REAL-TIME IMAGE SUPER-RESOLUTION RECONSTRUCTION FOR SYSTEM ON CHIP FPGA

Mr. Watchara Ruangsang

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

A Dissertation Submitted in Partial Fulfillment of the Requirements
for the Degree of Doctor of Philosophy in Electrical Engineering
Department of Electrical Engineering
FACULTY OF ENGINEERING
Chulalongkorn University
Academic Year 2021

Copyright of Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

การสร้างคืนภาพความละเอียดสูงยิ่งยวดเวลาจริงสำหรับเอฟพีจีเอแบบชิปที่มีระบบประมวลผล

นายวัชระ เรืองสังข์

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วิทยานิพนธ์นี้เป็นส่วนหนึ่งของการศึกษาตามหลักสูตรปริญญาวิศวกรรมศาสตรดุษฎีบัณฑิต
สาขาวิชาวิศวกรรมไฟฟ้า ภาควิชาวิศวกรรมไฟฟ้า
คณะวิศวกรรมศาสตร์ จุฬาลงกรณ์มหาวิทยาลัย
ปีการศึกษา 2564

ลิขสิทธิ์ของจุฬาลงกรณ์มหาวิทยาลัย

Thesis Title	REAL-TIME IMAGE SUPER-RESOLUTION RECONSTRUCTION FOR SYSTEM ON CHIP FPGA
By	Mr. Watchara Ruangsang
Field of Study	Electrical Engineering
Thesis Advisor	Associate Professor SUPAVADEE ARAMVITH, Ph.D.
Thesis Co Advisor	Professor Takao Onoye, D.Eng.

Accepted by the FACULTY OF ENGINEERING, Chulalongkorn University in Partial Fulfillment of the Requirement for the Doctor of Philosophy

.... Dean of the FACULTY OF ENGINEERING
(Professor SUPOT TEACHAVORASINSKUN, D.Eng.)

DISSERTATION COMMITTEE

.... Chairman
(Professor Prasit Prapingmongkolkarn, Ph.D.)
.... Thesis Advisor
(Associate Professor SUPAVADEE ARAMVITH, Ph.D.)
.... Thesis Co-Advisor
(Professor Takao Onoye, D.Eng.)
.... Examiner
(Assistant Professor SUREE PUMRIN, Ph.D.)
.... Examiner
(Associate Professor CHARNCHAI PLUEMPITIWIRIYAWEJ, Ph.D.)
.... External Examiner
(Sitapa Rujikietgumjorn, Ph.D.)

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

วัชระ เรืองสังข์ : การสร้างคืนภาพความละเอียดสูงยิ่งยวดเวลาจริงสำหรับเอฟพีจีเอแบบชิปที่มีระบบประมวลผล. (REAL-TIME IMAGE SUPER-RESOLUTION RECONSTRUCTION FOR SYSTEM ON CHIP FPGA) อ.ที่ปรึกษาหลัก : รศ. ดร.สุภาวดี อร่ามวิทย์, อ.ที่ปรึกษาร่วม : ศ. ดร.ทาคาโอะ โอโนเย

ในปัจจุบัน การสร้างคืนภาพความละเอียดสูงยิ่งยวดใช้เทคนิคภายใต้โครงข่ายประสาทเทียมแบบคอนโวลูชั่น ได้รับความสนใจอย่างมาก ในการใช้งานด้านคอมพิวเตอร์วิชชั่นและบริษัทที่ทำเกี่ยวกับปัญญาประดิษฐ์ อย่างไรตามวิธีการการสร้างคืนความละเอียดสูงยิ่งยวดต้องใช้ทรัพยากรในการประมวลที่ต้องใช้พลังงานสูง และการใช้หน่วยความจำจำนวนมาก ซึ่งปัญหาที่สำคัญ ในการนำไปใช้งานจริง คือ การออกแบบให้มีประสิทธิภาพและโมเดดมีขนาดเหมาะสม ปรับปรุงความคมชัดและคุณภาพของภาพไม่เสียไป โดยการแก้ไขดังกล่าว เราจึงเสนอโครงข่ายข้ามช่องสัญญาณตกค้างรวมหลายทาง จากการใช้สัญญาณตกค้างในสัญญาตกค้างภายใต้การเชื่อมต่อหลายทาง พร้อมทั้งคุณลักษณะพื้นฐานของชั้นข้อมูลก่อนหน้าในการสร้างคืนภาพความละเอียดสูง โดยใช้การฝึกฝนโมเดลกับฐานข้อมูลจากการใช้การ์ดประมวลผลภาพ และนำไปใช้งานกับเอฟพีจีเอแบบชิปที่มีระบบประมวลผลโดยการใช้เทคนิคการแบบนับ จากข้อมูลแบบอิงดรรชนีเป็นไม่อิงดรรชนีแบบจำนวนเต็ม ไปประมวลผลโดยใช้โมดูลดีพียู โดยผลการทดลองแสดงเห็นว่า วิธีการที่นำเสนอสามารถลดขนาดตัวแปรได้อย่างมีนัยสำคัญ (8.4 เท่าเมื่อเทียบกับวิธีการแบบอาร์ซ๊เอเอ็น) โดยที่รักษาคุณภาพของภาพและค่าพีเอสเอ็นอาร์กับเทคนิคทันสมัยต่างๆ อีกทั้งสามารถนำไปประมวลผลเอฟพีจีเอแบบชิปที่มีระบบประมวลผลด้วยความเร็ว 30 เฟรมต่อวินาที และภาพมีความถูกต้องในรูปแบบการวัดค่าแบบพีเอสเอ็นอาร์ไม่ลดลงน้อยกว่า 1 เดซิเบล

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

สาขาวิชา วิศวกรรมไฟฟ้า
ปีการศึกษา 2564

ลายมือชื่อนิสิต ....
ลายมือชื่อ อ.ที่ปรึกษาหลัก ....
ลายมือชื่อ อ.ที่ปรึกษาร่วม ....

5871440221 : MAJOR ELECTRICAL ENGINEERING
KEYWORD: Convolutional Neural Network, Hardware Implementation, Residual Network, Super Resolution, System on Chip FPGA

Watchara Ruangsang : REAL-TIME IMAGE SUPER-RESOLUTION RECONSTRUCTION FOR SYSTEM ON CHIP FPGA. Advisor: Assoc. Prof. SUPAVADEE ARAMVITH, Ph.D. Co-advisor: Prof. Takao Onoye, D.Eng.

In recent years, image super-resolution (SR) techniques based on Convolutional Neural Network (CNN) have achieved impressive attention from computer vision scholars and artificial intelligence (AI) companies. Due to the necessity of using the SR algorithms in real-world applications, designing an efficient and lightweight SR algorithm that improves the sharpness and visual quality of the SR results is a critical issue in real-time hardware implementation. To address these issues, we proposed the Multi-FusNet of Cross Channel Network (MFCC) network by constructing the groups of Residual-in-Residual architecture under the multi-path cascading framework. Additionally, a residual connection is used to transfer the low-level features of the early layer to the reconstructed SR image. The proposed SR model is initially trained with the GPU's training image dataset. To implement our trained model in System on Chip FPGA, the size of the proposed model is required to reduce. We convert the floating-point checkpoint into a fixed-point integer checkpoint in the quantization procedure. According to the experimental results, the proposed method reduces the number of network parameters significantly (8.4 times compared to RCAN), can execute fast in System On Chip FPGA, around 30 frames per second, and image accuracy of the proposed method in terms of PSNR value does not decrease over 1 dB.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Field of Study:	Electrical Engineering	Student's Signature ....
Academic Year:	2021	Advisor's Signature ....
		Co-advisor's Signature ....

ACKNOWLEDGEMENTS

First of all, I would like to express my gratitude to my thesis advisor Assoc. Prof. Dr. Supavadee Aramvith for her continuous support and untiring effort all the time during my study. This achievement would not have been possible without her guidance. I would like to thank my thesis co-advisor Prof. Takao Onoye for giving suggestions, guidance, idea, and viewpoint on my dissertation.

I also would like to thank Chairman Prof. Prasit Prapingmongkolkarn for giving suggestions on my dissertation and giving me a chance of working on the project. I would like to thank the other member of my thesis committee, Asst. Prof. Suree Pumrin,Asst. Prof. Charnchai Pluempitiwiriyawej, and Dr.Sitapa Rujikietgumjorn, for giving critical reviews and advice on my dissertation.

I also appreciate all members of the Video Technology Research Group (VTRG), seniors, and my friends for their support and spending their time for suggestions and guidance.

Last but not least, I would like to deepest gratitude to express my family members for their help, unfailing understanding, and affectionate encouragement during my graduate journey.

Watchara Ruangsang

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

TABLE OF CONTENTS

Page

.... iii

ABSTRACT (THAI) iii

.... iv

ABSTRACT (ENGLISH) iv

ACKNOWLEDGEMENTS v

TABLE OF CONTENTS vi

LIST OF TABLES ix

LIST OF FIGURES x

Chapter 1 Introduction 1

  1.1. Motivation and Problem Statement 1

  1.2. Objective 3

  1.3. Working Scopes 3

Chapter 2 Background 4

  2.1. Super-Resolution Algorithm 4

    2.1.1. Single Image Interpolation Based methods 4

    2.1.2. Multiple Image Reconstruction Methods 4

    2.1.3. Learning-Based Methods 5

      2.1.3.1. Classical Sparse Coding Method 6

      2.1.3.2. Anchored Neighborhood Regression Method 6

      2.1.3.3. Self-Exemplars Method 7

      2.1.3.4. Deep Learning Method 7

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2. Super-Resolution Algorithm in Deep Learning Method 8
  2.2.1. Linear Networks 8
  2.2.2. Residual Network 10
  2.2.3. Recursive Networks 17
  2.2.4. Progressive Reconstruction 20
  2.2.5. Densely Connected Networks 21
  2.2.6. Multi-Branch Designs 23
  2.2.7. Attention Based Networks 25
  2.2.8. Multiple Degradation Handing Networks 26
  2.2.9. Generative Adversarial Nets Network 27
2.3. Super-Resolution on Hardware 29
  2.3.1. Interpolation based on hardware 29
  2.3.2. Multiple frames based on hardware 30
  2.3.3. Learning-Based Method 31
  2.3.4. Deep Learning Method 32
Chapter 3 Image Super-Resolution 35
3.1. Multi-FusNet of Cross Channel Network 37
  3.1.1. Residual Group: 39
  3.1.2. Multipath Residual: 40
  3.1.3. Pixel Shuffle Fusion 41
3.2. Dataset Image 47
  3.2.1. Training Dataset image 47
    3.2.1.1. DIV2K [69] 47
  3.2.2. Testing Dataset Image 47

