific number of images, quality, and resolution. Handful datasets supply the HR image, whereas the LR image is built using bicubic interpolation, and only some of them also have the LR-HR image pairings. We have compiled a list of some of the most common and widely utilized image datasets in the image super-resolution task.

Dataset	Count	Purpose
DIV2K[22]	800	Training
Set5[1]	5	Testing
Set14[26]	14	Testing
B100[16]	100	Testing
Urban100[8]	100	Testing
Manga109[17]	109	Testing

Table 4.1: List of Datasets used for SISR task

Different datasets can be used for the image super-resolution benchmark, but these are the most prevalent. In addition, people employ a combination of multiple datasets to generate their training dataset by integrating photos from other datasets. Image data augmentation is a well-known and extensively used technique for artificially increasing the size of a training dataset. It generates a changed version of a single image, and the procedure can be performed for an entire image dataset. Training deep learning models and models on vast amounts of data may yield more resilient and generic models; augmentation approaches generate distinct versions of images depending on various transformations such as flip, rotation, shear, etc.

We trained our network on DIV2K[22] dataset, which consist of 800 training images to train the network. For testing we used the most popular benchmark test dataset for image super-resolution task i.e., Set5[1], Set14[26], Urban100[8], B100[16], and Manga109[17].

4.2 Training and Implementation details

We are using attention bins here, so for training we set the number of bins to 144. In the network we are using 5 non-local sparse attention blocks and 4 Residual Groups which contains 8 RCABs each. We used randomly cropped patches of size 48×48 as the training. To optimize the model we used ADAM[11] optimizer with the parameters $\beta_1 = 0.9$, $\beta_2 = 0.99$ and $\epsilon = 10^{-8}$. This architecture is implemented using pyTorch and trained on Tesla T4 GPU.

4.3 Results and Comparison

For evaluating the effectiveness of our network we examine the results of the proposed network with the state-of-the-art methods like,

1. SRCNN[3]: The First ever method to use deep learning method for SISR task.
2. VDSR[9]: Very deep learning architecture for Image Super-resolution.
3. EDSR[14]: The network which removes unnecessary modules from conventional residual networks and gives effective super-resolution image.
4. NLRN[15]: It is a Non Local Recurrent Network for image restoration which incorporates non local operation with the RNN(Recurrent Neural Network).
5. RNAN[29]: Residual Non Local Attention Network for image restoration.
6. SRFBN[13]: SRFBN(Super-resolution feedback network) is proposed to fine tune low-level representations with high-level information.
7. RDN[30]: Residual Dense Network(RDN) uses dense skip connections for SISR task.
8. RCAN[28]: Residual Channel Attention Network which uses channel attention mechanism with short and long skip connections for SISR task.

9. NLSN[18]: In this Non Local Sparse Network sparsity constraint is used with non local attention for image super resolution.

The results for the same are given in the Table 4.2 and 4.3, where the bold values are the highest result in that particular dataset and underlined result is the second highest result in that column.

As we can notice in Table 4.2 our method gives better results than the most of the existing methods like SRCNN[3], VDSR[9], EDSR[14], NLRN[15], RNAN[29], SRFBN[13], and RDN[30] in scale 2 in most of the datasets. And also in scale 3 it can get better results from the existiting methods like SRCNN[3], VDSR[9], EDSR[14], NLRN[15] and RNAN[29]. Table 4.3 shows the quantitative results of scale 4 of the existing methods and our method, we can notice that we get better results from some of the existing methods.

The qualitative performance is shown from Figure 4.1 to 4.4. Figure 4.1 shows the scale 2 results of the butterfly image of Set5. We can observe the edges of the cropped portion of an image. The edges of bicubic interpolated image is not clear, also in RCAN's visual results we can notice some artifacts around the edges, while in our results we can notice that the visual results are mostly similar as the results of NLSN[18] architecture.

For scale 3 we have chosen a comic.png from Set14 dataset which is shown in Figure 4.2. We can observe that visual result with our proposed model is better and visually appealing than the existing methods.

In scale 4, Figures 4.3 and 4.4 shows the qualitative results from Urban100 dataset's img_002.png and img_093.png.

Method	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	×2	33.66	0.9299	30.24	0.8688	29.56	0.8431	26.88	0.8403	30.80	0.9339
SRCNN[3]	×2	36.66	0.9542	32.45	0.9067	31.36	0.8879	29.50	0.8946	35.60	0.9663
VDSR[9]	×2	37.53	0.9590	33.05	0.9130	31.90	0.8960	30.77	0.9140	37.22	0.9750
EDSR[14]	×2	38.11	0.9602	33.92	0.9195	32.32	0.9013	32.93	0.9351	39.10	0.9773
NLRN[15]	×2	38.00	0.9603	33.46	0.9159	32.19	0.8992	31.81	0.9249	-	-
RNAN[29]	×2	38.17	0.9611	33.87	0.9207	32.32	0.9014	32.73	0.9340	39.23	0.9785
SRFBN[13]	×2	38.11	0.9609	33.82	0.9196	32.29	0.9010	32.62	0.9328	39.08	0.9779
RDN[30]	×2	38.24	<u>0.9614</u>	34.01	0.9212	32.34	<u>0.9017</u>	32.89	0.9353	39.18	0.9780
RCAN[28]	×2	<u>38.27</u>	<u>0.9614</u>	34.12	<u>0.9216</u>	<u>32.41</u>	0.9027	<u>33.34</u>	<u>0.9384</u>	<u>39.44</u>	<u>0.9786</u>
NLSN[18]	×2	38.34	0.9618	<u>34.08</u>	0.9231	32.43	0.9027	33.42	0.9394	39.59	0.9789
Our Method	×2	38.25	0.9613	33.91	0.9204	32.29	0.9008	32.74	0.9338	39.32	0.9784
Bicubic	×3	30.39	0.8682	27.55	0.7742	27.21	0.7385	24.46	0.7349	26.95	0.8556
SRCNN[3]	×3	32.75	0.9090	29.30	0.8215	28.41	0.7863	26.24	0.7989	30.48	0.9117
VDSR[9]	×3	33.67	0.9210	29.78	0.8320	28.83	0.7990	27.14	0.8290	32.01	0.9340
EDSR[14]	×3	34.65	0.9280	30.52	0.8462	29.25	0.8093	28.80	0.8653	34.17	0.9476
NLRN[15]	×3	34.27	0.9266	30.16	0.8374	29.06	0.8026	27.93	0.8453	-	-
RNAN[29]	×3	34.66	0.9290	30.52	0.8462	29.26	0.8090	28.75	0.8646	34.25	0.9483
SRFBN[13]	×3	34.70	0.9292	30.51	0.8461	29.24	0.8084	28.73	0.8641	34.18	0.9481
RDN[30]	×3	34.71	0.9296	30.57	0.8468	29.26	0.8093	28.80	0.8653	34.13	0.9484
RCAN[28]	×3	<u>34.74</u>	<u>0.9299</u>	<u>30.65</u>	<u>0.8482</u>	<u>29.32</u>	<u>0.8111</u>	<u>29.09</u>	<u>0.8702</u>	<u>34.44</u>	<u>0.9499</u>
NLSN[18]	×3	34.85	0.9306	30.70	0.8485	29.34	0.8117	29.25	0.8726	34.57	0.9508
Our Method	×3	34.67	0.9290	30.49	0.8439	29.19	0.8067	28.60	0.8601	34.23	0.9480