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

        3.2.2.1. Set5 [70] 47
        3.2.2.2. Set14 [71] 48
        3.2.2.3. BSD100 [72] 48
        3.2.2.4. Urban100 [19] 49
        3.2.2.5. Manga109 [73] 49
    3.2.3. Testing Dataset Image of HW 50
  3.3. Hardware 50
    3.3.1. System on Chip (SoC) FPGA 50
    3.3.2. Hardware Design Flow 53
  3.4. Performance Evaluate 56
    3.4.1. Peak Signal to Noise Ratio: PSNR 56
    3.4.2. Structural similarity (SSIM) 56
    3.4.3. Frame Per Second (FPS) 57
Chapter 4 Experimental Results and Discussions 58
  4.1. Experimental Setup 58
  4.2. Result Analysis of Network 58
  4.3. Result Analysis of Hardware 66
Chapter 5 Conclusion and Future Works 69
REFERENCES 70
VITA 79

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

LIST OF TABLES

Page

Table 1 BASED LINE MODEL AT 50 EPOCH ON SET5 (2×) 44
Table 2 THE EXTENSION OF RESIDUAL IN BASED LINE MODEL 46
Table 3 RESULTS QUANTITATIVE OF MODEL X2 59
Table 4 RESULTS QUANTITATIVE OF MODEL X3 60
Table 5 RESULTS QUANTITATIVE OF MODEL X4 61
Table 6 RESULTS QUANTITATIVE OF MODEL X8 62
Table 7 Hardware implementations average PSNR in with a scaling factor 2 of Set5 67
Table 8 Hardware implementations comparison 68

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

LIST OF FIGURES

Page

Figure 1 Example CCTV Image 1

Figure 2 Super-Resolution Convolutional Neural Network PSNR Performance [5] 2

Figure 3 Learning a Deep Convolutional Network for Image Super-Resolution (SRCNN) [21] 8

Figure 4 Learning a Deep Convolutional Network for Image Super-Resolution Network Structure (FSRCNN) [22] 9

Figure 5 Very Deep Convolutional Networks (VDSR) Network Structure [7] 10

Figure 6 Enhanced Deep Residual Networks (EDSR) for Single Image Super-Resolution Structure [25] 11

Figure 7 Cascading Residual Network (CARN) [32] 12

Figure 8 Multi-scale Residual Network for Image Super-Resolution (MSRN) [33] 13

Figure 9 Super-Resolution Feedback Network (SRFBN) [34] 14

Figure 10 Adaptive Weighted Super-Resolution Network (AWSRN) [35] 14

Figure 11 Multi-scale inception-based super-resolution Structure (MSISRD) [36] 15

Figure 12 Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning (A2F) [37] Structure 16

Figure 13 Multi-objective reinforced evolution in mobile neural architecture search (MoreMNAS) Architecture [38] 17

Figure 14 Deeply-Recursive Convolutional Network (DRCN) Structure [8] 18

Figure 15 Deep Recursive Residual Network (DRRN) Structure [40] 19

Figure 16 Memory Network for Image Restoration (MemNet) Structure [23] 20

Figure 17 Laplacian Pyramid Super-Resolution Network (LapSRN) [6] 21

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Figure 18 Deep Residual Dense Network (RDN) Structure [26] 22

Figure 19 Deep Back-Projection Networks (DBPN) for super-resolution [27]....... 23

Figure 20 Fast and Accurate Single Image Super-Resolution via Information Distillation Network (IDN) [41] 24

Figure 21 Image Super-Resolution with Information Multi-distillation Network (IMDN) [41] 25

Figure 22 Deep residual learning for image recognition Network (SelNet) [42]....... 25

Figure 23 Residual Channel Attention Network (RCAN) [28] Structure....... 26

Figure 24 Super-Resolution Network for Multiple Degradations Structure (SRMDNF) [24] 27

Figure 25 Generative Adversarial Nets Network Structure [29] 28

Figure 26 Interpolation based on Hardware [43] 29

Figure 27 Interpolation based on Hardware [44] 30

Figure 28 Multiple Frame based on Hardware [45] 30

Figure 29 Multiple-frame registration on hardware [46] 30

Figure 30 Block Diagram of Learning-Based Method on Hardware [47] 31

Figure 31 Block Diagram of Learning-Based Method on hardware [48] 31

Figure 32 Field Programmable Gate Array (FPGA) for real-time Video SR [49]....... 32

Figure 33 Block Diagram of Deep Learning Method on Hardware [50] 33

Figure 34 Block Diagram of Deep Learning Method on Hardware [51] 33

Figure 35 Block Diagram of Deep Learning Method on Hardware [52] 34

Figure 36 Multi-FusNet of Cross Channel Network for Image Super-Resolution 38

Figure 37 PSNR vs. Parameter 43

Figure 38 Multi-Residual Networks [68] Structure 45

Figure 39 An example of an image in DIV2K image 47

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Figure 40 An example of an image in Set5 47

Figure 41 An example of an image in Set14 48

Figure 42 An example of an image in BSD100 48

Figure 43 An example of an image in Urban100 49

Figure 44 An example of an image in Manga109 49

Figure 45 Example of an image in Testing Dataset Image of HW 50

Figure 46 Kria KV260 specifications 52

Figure 47 Hardware Design Flow 53

Figure 48 Kria KV260 Model 54

Figure 49 Qualitative comparison over datasets for scale ×4. The red rectangle indicates the area of interest for zooming. 63

Figure 50 Qualitative comparison over datasets for scale ×4. The red rectangle indicates the area of interest for zooming. 64

Figure 51 Qualitative comparison of img093 belongs to Urban100 dataset for scale ×8. The yellow rectangle indicates the area of interest for zooming 65

Figure 52 Qualitative comparison of an image belongs to Manga109 dataset for scale ×8. The yellow rectangle indicates the area of interest for zooming. 65

Figure 53 Target Resolution 66

Figure 54 Result of KV260 processing super resolution method MFCC 2x 66

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter 1
Introduction

1.1. Motivation and Problem Statement

A closed-circuit television (CCTV) is used for urban security and environmental monitoring. In the setup of smart cities, home video surveillance systems provide the necessary visual information to improve situational awareness, enhance security, deter potential theft, and provide remote monitoring. But the output image of CCTV may have low resolution (LR) or loss of information because of the camera setup and lighting [1] as shown in Figure 1.

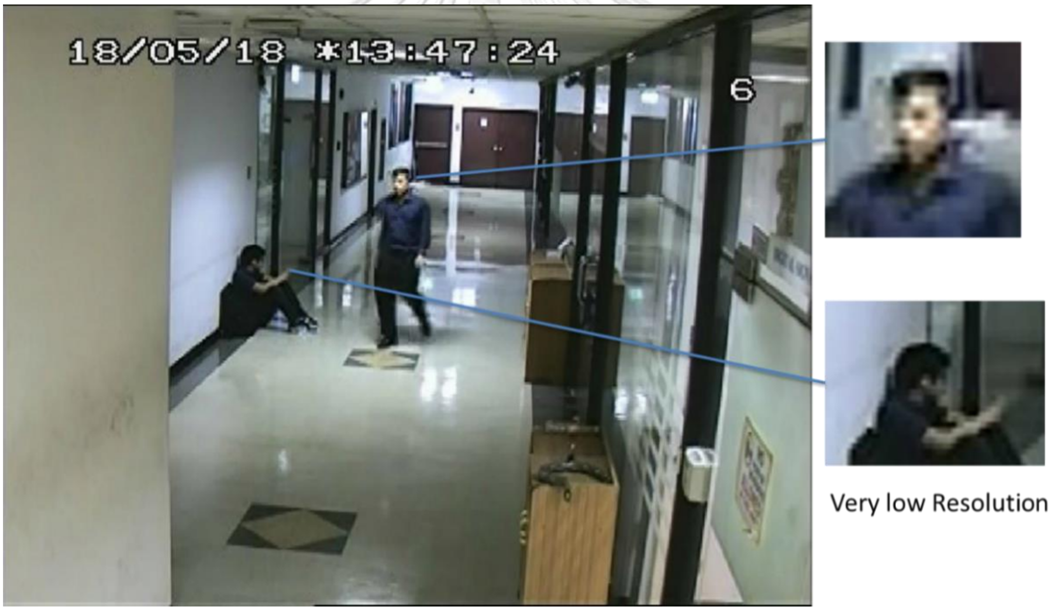


Figure 1 *Example CCTV Image*

Face recognition and understanding information such as texts and tasks are applied in other aspects. If a person's face is taken from a long distance, the output image is hard to detect and learn to recognize faces. The computer algorithm for improving the resolution of the LR images is known as Super-Resolution (SR) [2].

Even though bicubic interpolation [3] is widespread in SR because of its fast speed, the HR image output is not clear and begins to blur. The dictionary method, also called sparse coding (SC) [4] can produce better quality results than the bicubic interpolation. SC utilizes the relationship between LR and HR images. At present, the most widely used method is the convolution neural network (CNN) [5]. This method is also used to learn from image datasets. The HR image output of the CNN method has a higher peak signal-to-noise ratio (PSNR) value than the bicubic and the sparse coding methods, as shown in Figure 2. Hence, the image quality of using CNN is better than other methods.

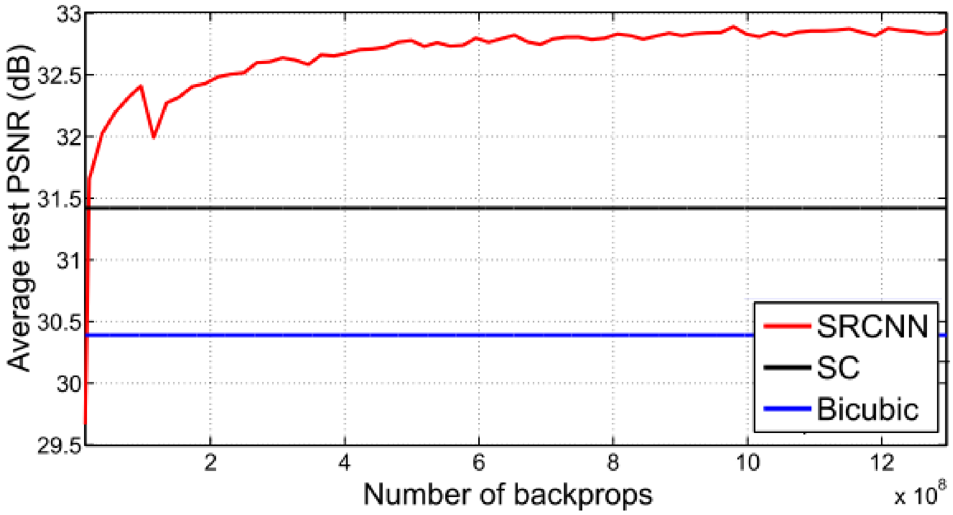


Figure 2 Super-Resolution Convolutional Neural Network PSNR Performance [5]

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

In [6-8], graphics processing unit (GPU) processing to process CNN makes them more expensive. Typically, a convolution process uses the whole image. If the method uses patch processing, there is some processing time for the path sequence process. To solve this problem, the advantage of parallel processing in field-programmable gate array (FPGA) can help the repetitive process faster. Moreover, FPGA can design on the chip for some applications. Integration of FPGA with central processing unit (CPU), memory, and input/output port by low power consumption is called System on Chip (SOC). The objective of combined SoC with FPGAS is to use some libraries or processing from CPU with parallel processing of FPGA for faster processing.