Table 4.2: Quantitative analysis of different Architectures (Scale 2 & 3)

Method	Scale	Set5		Set14		B100		Urban100		Manga109	
		PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM	PSNR	SSIM
Bicubic	×4	28.42	0.8104	26.00	0.7027	25.96	0.6675	23.14	0.6577	24.89	0.7866
SRCNN[3]	×4	30.48	0.8628	27.50	0.7513	26.90	0.7101	24.52	0.7221	27.58	0.8555
VDSR[9]	×4	31.35	0.8830	28.02	0.7680	27.29	0.0726	25.18	0.7540	28.83	0.8870
EDSR[14]	×4	32.46	0.8968	28.80	0.7876	27.71	0.7420	26.64	0.8033	31.02	0.9148
NLRN[15]	×4	31.92	0.8916	28.36	0.7745	27.48	0.7306	25.79	0.7729	-	-
RNAN[29]	×4	32.49	0.8982	<u>28.83</u>	0.7878	27.72	0.7421	26.61	0.8023	31.09	0.9149
SRFBN[13]	×4	32.47	0.8983	28.81	0.7868	27.72	0.7409	26.60	0.8015	31.15	0.9160
RDN[30]	×4	32.47	0.8990	28.81	0.7871	27.72	0.7419	26.61	0.8028	31.00	0.9151
RCAN[28]	×4	32.63	0.9002	28.87	<u>0.7889</u>	<u>27.77</u>	<u>0.7436</u>	<u>26.82</u>	<u>0.8087</u>	<u>31.22</u>	<u>0.9173</u>
NLSN[18]	×4	<u>32.59</u>	<u>0.9000</u>	28.87	0.7891	27.78	0.7444	26.96	0.8109	31.27	0.9184
Our Method	×4	32.43	0.8973	28.73	0.7853	27.63	0.7372	26.39	0.7927	30.94	0.9125

Table 4.3: Quantitative analysis of different Architectures(Scale 4)

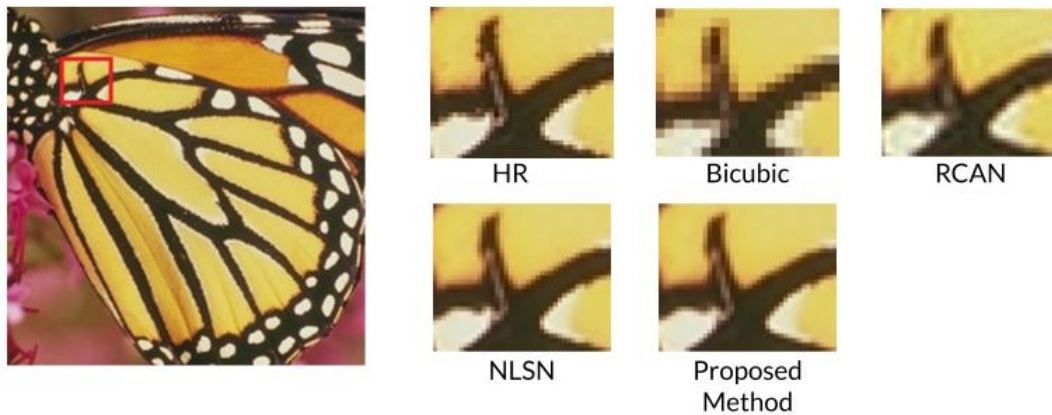


Figure 4.1: Qualitative Results on Scale 2 (Set5: Butterfly.png)

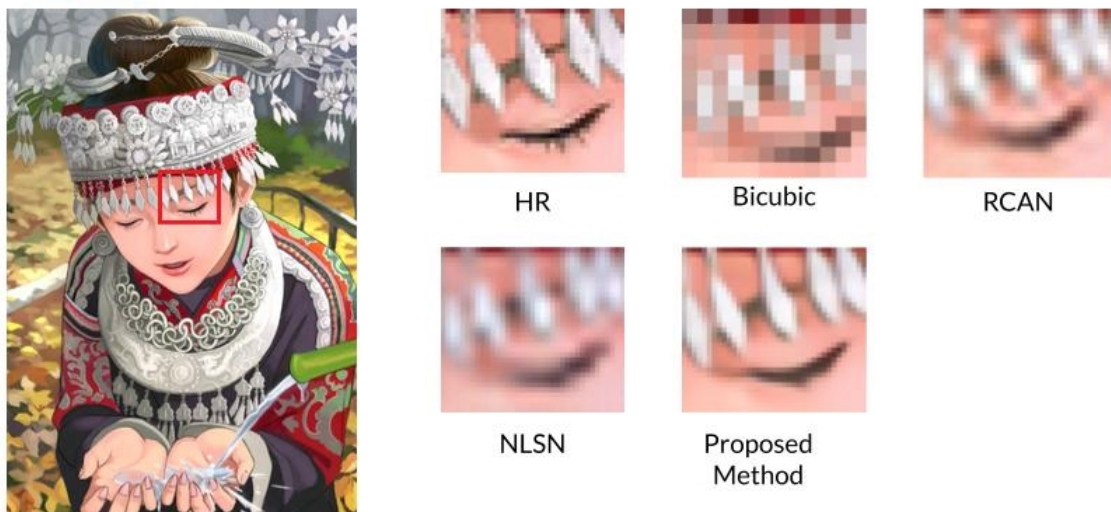


Figure 4.2: Qualitative Results on Scale 3 (Set14: comic.png)

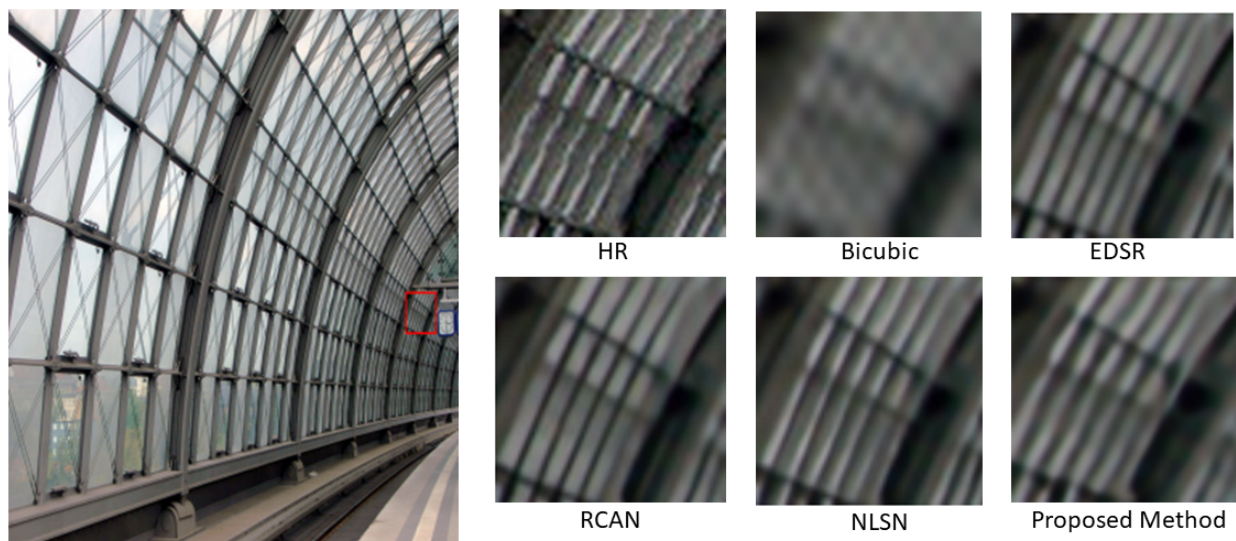


Figure 4.3: Qualitative Results on Scale 4 (Urban100: img_002.png)

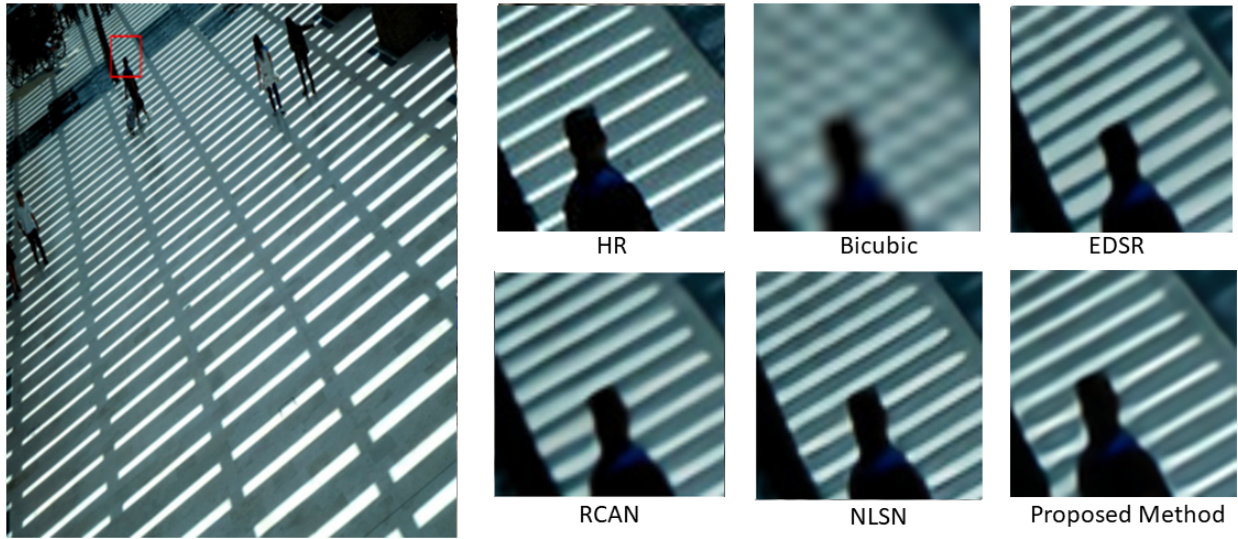


Figure 4.4: Qualitative Results on Scale 4 (Urban100: img_093.png)

4.4 Other Experiments

4.4.1 Other Experiments

Other than the proposed method we have done some experiments which are briefly described below. We first learn about channel attention, pixel attention[31], and spatial attention[24] in brief. As seen in the figure 4.5, channel attention seeks to produce a 1D ($C \times 1 \times 1$) attention feature vector, whereas spatial attention produces a 2D ($1 \times H \times W$) attention map. As opposed to this, pixel attention can produce a 3D ($C \times H \times W$) matrix as its attention features. Therefore, attention coefficients are generated for each pixel in the feature map using pixel attention. Here, C stands for the number of channels, and H and W stand for the features' respective height and width.