In this research, we are interested in HR imaging using SR based on CNN [5]. SR is an algorithm to reconstruct HR images. The challenge is to be able to perform in real-time. An embedded super-resolution approach can be implemented and connected directly to CCTV to solve that problem. This approach enables real-time processing with a faster and more optimized algorithm version. There are several CNN layers to learn the features of the image for reconstruction in a typical CNN-based SR algorithm. This network structure requires a considerable amount of processing time.

1.2. Objective

1. Develop the algorithm for fast image super-resolution based on a convolution neural network.
2. Proposed super-resolution on hardware for fast image super-resolution.

1.3. Working Scopes

1. This work used processing from image frame by frame.
2. The proposed algorithm can process fast super-resolution in System On Chip FPGA at least 30 frames per second.
3. The image accuracy of the proposed method in terms of PSNR value does not decrease over 1 dB.
4. The performance measurement of the algorithm is tested with:

-> Set5 image dataset (5 images)

-> Set14 image dataset (14 images)

-> Urban100 image dataset (100 images)

Next, we will describe the related work in the following section:

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Chapter 2
Background

2.1. Super-Resolution Algorithm

The SR approaches of using a single LR image to create an HR image are more accessible and faster than using multiple LR images to create HR images. The multiple images can find more lost data but take exceedingly long processing time and are more complex. The SR method can be divided into three techniques: single image interpolation-based, multiple image reconstruction, and learning-based.

2.1.1. Single Image Interpolation Based methods

Single Image Interpolation methods use a single image to create HR image. The process is simple, but some information may be lost. The details of the output image are not quite the same as the original image. This method can be divided into as follows. Smoothing [9] makes the HR image with less or disappear noise. The method aims to eliminate or reduce noise. Examples of smoothing are the Gaussian, Median, and Wiener filters. However, in these processes, some details of the image will be lost. The sharpening process will not increase the contrast ratio. But the process will add the details to the HR image. The advantage is adjusting the contrast and noise does not increase.

Moreover, the interpolation has a process that uses pixels or blocks in the image, such as the nearest neighbor, bilinear interpolation [10] and bicubic interpolation [3]. These methods will include the smoothing and sharpening process. But the details of the HR image on borders, corners, and surfaces in the picture are unclear.

2.1.2. Multiple Image Reconstruction Methods

Multiple Image Reconstruction Method [11] uses multiple LR images. Where the images are taken at different angles of the same scene, each LR image is an image that is in a different position. If the position of the image is in the same position, the HR image acquired has the same information as the LR image. This method is used for creating the HR image when LR images lose a lot of information. To use this method, the first step is to define the pixel's position by assigning a one-pixel image reference. But the real LR image is difficult to define the position. Suppose the position of the pixel is an error or wrong position; it has a huge effect on the efficiency and accuracy of the HR image. The second step is to compare the pixel's position with the pixel in HR images. But the information of LR images changed position without a pattern. It cannot be matched with the pixels of the HR image—the estimated need to have an area of HR images from some parts of LR images. The whole process will improve the image resolution. But in the first step, the image is blurred. Therefore, it is resolved using the procedure to estimate the range in the next step. The last step is improving the HR image's blurring by the pixel area's average value. To use multiple LR images for the SR method when the input image has very LR. The HR image will have bad quality. This method needs to consider the special characteristics of the objects, especially the face image, license plate image, and gestures that arise.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.1.3. Learning-Based Methods

Typically, the image has a lot of details. When the analysis is complex and there are tricky, several research works [12-16] believe the relationships between images in LR and HR images can be learned. The method used image data set to determine the relationship. The process that studies how to get all the information from HR images from LR images is called the Learning-Based Method.

The disadvantage of this method is determining the relationship of LR images compared with the HR images required to study a sample data set. That process uses a lot of processing time. It must be finding a way to improve this process to take less.

Learning-based method for SR will divide the image into patch images. Typically, this method is used to find the same information and use the data sets using pre-processing. Under this learning-based method can find missing data of HR images as well. The related research on the learning-based method is related research work as follows.

2.1.3.1. Classical Sparse Coding Method

The concept of the super-resolution (SR) method is to reconstruct high resolution (HR) image by multiplying LR image $\boldsymbol{Y}$ with blurred image $\boldsymbol{H}$ and down-sampled factor $\boldsymbol{D}$, resulting in the HR image $\boldsymbol{X}$, as shown in (1)

$$Y = DHX \tag{1}$$

Yang et al. [17] proposed a sparse representation for dictionary coefficient from the trained image data and patch processing to learn the relationship between LR and HR images. As shown in (2) be composed of operator feature extraction *F, LR* dictionary coefficient D_l, sparse representation α, LR input patch *y,* and weight value $\boldsymbol{\lambda}$. After that, HR is reconstructed from sparse representation with dictionary coefficient. This method can improve resolution quality but has more processing time.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

$$\min_{\alpha} \| FD_l \alpha - Fy \|_2^2 + \lambda \| \alpha \|_0 \tag{2}$$

2.1.3.2. Anchored Neighborhood Regression Method

Timofte et al. [18] proposed Anchored Neighbor Regression (ANR) as the fast learning-based method. The method reconstructs HR from LR patches from neighborhoods exemplar using ridge regression (l_2-norm) as shown in equation (3). N_l is neighborhood LR image correlation and β is weight vector.

$$\min_{\beta} \| IF - N_l \beta \|_2^2 + \lambda \| \beta \|_2 \tag{3}$$

2.1.3.3. Self-Exemplars Method

Huang et al. [19] proposed self-exemplars for creating an HR image. These methods are based on learning an image by itself without training data. HR image output used high processing time.

2.1.3.4. Deep Learning Method

In recent years, deep Convolutional Neural Networks (CNN) have been useful for SR with a higher reconstruction quality than the learning-based method. CNN usually has complex structures; thus, the computational time can also be higher. Typically, CNN consists of convolution and deconvolution layers as in (4).

$$X_i = max(0, w_i * X_{i-1} + b_i) \tag{4}$$

Where X_i and X_{i-1} are the output and input signals of an i^{th} layer. Each layer's biases and weight values are W_i and bi, respectively. The last parameter (*) is the convolution operation. The reconstruction layer represents in (5).

$$X_i = w_i * X_{i-1} + b_i \tag{5}$$

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

2.2. Super-Resolution Algorithm in Deep Learning Method

The excellent network [20] based on CNN can group into simple networks [5, 21-24], residual networks [6-8, 25-28] and generative adversarial networks [29]. Single Image Super-Resolution (SISR) [30] extensively uses deep Convolutional Neural Networks (CNN) learning to represent networks of image super-resolution by the design networks, respectively.

2.2.1. Linear Networks

The simple network is Super-Resolution using deep Convolutional Neural Networks (SRCNN) [5, 21] the first state-of-the-art CNN for SR methods. SRCNN is based on CNN, and the HR output achieves high quality in terms of a peak-signal-to-noise ratio (PSNR). The network consists of a feature extraction layer to build an outline input image, a non-linear mapping layer that shows low-resolution features to high-resolution features, and a reconstruction layer to crate HR images from LR patches. The SRCNN uses bicubic interpolation to build an HR image of the same size due to an end-to-end network with Mean Squared Error (MSE) loss function.

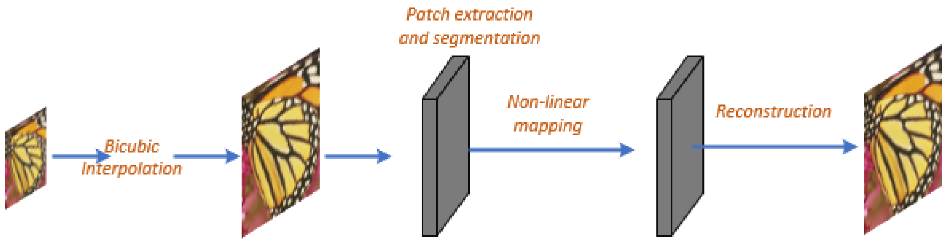


Figure 3 Learning a Deep Convolutional Network for Image Super-Resolution (SRCNN) [21]

After that, Fast-SRCNN (FSRCNN) [22] develops quality with a fast process for real-time applications. The network has four layers of convolution and one deconvolution to an upscale image. The feature extraction layer uses LR images to map features. After that, shrinking is utilized to decrease a very large LR feature dimension. Expanding is the penultimate step used to achieve high quality in the HR part by adding a layer of a feature. Finally, deconvolution is the important step in creating an HR image with each pixel from the feature value. The idea is to add sixteen recursive layers to improve accuracy. This process adds a layer without a new parameter. The difference with SRCNN is the input size using a smaller filter and extending mapping layers. The approach increases the layer for improved accuracy. Output HR image reconstruction by combining residual learning with bicubic interpolation is shown in Figure 2.

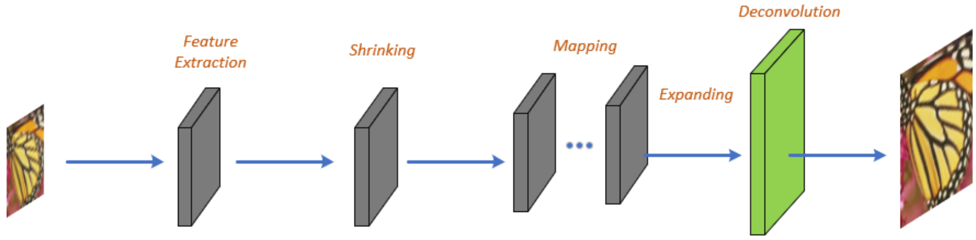


Figure 4 Learning a Deep Convolutional Network for Image Super-Resolution Network Structure (FSRCNN) [22]

FSRCNN offers a higher speed and a better restoration quality than SRCNN by using a smaller filter size and extending mapping layers. The feature extraction layer uses LR images to map features. After that, shrinking is utilized to decrease a very large LR feature dimension. Expanding is the penultimate step used to achieve high quality in the HR part by adding a layer of a feature. Finally, deconvolution is the important step to creating an HR image with each pixel from the feature value as shown in (6), where y_k is HR in k patch of feature pixel in x_i at the position of m and n and weight w_i is weight value at an i^{th} layer of a feature.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

$$y_k = x_i(m,n) \times w_i \tag{6}$$

Very Deep Super-Resolution (VDSR) [7] increases the layer for improved accuracy. Output HR image reconstruction by combining residual with bicubic interpolation. That network names from team join competition represent convolution same size to fend off convergent.

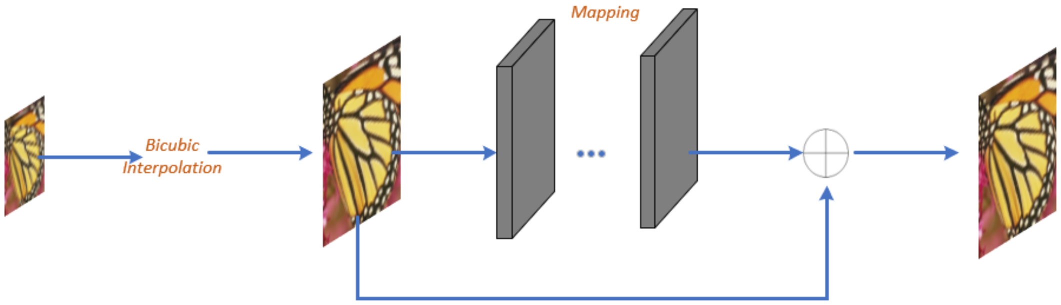


Figure 5 Very Deep Convolutional Networks (VDSR) Network Structure [7]

2.2.2. Residual Network

Deep Super-Resolution (EDSR) [25] used residual nets (ResNet) [31] by deleting Batch Normalization layers and ReLU activation to a lower number of parameters. EDSR method exceeds in performance of current SOTA models. The significant performance improvement is due to optimization by removing avoidable studies programs widely used residual networks and expanding the model's size for setting up the training course. It also suggests a new system MDSR and training strategy for reconstructing HR images for different scaling factors in a single model rather than multiple models. This method carefully designs the configuration and uses practiced operations, which are the key requirement for training the meshes. This model solves two problems: first, for training the model for altered scaling aspects in a particular model, and the second issue is the increased computation time and memory. These two problems can be solved by reducing unnecessary modules in SRResNet [29] architecture and using appropriate loss functions while training the model. This method uses scale-independent information, i.e., the training of various up-scale models from previously trained down-scale prototypes. Then

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

again, it uses a multi-scale architecture that shares parameters across different scales. An enhanced super-resolution algorithm is proposed in this paper.

Improved results have been achieved while compacting the model and abolishing avoidable modules from the current widely used ResNet example. Residual scaling techniques are also employed to train bigger miniatures securely. The suggested SISR miniature overperforms present models and attains SOTA comparable achievements. Additionally, for further cutting down the model size and time utilized in training, there occurs an advancement of a multi-scale (MS) SR framework. In a unified net, this MS model can develop a productive pact with varied scales of SR with scale-subjected calibers and collective main nets. The model construction is concurrent to SRResNet, but the activation function ReLU layers are not included in the exterior of the residual blocks. Also, the network dummy involves a residual scaling panel since it uses 64 feature maps for each convolution panel.

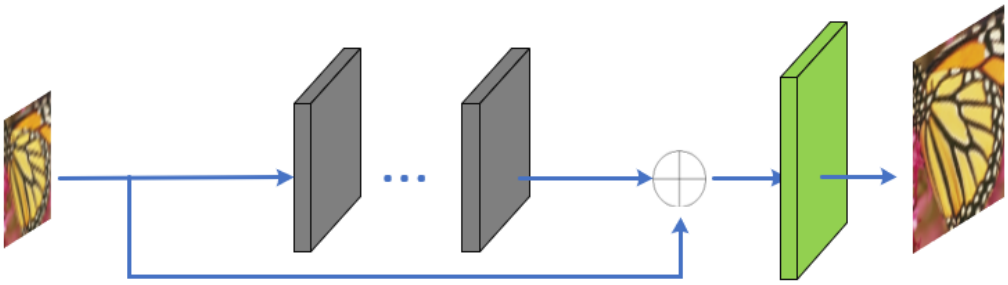


Figure 6 Enhanced Deep Residual Networks (EDSR) for Single Image Super-Resolution Structure [25]

This novel method [32] proposes a cascading mechanism for reducing the model's weight and improving efficiency. Most SOTA models have larger computations which are not good for user experience. Hence we propose Cascading Residual Network to increase the performance and speed by suppressing the network's computations for real-world applications. CARN uses ResNet architecture and a cascading mechanism at local and global levels to include features from all layers. Also, the CARN-M model shows the trade-off between the model's performance and weight. The CARN-M technique can further be applied to video data since many streaming services require storage for high-quality videos. Thus, a design can be made in stores LR videos; after passing through SR, the system produces high-quality videos. The CARN model out-performs all the state-of-art models having less than 5M parameters. It has an efficient computational cost compared to other SOTA models because the basis of comparison is multi-adds (multiply-operations). The size of CARN-M is 1.6MB which is good for use on mobile devices.

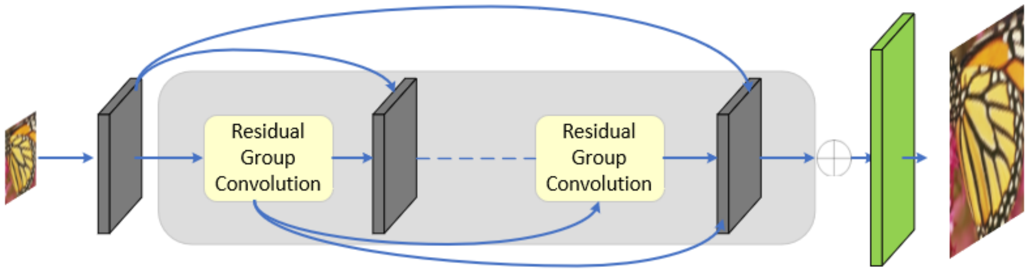


Figure 7 Cascading Residual Network (CARN) [32]

Most of the existing SR models during the reconstruction have the following problems: Poor Feature Utilization; difficulty to adapt with arbitrary scaling factors. A novel multi-scale residual network (MSRN) [33] has been proposed to address the abovementioned issues. -This model can fuse the image features at different scales. It is the first multi-scale module based on a residual structure which is very easy to train -It is a basic skeleton for stratified feature union and is flexible with different up-scaling factors. The model shows superior performance over the SOTA model on various benchmark datasets. Also, this model can achieve competitive results by increasing the number of MSRBs.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

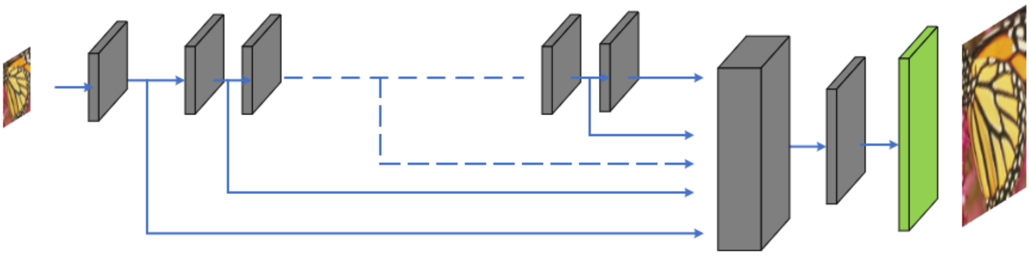


Figure 8 Multi-scale Residual Network for Image Super-Resolution (MSRN) [33]

In image SR, multiple HR images may often result in identical LR images. To solve this issue, multiple solutions have already been proposed. One is Feedback Network for Image SR [34] in which a feedback block is designed to handle feedback connections and generate high-quality reconstruction results. The feedback manner is achieved by using hidden states in an RNN. The skip connections and depth connections play a vital role in any of the deep CNNs. The Depth connections give more contextual information, whereas the skip connections reduce the gradient vanishing problem of deeper networks. The proposed method further introduces a curriculum learning strategy that helps to learn complex degradation with difficulties in HR reconstruction. This model also uses data augmentation. The main contribution of this model design are:

- Employing a feedback mechanism in which high-level information flows in top-down feedback flow, thereby increasing the reconstruction ability with fewer parameters

- Enriches HR through up and down sampling and dense skip connections.

- Applies curriculum learning strategy to learn complex degradations.

Experimental results show superiority of the proposed model over SOTA models in network parameters with BI degradation model, BD, and DN degradation model.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

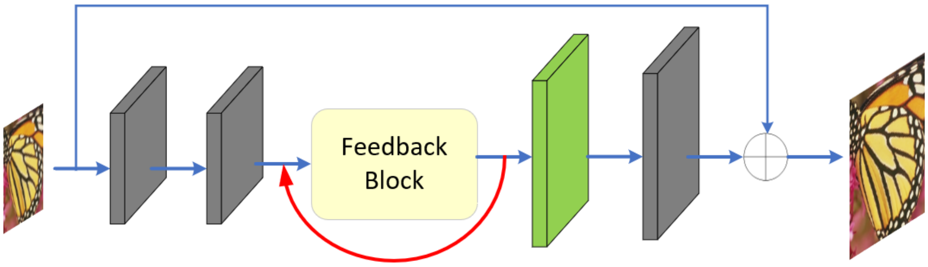


Figure 9 Super-Resolution Feedback Network (SRFBN) [34]

With the fast-growing techniques in the field of super-resolution, Deep Learning has been successfully introducing many model designs to emend the achievement of SR prototypes. However, most CNN-based networks require heavy computation and thus have limited real-world application. The model Adaptive Weight SR network (AWSRN) [35] is designed to resolve this subject. This model consists of Local Fusion Blocks (LFB) designed with residual learning-based embryonic adaptive voluminous residual units (ARWU) and a local residual fusion entity (LRFU). Apart from LFB, it also contains an adaptive weight multi-scale module (AWMS) for enhancing the reconstruction layer to extract better features. AWMS is an important contributing factor to making the network a lightweight structure. AWSRN consists of distinctive blocks: element abstraction, non-linear mapping, and adaptive weight multi-scale modules. The model shows better results of PSNR and SSIM on std. Benchmark library as compared to SOTA formulas.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

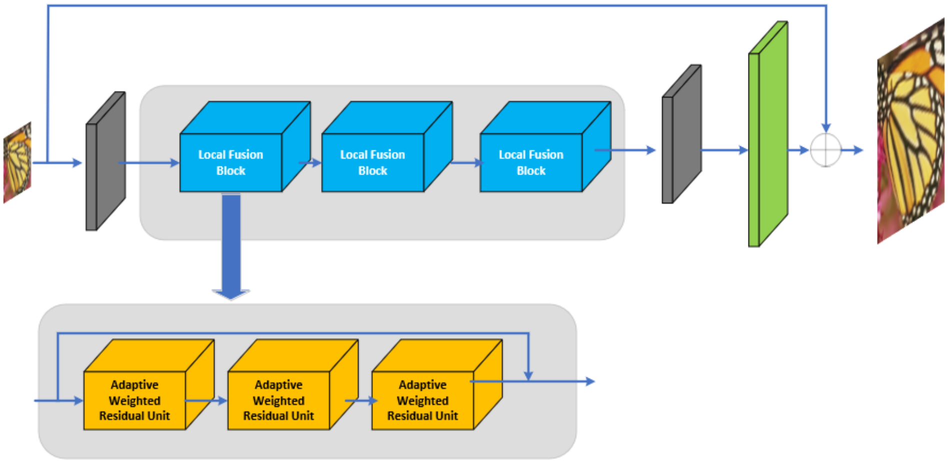

Local Fusion Block
Local Fusion Block
Local Fusion Block
Adaptive Weighted Residual Unit
Adaptive Weighted Residual Unit
Adaptive Weighted Residual Unit

Figure 10 Adaptive Weighted Super-Resolution Network (AWSRN) [35]

Inception Network [36] uses a residual network using asymmetric convolution to reduce the number of parameters. The short and long feature information is extracted using locally residual asymmetric convolutional block and inception-based asymmetric convolutional block architecture by model directly. Compared to the set of SISR models, the MS models remain compact and show competitive results compared to SISR models. The dictated method was evaluated on the standard benchmark dataset and DIV2K dataset. The evaluation was ranked 1st and 2nd in NTIRE SR Challenge. This method is better than the state-of-the-art quantitative and qualitative analysis model for 2, 4, and 8 scaling factors on five benchmark datasets. The model construction is concurrent to SRResNet, but the activation function ReLU layers are not included in the exterior of the residual blocks. Also, the dummy involves a residual scaling panel since it uses 64 feature maps for each convolution panel.