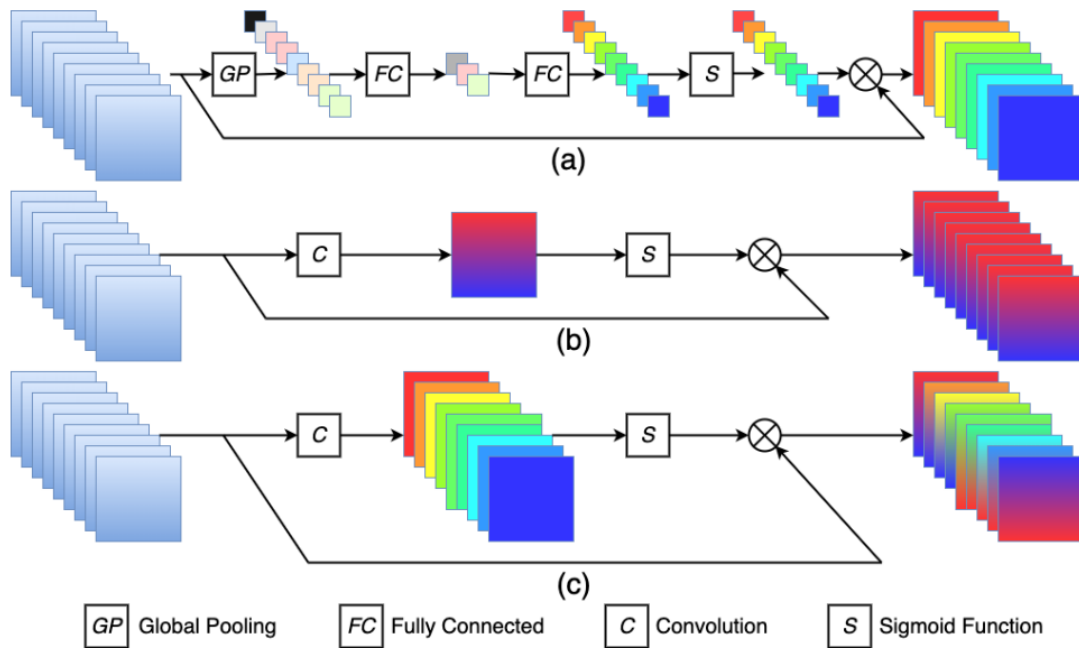


Figure 4.5: (a) CA: Channel Attention; (b) SA: Spatial Attention; (c) PA: Pixel Attention. [31]

4.4.2 RCAN_Dense

From RCAN[28] we notice that in RCAN features are extracted only from the previous RCAB. So for hierarchical feature extraction we applied the residual

skip connections between Residual Channel Attention Blocks in the each Residual Group of the network. The architecture of RCAN dense network is given in the figure 4.6

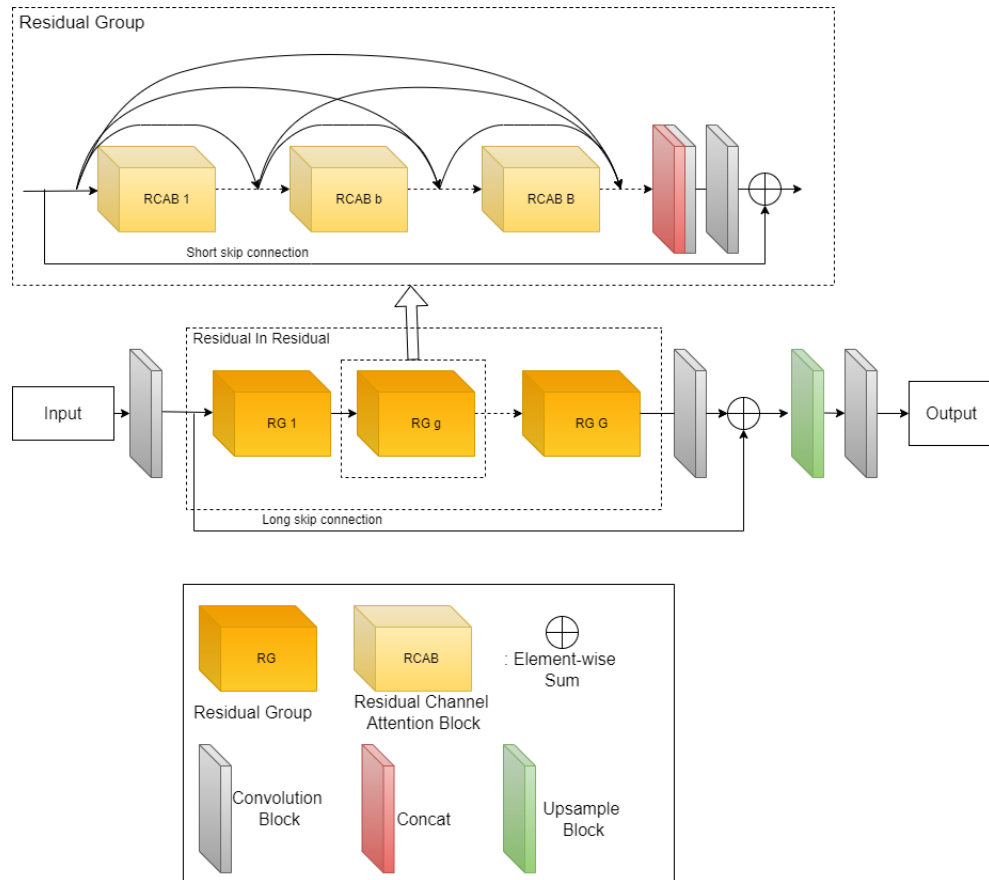


Figure 4.6: RCAN_dense Architecture

4.4.3 RCAPAN(Residual Channel Attention and Pixel Attention Network)

In Residual Channel Attention Network[28] they focused on only channel attention. So for this experiment we tried to incorporate pixel attention with the channel attention. We tried to make residual pixel attention blocks(RPAB) as shown in figure 4.7. Here 10 residual groups are used in which first 10 RCABs are there and then 10 RPABs are there.

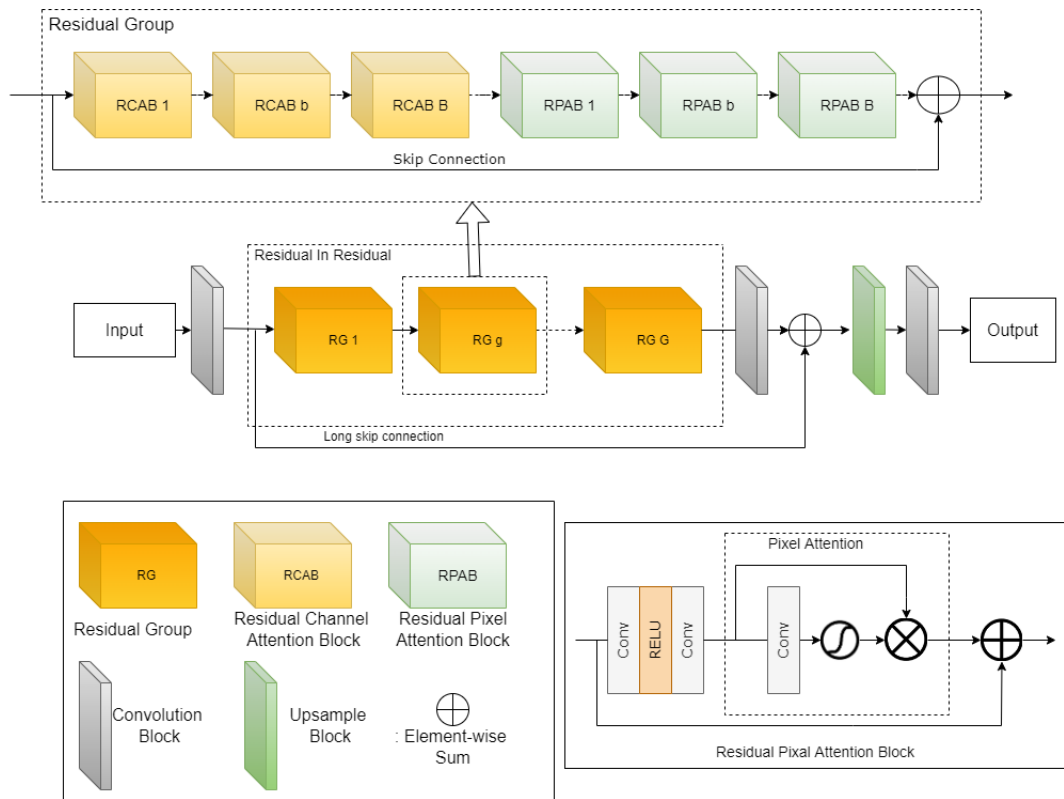


Figure 4.7: RCAPAN Architecture

4.4.4 RPCSN(Residual Channel Pixel Spatial Network)

After the improvement of previous experiment's result we tried to implement spatial attention with channel attention and pixel attention as shown in figure 4.8. We used each block serially in this experiment i.e., first RCAB, then RPAB and then Channel Spatial Attention Blocks(CSAB[19]). The number of this blocks are 7 and the number of residual groups are 10.