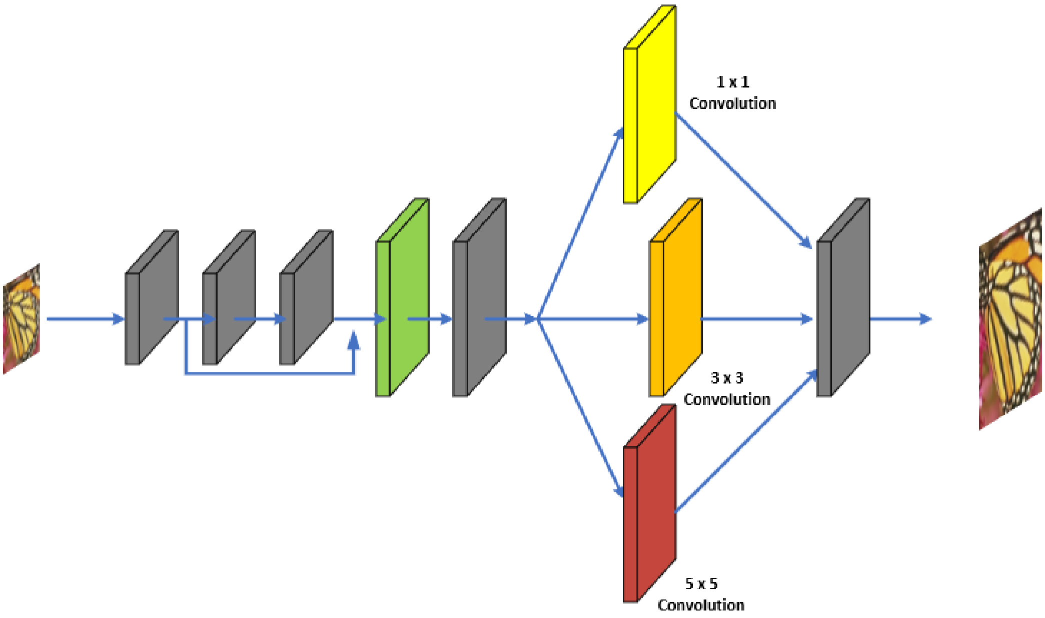


Figure 11 *Multi-scale inception-based super-resolution Structure (MSISRD) [36]*

The network Lightweight Single Image, SR Network with Attentive Auxiliary Feature Learning [37] shows that model performance can be greatly improved by using auxiliary features and channel attention since the studies show that the lesser the number of auxiliary features, the lesser the high-frequency information. To check the effectiveness of auxiliary features, we remove the auxiliary feature branch and check the results with only the attention branch. Even though CNNs have boosted the performance of SISR, they have much higher computational costs restricting their practical applications. Thus A2F model for SISR extracts auxiliary features and projects them to a common space. The channel attention mechanism enhances the use of these identified auxiliary features. Thus, the A2F model, which is fast, accurate, and with low parameters, shows comparable results on five benchmark datasets and better visual results. Also, when the parameters are less than 320K, A2F outperforms SOTA methods for all scales, proving its ability to utilize the auxiliary features better.

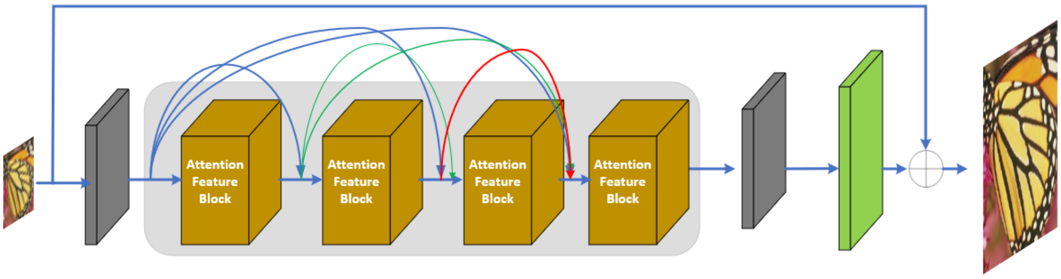


Figure 12 Lightweight Single-Image Super-Resolution Network with Attentive Auxiliary Feature Learning (A2F) [37] Structure

The design of Neural networks for mobile devices with constrained resources is difficult. A new multi-objective-oriented algorithm, MoreMNAS [38] has been proposed to address this issue. This model performs multi-objective NAS using advantages from both reinforcement learning and NSGA-II. It means it harnesses benefits from both evolution algorithm (EA) and reinforced learning (RL). For speeding up the selection, this method constructs a cell-based search allowing mutation. These natural mutations and reinforced controls maintain a delicate balance between exploration and exploitation. Thus, protect the model from degradation during evolution and make better use of learned knowledge. Experiments show comparable outcomes on SOTA methods with fewer FLOPS on benchmark datasets.

This method is not just limited to mobile device applications but also dominates in other domains.

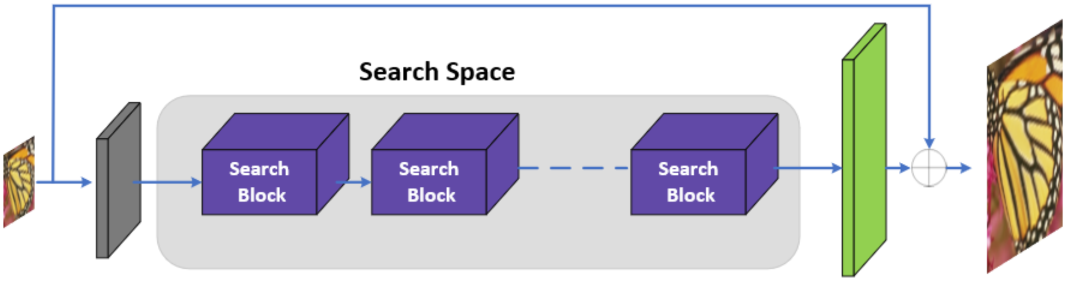


Figure 13 Multi-objective reinforced evolution in mobile neural architecture search (MoreMNAS) Architecture [38]

The FALSR method [39] Fast and Lightweight SR with Neural Architecture Search contributes to maximizing the balance between the image restoration and the models' weight. For the SOTA model, it is difficult to reform the realization of the prototype and constrain the resources. The methods successfully improve the performance of this model with constrained resources using neural architecture search. Usually, in SOTA models, the SR is achieved using a multi-objective approach. Still, in the proposed model, an elastic search approach is used, which is based on a hybrid controller for both micro and macro levels. This elastic search helps reduce the model's computational cost and boosts its capacity. Except for deeper networks such as RDN or RCAN model gives better results than other SOTA methods with comparable FLOPS.

2.2.3. Recursive Networks

Residual learning helps accelerate the learning process as a Deeply Recursive Convolutional Network for Image Super-Resolution (DRCN) [8] with a recurrent block by an individual convolution layer. Moreover, Kim et al. [8] proposed Deeply-Recursive Convolutional Network for Image Super-Resolution (DRCN). The idea is to add sixteen recursive layers to improve accuracy. This process adds a layer without a new parameter.

Chulalongkorn University

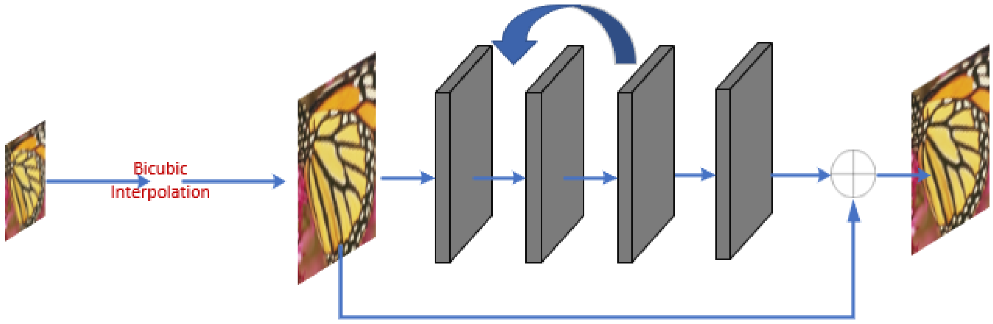


Figure 14 Deeply-Recursive Convolutional Network (DRCN) Structure [8]

A Deep Recursive Residual Network [8] consisting of almost 52 convolutional layers is being proposed in this network that strives for a deep yet concise network to be applicable in mobile systems since they use less storage space. It adopts residual and recursive learning. Residual Learning: Global Residual Learning is specifically adopted to remedy the difficulties of training deep networks, whereas Local Residual Learning is used to address the issue of performance degradation in visual recognition and image restoration and for controlling the model parameters while increasing the depth. Every amorphic layer uses LRL, whereas GRL is carried off among LR and HR data. The DRRN substantially exceeds SOTA in SISR while using slighter panoramas; this has been indicated by extensive benchmark evaluation. A residual entity infrastructure is learned iteratively in a residual section in DRRN. Stacking a handful of iterative sections together to prepare the residual data is done. After this, the remaining data is concatenated with the LR data from an identity section for regenerating the HR pictures. This method effectively learns lateral calibration from the LR intake data to the HR destination image. DRRN is a rooted, brief, and preferable example in the field of SISR; benchmark experiments and analysis have extensively proved this.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