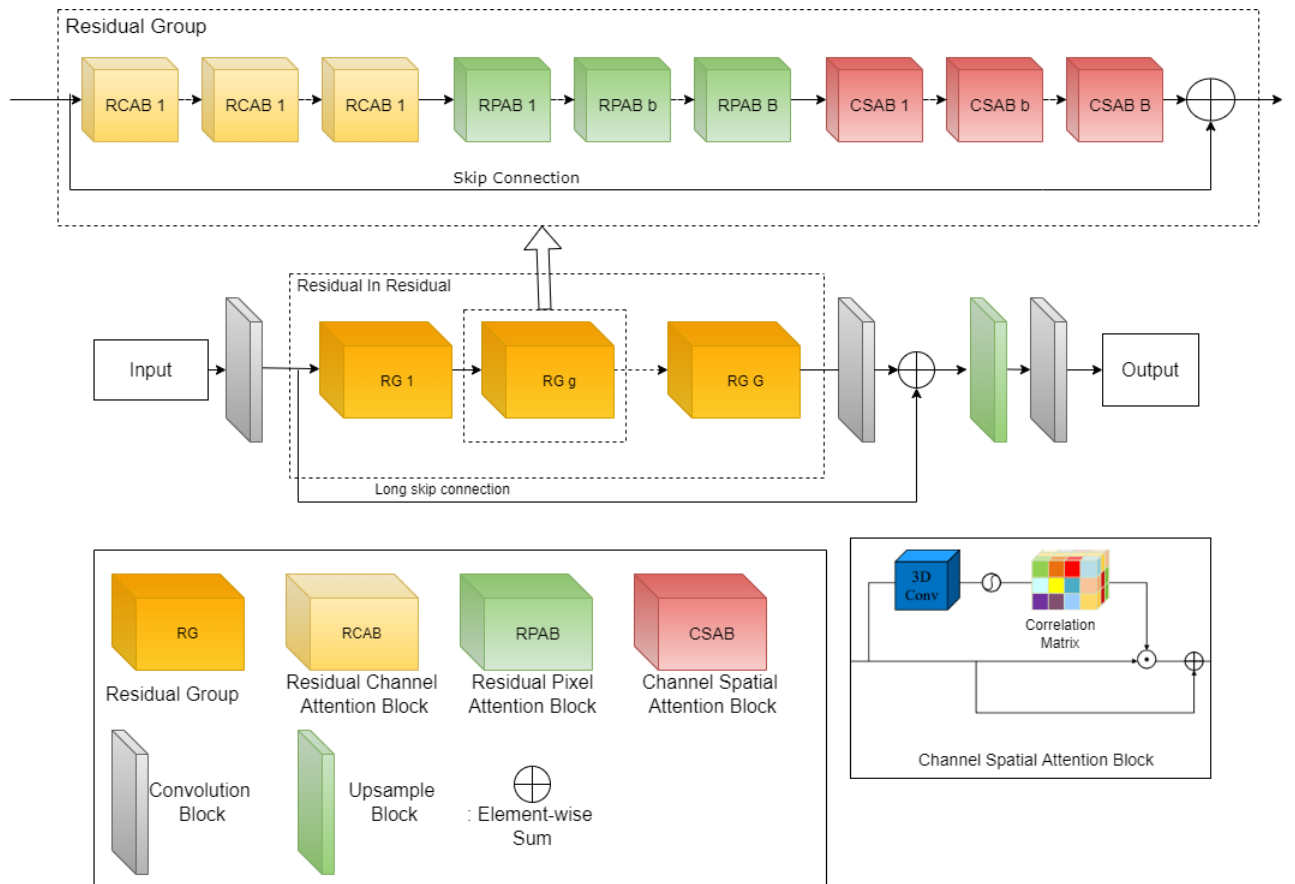


Figure 4.8: RPCS Architecture

4.4.5 Results of Experiments

Results of the experiments are given in table 4.4. These experiments are performed on scale 2.

Dataset	Scale	PSNR	SSIM
RCAN_Dense			
Set5	X2	37.930038	0.960345
Set14	X2	33.444639	0.916403
Urban100	X2	31.915354	0.925503
Manga109	X2	38.174127	0.976865
B100	X2	32.034723	0.901073
RCAPAN			
Set5	X2	38.085357	0.960697
Set14	X2	33.721527	0.918634
Urban100	X2	32.366394	0.930623
Manga109	X2	38.681300	0.977189
B100	X2	32.209183	0.899827
RCPSN			
Set5	X2	38.056365	0.960782
Set14	X2	33.635946	0.917732
Urban100	X2	32.386882	0.930770
Manga109	X2	38.675729	0.977264
B100	X2	32.204666	0.899918

Table 4.4: Other Experiments’ Quantitative results

4.5 Experiments with other Image Modalities

4.5.1 Depth Map

When comprehending a scene, people are able to capture the depth information necessary to produce stereo perception in addition to the scene’s appearance (such as colour and texture). Numerous research areas that depend on high-quality depth data, such as autonomous navigation and 3D reconstruction, can be facilitated by a better understanding of the scene. Portable consumer-grade depth cameras, like Microsoft Kinect and Lidar, have become increasingly common and offer great convenience for quickly determining the depth of a scene. The resolution of a depth map, even when combined with a high-resolution colour image,

is typically constrained due to the imaging limitations of depth cameras. Depth map super-resolution (SR) technique has drawn increasing attention as a potential solution to the urgent need for high-quality depth maps in applications.

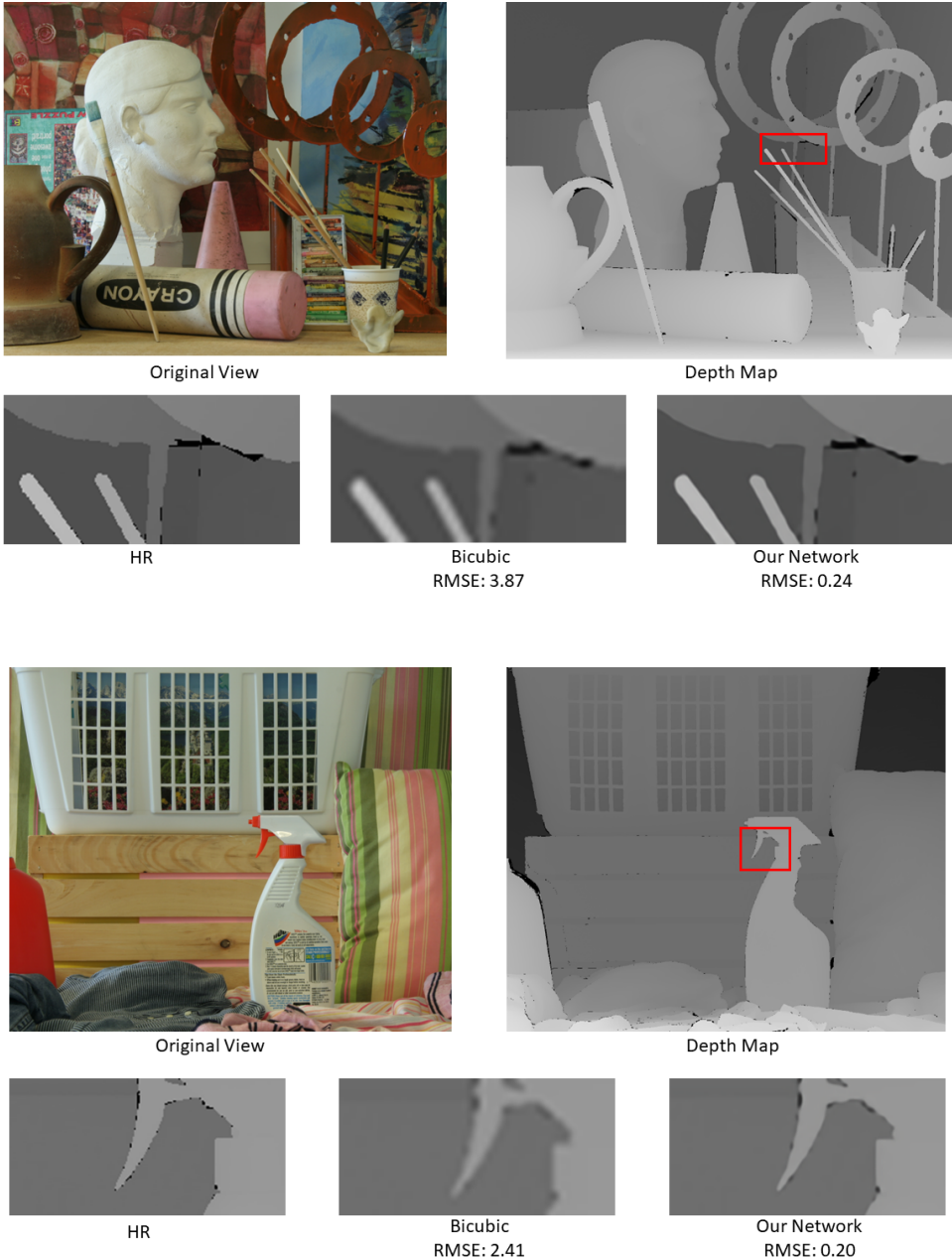


Figure 4.9: Qualitative results of our network on scale 4 on Middlebury 2005 Art and Laundry image[21]

The depth map SR is evaluated using Root Mean Squared Error(RMSE) values.