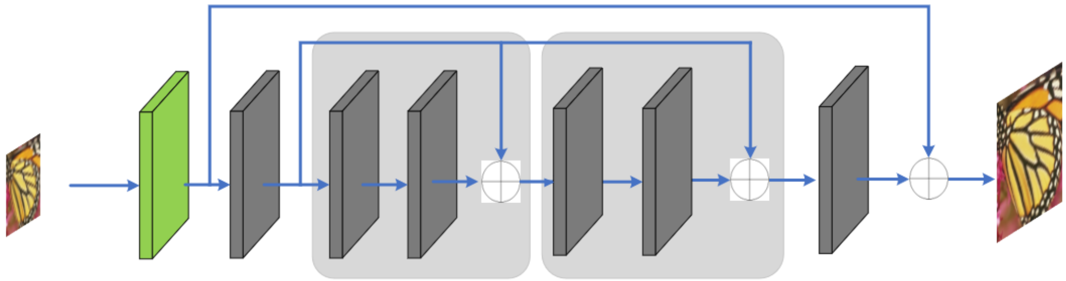


Figure 15 *Deep Recursive Residual Network (DRRN) Structure [40]*

A persistent memory network for image restoration (MemNet) [23] used a memory block in the network of SRCNN. This paper proposes an extremely rooted reminiscence network called MemNet. It familiarizes a recursive and a gate unit to store the information through an adaptive learning process. Most state-of-art models, such as VDSR and DRCN, use short-term memory since each state is directly linked with its former state. However, some models like ResNet have skip connections, which pass information for many layers. Thus, they are restricted to long-term memory. Deeper networks lack long-term memory. To remove this problem, MemNet has been introduced, which consists of a recollection block having a recursive unit and gate unit. The recursive unit is for storing short-range cognizance. The recollection blocks contain the longstanding cognizance. The information from recollection blocks and the recursive units is concatenated at the gate to maintain a persistent memory. So far, MemNet is the most rooted mesh for image restoration. The MemNet attains SOTA completion in denoising the image, SR, and JPEG unblocking. MemNet shows superiority over state-of-art methods on benchmark datasets.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

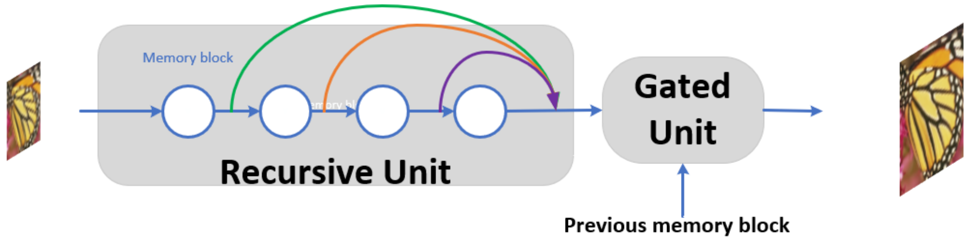


Figure 16 Memory Network for Image Restoration (MemNet) Structure [23]

2.2.4. Progressive Reconstruction

Lai et al. [6] proposed Laplacian Pyramid Super-Resolution Network (LapSRN). The method used a convolution layer to upscale the filter step to obtain the loss function. Then each output scale image result combines the loss. The result is high-quality, but the configuration is a high parameter number. This paper proposes a Laplacian Pyramid Framework (LapSRN) which uses progressive up-sampling to reconstruct fast and accurate residuals of HR images. The drawbacks of previous state-of-art models, such as high computational cost, blurry images, and learning difficulty, are overcome by LapSRN. This method uses cascaded CNNs to predict sub-band surplus in rough to fine texture. The model uses the Charbonnier loss function instead of L2 loss to reduce the prediction of blurry visuals. Instead of using predefined up-sampling methods like Bicubic Interpolation, the method uses progressive up-sampling LapSRN performs better in terms of accuracy, speed, and progressive reconstruction. The LapSRN method has 27 layers overall and takes LR as input, uses residual learning, and progressive reconstruction with a char bonnier loss function. LapSRN has two branches in its network architecture: Feature Extraction and Image Reconstruction. The proposed method constructs high-quality HR images at a relatively faster speed than other state-of-art methods. It also helps to remove the blur kernels. The only problem with the model is that it does not hallucinate fine details over large scales.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

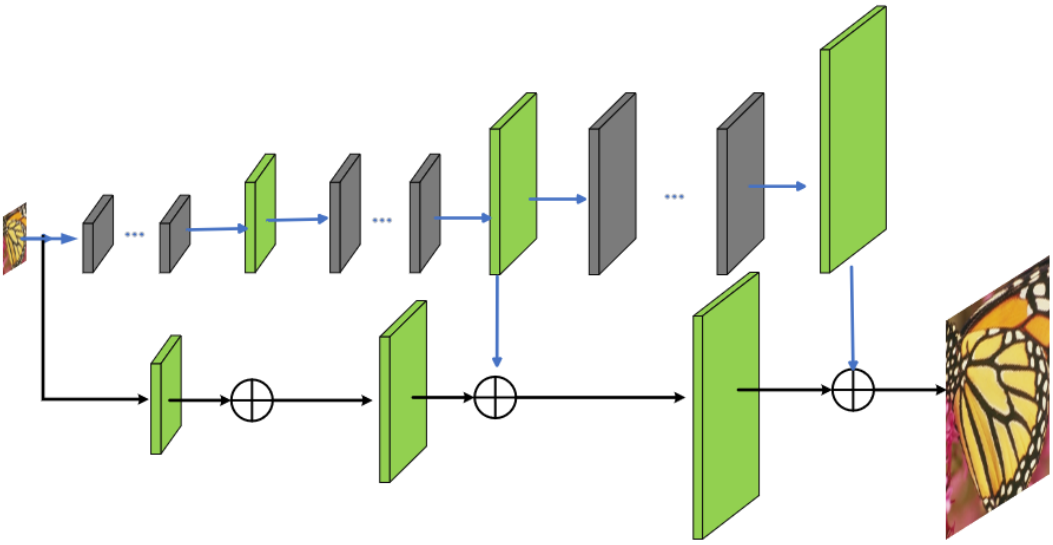


Figure 17 Laplacian Pyramid Super-Resolution Network (LapSRN) [6]

2.2.5. Densely Connected Networks

A residual connection connects dense connections in the Residual Dense Network (RDN) [26]. Most state-of-the-art methods do not use the information of each layer of the network, which can also be termed hierarchical information. Hierarchical features from deeper networks can give more clues for reconstructing HR images. The proposed method thus addresses these drawbacks and utilizes the information from each subsequent layer, or it could be said it does use the hierarchical features using the dense block with a low growth rate, making the training easier for the network. This model consists of a Residual Dense Block (RDB) acting as a building module for the RDN. RDB further consists of Local Feature Fusion (LFF) and densely connected layers. The output of one RDB is directly fed into the next RDB resulting in a continuous state known as Contiguous Memory (CM). This extracted info is adaptively preserved. It also preserves the hierarchical features extracted by Global Feature Fusion (GFF). This network extracts and adaptively fuses features from all layers in LR space. The Local Feature Fusion (LFF) stabilizes wider networks' training and adaptively controls the information preserved from present and previous RDBs. The Local Residual Learning used in the network improves the

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

flow of information and gradient. By fusing the features extracted from local and global feature fusions, RDN leads to deep image supervision. The model achieves superior performance over five benchmark datasets compared to state-of-art models.

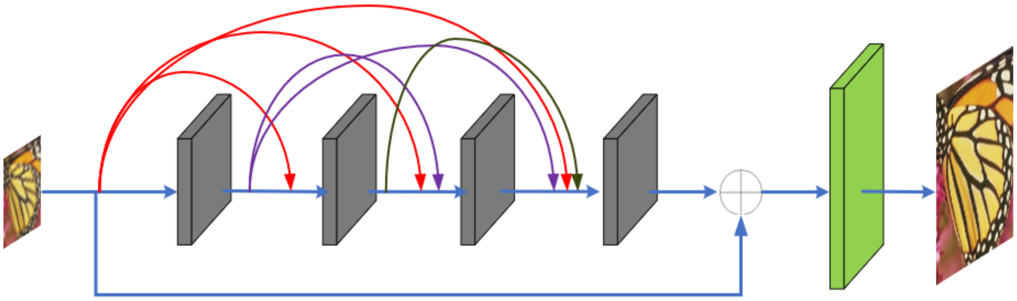


Figure 18 *Deep Residual Dense Network (RDN) Structure [26]*

The network represents flexibility in SR images. Also, Dense Deep Back-Projection Network for super-resolution (D-DBPN) [27] adds feedback from LR and HR images with multiple up and down scale blocks. But the method-made network has high complexity. The proposed Deep Back Projection Networks focuses on the mutual dependencies of the LR and HR images by exploiting up-sampling and down-sampling layers of the network and by providing a feedback mechanism for projecting errors in every stage. Unlike other feed-forward networks, methods use multiple iterative up and down sampling stages and a feedback network to increase the extracted SR features. Network architecture consists of three parts:

1.) Initial feature extraction-This stage reduces the dimension of the LR image to make it suitable for entering the projection stage.,

2.) back-projection stage- This unit alternates the construction between the LR and HR design types in which each unit has access to the output of all previous units.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.) Reconstruction stage- In this stage, finally, the HR picture is obtained by concatenating the design maps created in each protrusion entity.

The proposed network shows better results on large scaling factor 8 enlargement and five benchmark datasets than other state-of-art methods.

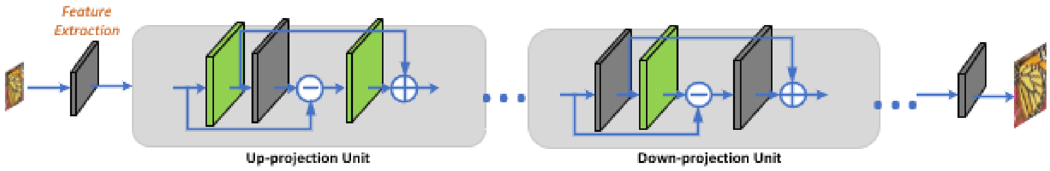


Figure 19 Deep Back-Projection Networks (DBPN) for super-resolution [27]

2.2.6. Multi-Branch Designs

The network is a well-known fact that all the CNN-based models have faced the problem of computation complexity and memory consumption. To solve this drawback, a deep but compact CNN has been proposed. This IDN model [41] consists of distinctive parts: -n element abstraction module, linearly shaped information refining, and a representative block. The proposed IDN has multiple DBlocks (enhancement unit + compression unit) for extracting short and long feature maps directly from LR images and generating residual reconstructions. IDN is a concise structure, so it is much better in speed and accuracy than several CNN-based SR methods. The proposed method uses group convolutions and performs favorably great in terms of fast execution and fear number of filters per layer on most datasets. This method also performs great in the Information Fidelity Criterion (IFC) metric, which correlates with the human perception of images.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

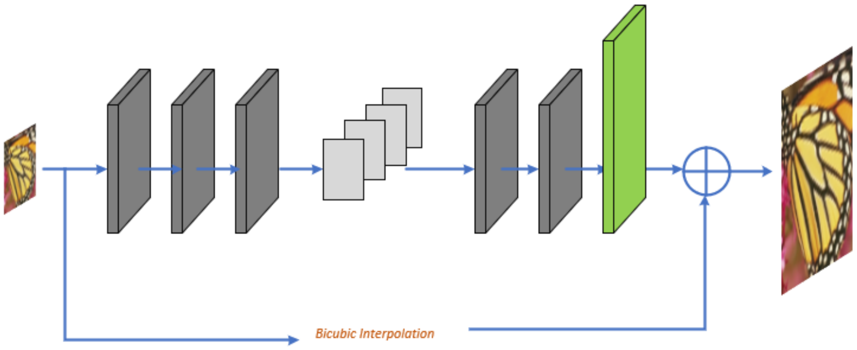


Figure 20 Fast and Accurate Single Image Super-Resolution via Information Distillation Network (IDN) [41]

Recently designed Deep Networks have powerful representational capabilities but excessive computations and cannot be implemented over arbitrary scaling factors. To address these drawbacks, a lightweight information multi-distillation network (IMDN) [41] has been proposed containing selective and distillation parts. The distillation parts extract the hierarchical features, and the fusion module aggregates them. The fast and accurate results of the model are due to the information provided by the multi-distillation block IMDB and the Contrast wave attention layer CCA. The method incorporates an adaptive cropping strategy (ACS), allowing the network to have down sampling operations to work with images of any arbitrary scale. This model can reduce the computational cost, memory occupation, and inference time by using ACS. By using this model, the actual factors affecting the inference time can be interpreted. Experimental evaluations show that the proposed network shows great results in visual quality, speed, and memory consumption; also, this method achieves balance in all the factors affecting practical use.