We tested our network for Middlebury dataset’s Art, Books, Laundry, Reindeer and teddy images we took average of the both depth images present in that set.

Data	Bicubic	Our Network
Art	3.87	0.24
Books	1.61	0.26
Laundry	2.41	0.20
Reindeer	2.81	0.19
Teddy	2.86	0.29

Table 4.5: Quantitative results of Depthmap SR on scale 4 upsampling on Middlebury dataset in terms of RMSE values

4.5.2 X-Ray Images

To perform experiment on medical image dataset, we used COVID-19 image dataset[2] which consists of chest X-ray and computed tomography (CT) images. Electromagnetic waves are a category of radiation that includes X-rays. X-ray imaging produces images of what’s in your body. The images depict the various body parts in various shades of black and white. This is due to the fact that different tissues absorb radiation in different ways. Because calcium in bones absorbs the most x-rays, bones appear white. Fat and other soft tissues have a grey appearance and absorb less. Lungs appear black because air absorbs the least. The dataset is updated frequently, and it’s important to note that each image’s resolution varies. The results of our network on the some images of COVID-19 image dataset [2] are given in the figure 4.10. The quantitative results of these 4 images’ average in PSNR are given in the 4.6.

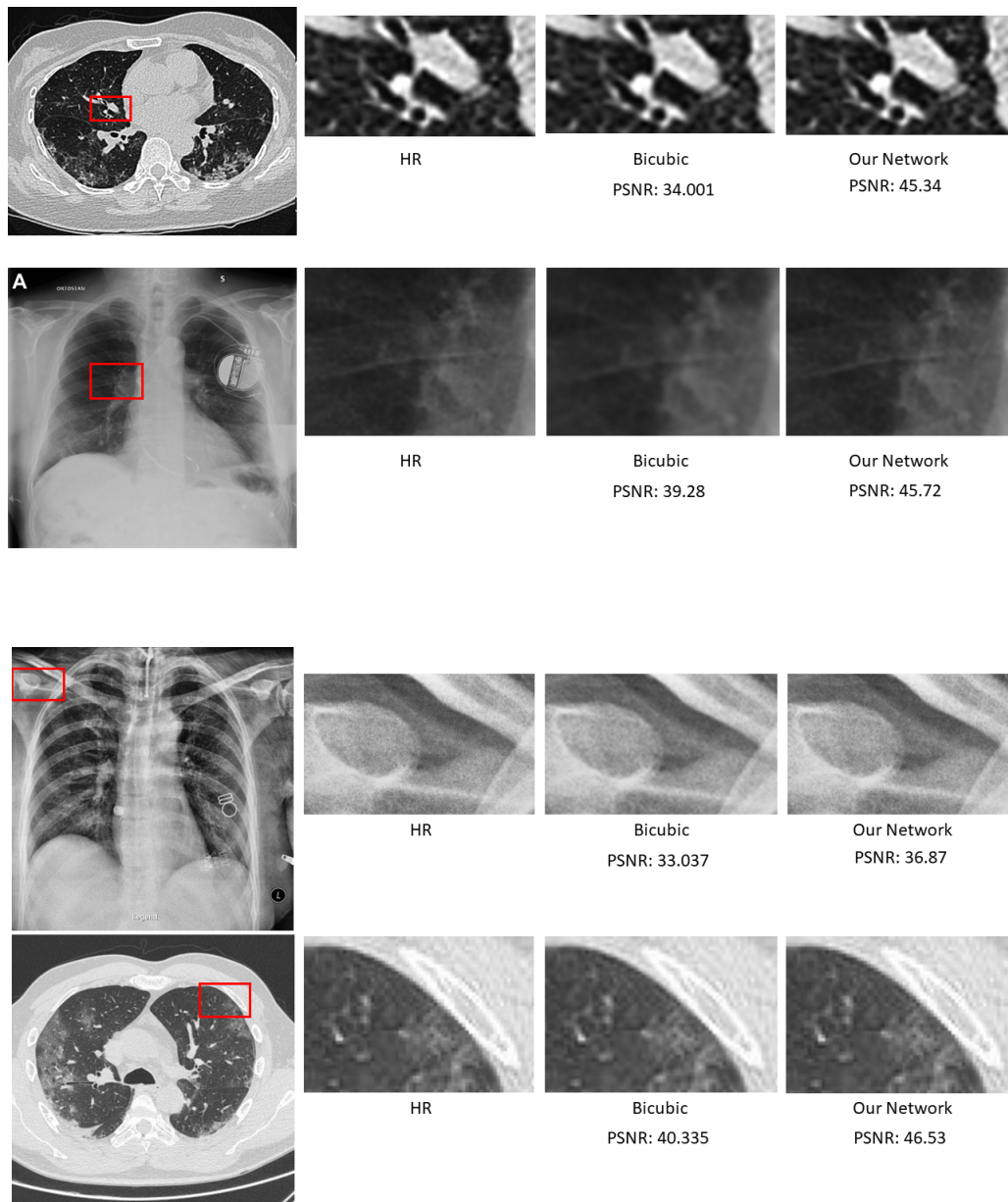


Figure 4.10: Qualitative results of our network on scale 2 upsampling

Data	Bicubic	Our Network
COVID-19	36.67	43.62

Table 4.6: Quantitative results of this on scale 2 upsampling in terms of PSNR

CHAPTER 5

Conclusion & Future Scope

5.1 Conclusion

In the proposed method, we used non-local sparse attention for single image super resolution networks, that simultaneously adopts the benefits of sparse representations and non-local similarity. Furthermore, to improve ability of the network, we suggest channel attention mechanism to adaptively rescale channel-wise features by taking into account inter-dependencies among channels. Our proposed method produces super-resolved results that are comparable with the state-of-the-art architectures in terms of qualitative and quantitative evaluation.