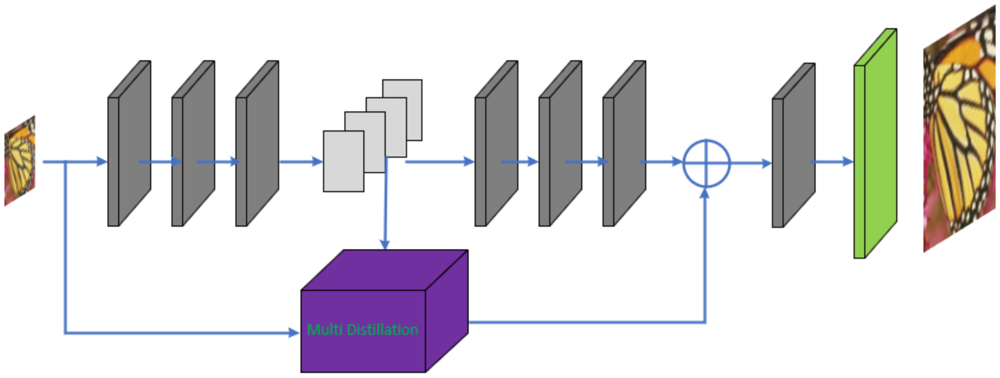


Figure 21 Image Super-Resolution with Information Multi-distillation Network (IMDN) [41]

2.2.7. Attention Based Networks

The SelNet [42] Deep CNN with Selection Units for SR is motivated by the linear mapping techniques used in other CNN architectures. The Rectified Linear Unit (RLU) was used for linearly mapping the LR images, inspired to create a non-linear unit known as Selection Unit (SU). Since SU combines identity mapping and a sigmoid switching function, it has better control over data passed through than ReLU. The SU can better handle the non-linearity functionality and is more flexible than ReLU. Thus, this model gives better qualitative and perceptual results. Results show the proposed network has much lower computational complexity and outperforms baseline only with ReLU and other SOTA DL-based SR methods.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

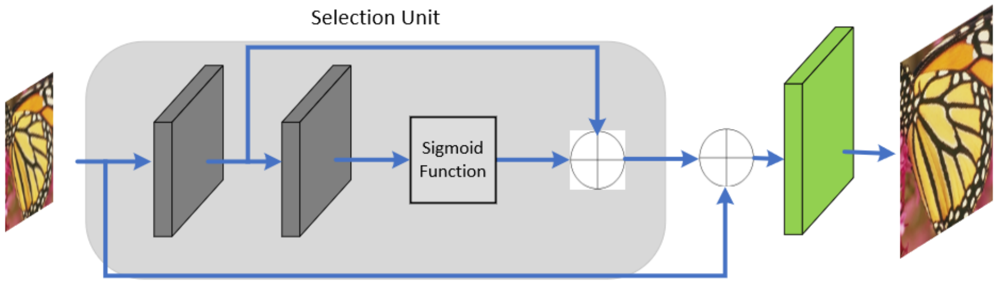


Figure 22 Deep residual learning for image recognition Network (SelNet) [42]

The result is high-quality, but the configuration is a high parameter number. Attention Network Currently, state-of-the-art of SR [20, 30] is Residual Channel

Attention Network (RCAN) [28] utilizes residual in residual (RIR) consisting of Residual Group (RG) and Residual Channel Attention Block (RCAB). The structure RG used a short skip connection network as a residual component, RCAB aspects long skip connection network to target feature component of Low Resolution (LR). Addition channel attention (CA) introduces to affect the feature rescale channel. The complexity of the processing takes time with a higher number of parameters.

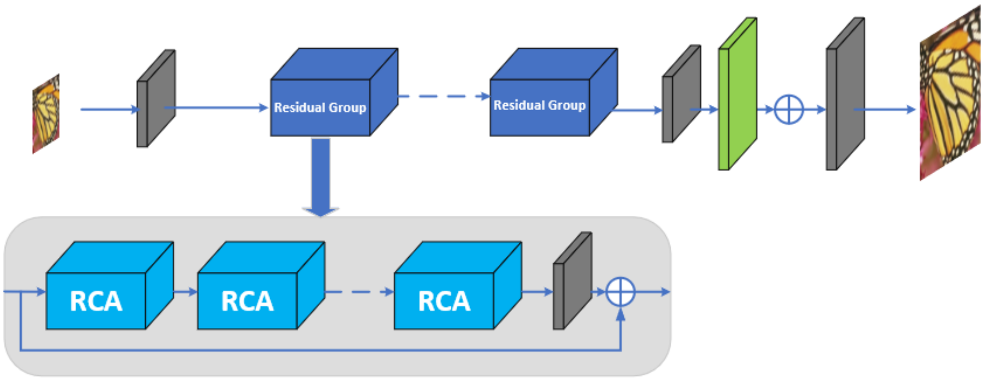


Figure 23 Residual Channel Attention Network (RCAN) [28] Structure

2.2.8. Multiple Degradation Handing Networks

Also, Learning a single convolutional super-resolution network for multiple degradations (SRMDNF) [24] applied noise-free degradations to unique efficiency PSNR. A Convolutional SR network for Multiple Degradation has been proposed to deal with multiple degradation processes of SISR. This network has high flexibility in supervising many deteriorations using a SISR model. This technique mainly focuses on exploiting the reprisal of CNN, such as alacrity, computing, exactitude, and advances in training design networks. While these CNNs give great results for a single simplified degradation process, these CNNs fail to give good results for multiple degradations or if the degradation deviates from a simplified degradation.

To solve these issues, CVPR has been proposed. -It gives an elementary, compelling, and rooted CNN frame for single model SR. This model works for simple

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

bicubic degradation and applies for numerous and spatially dissimilar deteriorations. - This technique also gives a stretching strategy to look for dissimilarity in the dimensions of LR input data, unclear kernels, and noise levels. We show that the dictated framework obtained by learning from unnatural image data also gives competitive results for real LR images.

The proposed method takes the concatenation of LR images and degradation maps (i.e., unclear kernels and turbulence levels) as input. It allows a SISR model to work with diverse and spatial degradation processes. The proposed model favors multiple and special variant degradations; evaluations are based on benchmark datasets.

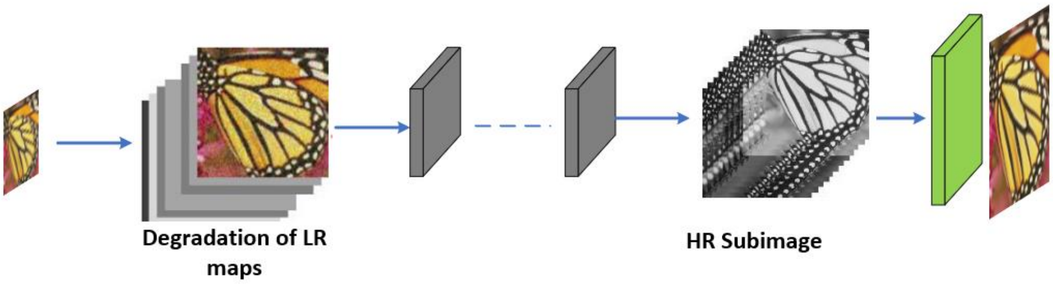


Figure 24 Super-Resolution Network for Multiple Degradations Structure (SRMDNF) [24]

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.2.9. Generative Adversarial Nets Network

Lastly, Generative Adversarial Nets Network [29] creates SR near-natural image by MSE loss with matric similarity and adversarial loss. Although the outputs of this network have better visual quality, the PSNR values are lower than other methods. This paper presents SRGAN Generative Adversarial Network for SR. This method is used to resolve finer details over large scaling factors. For the other state-of-art models, capturing perceptual differences with great texture details is hard. This framework has been proposed with perceptual loss functions consisting of adversarial and content loss. The adversarial loss uses a discriminator network and gives results like natural images. The content loss is used, which gives similarity in perception instead of similarity in pixels. Thus, GAN networks are good for generating beautiful natural images with high perceptual quality. The proposed network shows great performance in MOS Testing. This method also shows that the standard qualitative measures like PSNR or SSIM fail to capture the image quality w.r.t the human visual system. The proposed SRGAN show comparable results on public benchmark datasets. It also shows some limitations of qualitative measures of image SR. Finally, it shows great reconstructions of images for large upscaling factors 4.

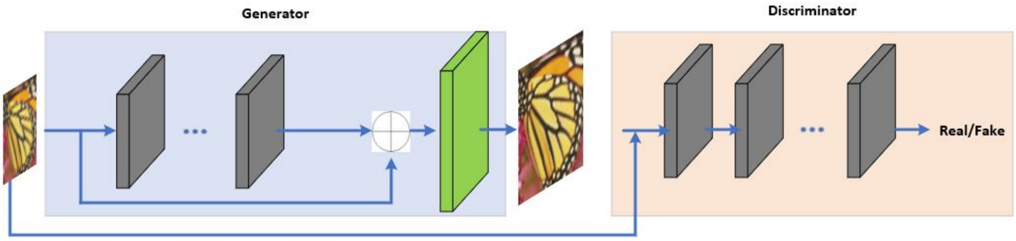


Figure 25 Generative Adversarial Nets Network Structure [29]

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

2.3. Super-Resolution on Hardware

The SR research on the hardware can be into three methods, similarly to the previous section.

2.3.1. Interpolation based on hardware

Bowen et al. [43] used the multi-frame interpolation based on the total weight mean SR with an iterative, as shown in Figure 26. The testing was done on the Xilinx Virtex II XC2V6000 FPGA board with 4 MB internal and external ram 8MB HR output with HD (1280x720) at 60 fps. From the experimental results, the HR image quality is not good, and noise occurs.