5.2 Future Work

1. We can try to modulate the convolution kernel and generate the adaptive context information and then use it in our architecture to improve the results.
2. We can also try to use overlapping of patches in training to further improve the results in existing architecture.
3. Using channel attention and pixel attention in parallel can also influence and improve results.

References

- [1] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel. Low-complexity single-image super-resolution based on nonnegative neighbor embedding. 2012.
- [2] J. P. Cohen, P. Morrison, and L. Dao. Covid-19 image data collection. *arXiv 2003.11597*, 2020.
- [3] C. Dong, C. C. Loy, K. He, and X. Tang. Image super-resolution using deep convolutional networks. *IEEE transactions on pattern analysis and machine intelligence*, 38(2):295–307, 2015.
- [4] C. Dong, C. C. Loy, and X. Tang. Accelerating the super-resolution convolutional neural network. In *European conference on computer vision*, pages 391–407. Springer, 2016.
- [5] S. Farsiu, D. Robinson, M. Elad, and P. Milanfar. Advances and challenges in super-resolution. *International Journal of Imaging Systems and Technology*, 14(2):47–57, 2004.
- [6] S. Farsiu, M. D. Robinson, M. Elad, and P. Milanfar. Fast and robust multiframe super resolution. *IEEE transactions on image processing*, 13(10):1327–1344, 2004.
- [7] D. Glasner, S. Bagon, and M. Irani. Super-resolution from a single image. In *2009 IEEE 12th international conference on computer vision*, pages 349–356. IEEE, 2009.
- [8] J.-B. Huang, A. Singh, and N. Ahuja. Single image super-resolution from transformed self-exemplars. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 5197–5206, 2015.

- [9] J. Kim, J. K. Lee, and K. M. Lee. Accurate image super-resolution using very deep convolutional networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2016.
- [10] J. Kim, J. K. Lee, and K. M. Lee. Deeply-recursive convolutional network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 1637–1645, 2016.
- [11] D. P. Kingma and J. Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [12] X. Li, Y. Hu, X. Gao, D. Tao, and B. Ning. A multi-frame image super-resolution method. *Signal Processing*, 90(2):405–414, 2010.
- [13] Z. Li, J. Yang, Z. Liu, X. Yang, G. Jeon, and W. Wu. Feedback network for image super-resolution. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3867–3876, 2019.
- [14] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee. Enhanced deep residual networks for single image super-resolution. *CoRR*, abs/1707.02921, 2017.
- [15] D. Liu, B. Wen, Y. Fan, C. C. Loy, and T. S. Huang. Non-local recurrent network for image restoration. *Advances in neural information processing systems*, 31, 2018.
- [16] D. Martin, C. Fowlkes, D. Tal, and J. Malik. A database of human segmented natural images and its application to evaluating segmentation algorithms and measuring ecological statistics. In *Proceedings Eighth IEEE International Conference on Computer Vision. ICCV 2001*, volume 2, pages 416–423. IEEE, 2001.
- [17] Y. Matsui, K. Ito, Y. Aramaki, A. Fujimoto, T. Ogawa, T. Yamasaki, and K. Aizawa. Sketch-based manga retrieval using manga109 dataset. *Multimedia Tools and Applications*, 76(20):21811–21838, 2017.
- [18] Y. Mei, Y. Fan, and Y. Zhou. Image super-resolution with non-local sparse attention. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3517–3526, 2021.

- [19] B. Niu, W. Wen, W. Ren, X. Zhang, L. Yang, S. Wang, K. Zhang, X. Cao, and H. Shen. Single image super-resolution via a holistic attention network. In *European Conference on Computer Vision*, pages 191–207. Springer, 2020.
- [20] S. C. Park, M. K. Park, and M. G. Kang. Super-resolution image reconstruction: a technical overview. *IEEE signal processing magazine*, 20(3):21–36, 2003.
- [21] D. Scharstein and C. Pal. Learning conditional random fields for stereo. In *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pages 1–8. IEEE, 2007.
- [22] R. Timofte, E. Agustsson, L. Van Gool, M.-H. Yang, and L. Zhang. Ntire 2017 challenge on single image super-resolution: Methods and results. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, July 2017.
- [23] Z. Wang, A. Bovik, H. Sheikh, and E. Simoncelli. Image quality assessment: from error visibility to structural similarity. *IEEE Transactions on Image Processing*, 13(4):600–612, 2004.
- [24] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon. Cbam: Convolutional block attention module. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.
- [25] J. Yang, J. Wright, T. S. Huang, and Y. Ma. Image super-resolution via sparse representation. *IEEE transactions on image processing*, 19(11):2861–2873, 2010.
- [26] R. Zeyde, M. Elad, and M. Protter. On single image scale-up using sparse-representations. In *International conference on curves and surfaces*, pages 711–730. Springer, 2010.
- [27] L. Zhang and W. Zuo. Image restoration: From sparse and low-rank priors to deep priors [lecture notes]. *IEEE Signal Processing Magazine*, 34(5):172–179, 2017.
- [28] Y. Zhang, K. Li, K. Li, L. Wang, B. Zhong, and Y. Fu. Image super-resolution using very deep residual channel attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, September 2018.

- [29] Y. Zhang, K. Li, K. Li, B. Zhong, and Y. Fu. Residual non-local attention networks for image restoration. *arXiv preprint arXiv:1903.10082*, 2019.
- [30] Y. Zhang, Y. Tian, Y. Kong, B. Zhong, and Y. Fu. Residual dense network for image super-resolution. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 2472–2481, 2018.
- [31] H. Zhao, X. Kong, J. He, Y. Qiao, and C. Dong. Efficient image super-resolution using pixel attention. In *European Conference on Computer Vision*, pages 56–72. Springer, 2020.
- [32] W. W. Zou and P. C. Yuen. Very low resolution face recognition problem. *IEEE Transactions on image processing*, 21(1):327–340, 2011.