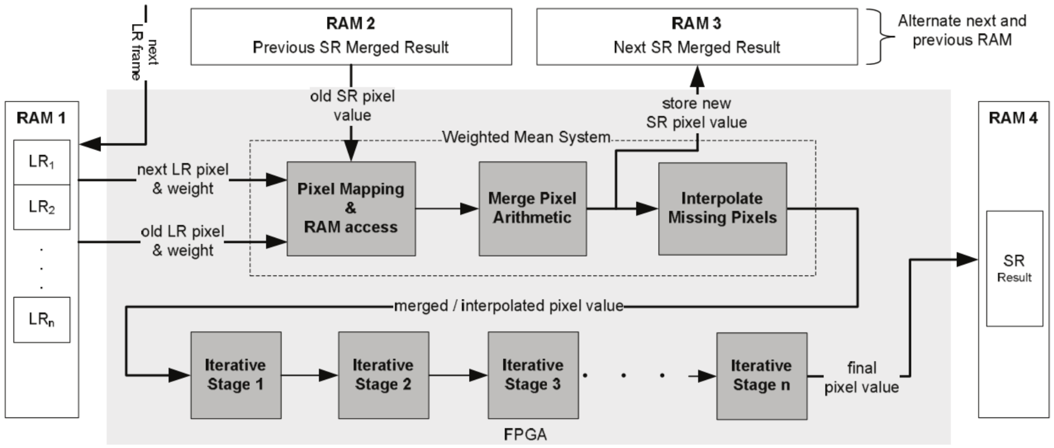


Figure 26 Interpolation based on Hardware [43]

Lee et al. [44] proposed to upscale the image from Full High Definition (FHD 1920x1080P) to 4K ultra-high-definition (3840x2160P). The system upscales the image using Lagrange interpolation, and the interpolation error improves the sharpening filter, as shown in Figure 27. The operation in the system use adds and shifts for computation. HR image used a processing time of around 30 fps.

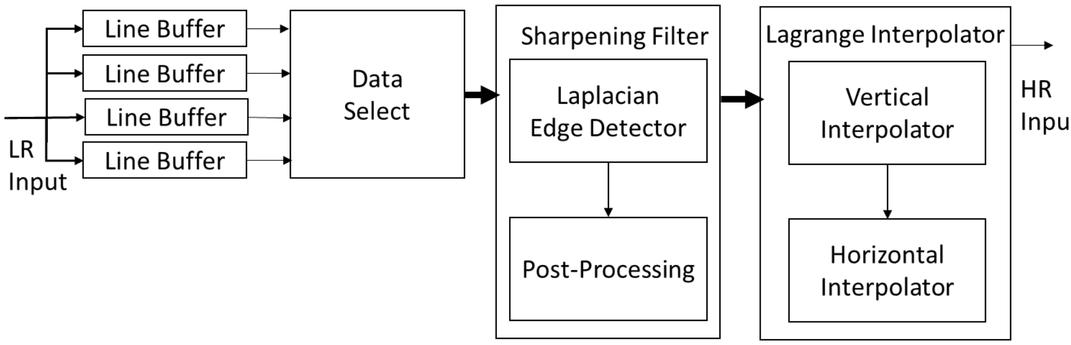


Figure 27 Interpolation based on Hardware [44]

2.3.2. Multiple frames based on hardware

Seyid et al. [45] proposed creating HR using multiple frames, as shown in Figure 28. The sub-pixel differences and the iterative back-projection are parallel processing on FPGA. The image HR output size is 512x512 pixels at 25 fps from processing 20 LR images.

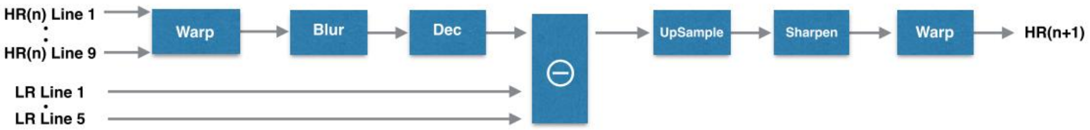


Figure 28 Multiple Frame based on Hardware [45]

Redlich et a.l [46] used infrared cameras size 160x120 pixel to create HR size 640x480 pixels using multiple-frame registration. As Figure 29, the pyramid generator reduces the image size to help in registration images, which are processed at 150 FPS since the images are small. The implement board is Xilinx Spartan-6 LX45 FPGA.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

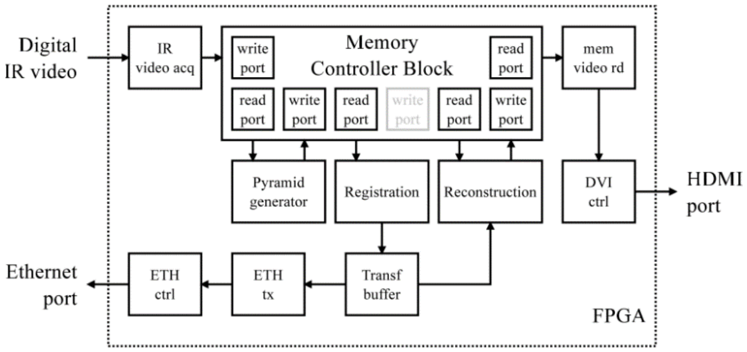


Figure 29 Multiple-frame registration on hardware [46]

2.3.3. Learning-Based Method

Yang et al. [47] proposed an SR system that uses five input line buffers and eight output line buffers for output image FHD size at 60 fps using a dictionary learning-based approach. They used anchored neighborhood regression (ANR) to create output with feature extraction from principal component analysis (PCA), reducing dimension for producing HR patch. As a result, the method used high memory and logic gates in FPGA.

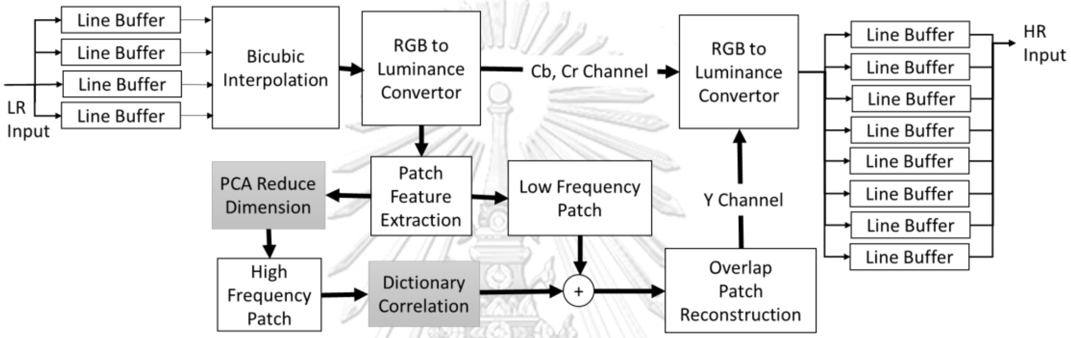


Figure 30 *Block Diagram of Learning-Based Method on Hardware [47]*

Kim et al. [48] used application-specific integrated circuit (ASIC) and FPGA to implement with lower memory and gate. The method used two input line buffers. The four output line buffers were used for 4 K UHD at 60 fps using an edge orientation in the training step.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

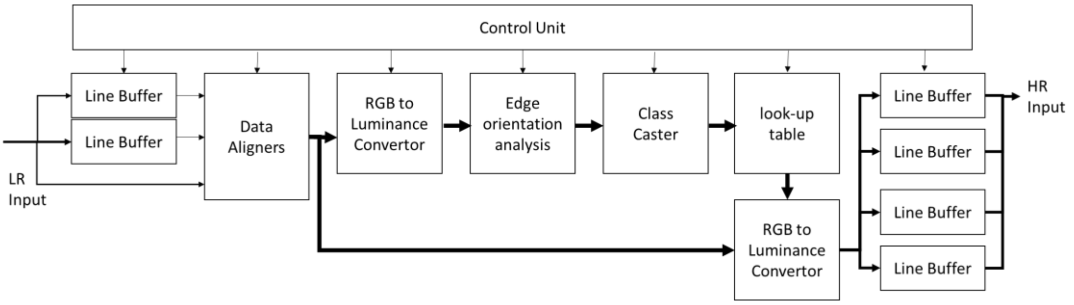


Figure 31 *Block Diagram of Learning-Based Method on hardware [48]*

This novel design is based on Field Programmable Gate Array (FPGA) for real-time Video SR [49]. The learning-based VSR techniques suffer from the following two problems: -large computational complexity -temporal inconsistency. To solve these issues, ERVSR exploits the temporal information of LR space in a hidden way and represents HR in the residual framework. It improves model efficiency and reduces the parameters by performing convolutions on LR input and hidden state and introducing channel modulation coefficient. Statistical normalization and fixed-point quantization compress the hidden state and reduce memory consumption. And to reduce the network complexity, separable and group convolutions are adopted. The model ERVSR is evaluated on multiple datasets. Results demonstrate it performs better than SOTA models. The proposed model also helps improve the temporal consistency by using lightweight recurrent architecture.

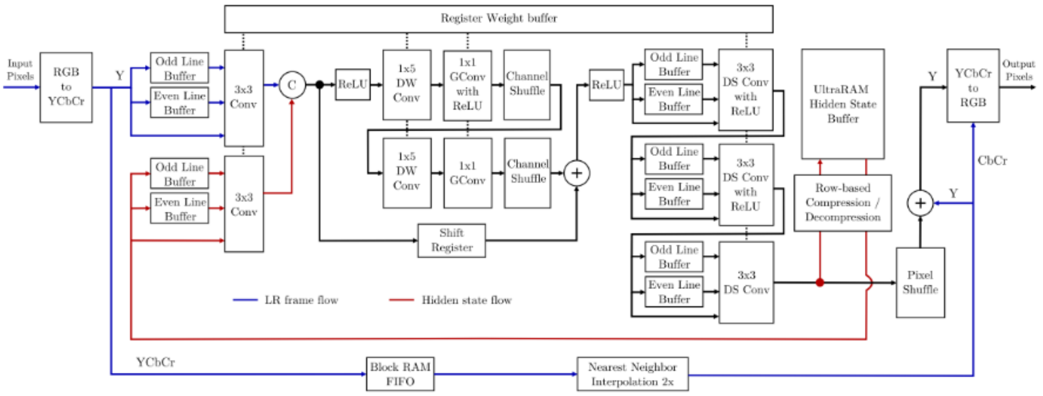


Figure 32 Field Programmable Gate Array (FPGA) for real-time Video SR [49]

2.3.4. Deep Learning Method

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

The classification CNNs and the Super Resolution CNNs [50] require a lot of memory access which is not good for the mobile user experience. Even though FPGAs are good replacements, they still face the issue because of large on-chip memory and can produce the throughput of 5fps, which is insufficient for small edge devices. To resolve these issues, three techniques have been classified. a.) Layer fusion based on selective caching, b.) memory compaction, c.) cyclic ring core architecture. The proposed Processor infrastructure consists of a controller, local rings having a convolutional core, weight memory parallel processing engines. Router and an accumulator, all fabricated in CMOS process having 50-200 MHz clock frequency. The designed processor shows 60fps throughput for scale x4 and 25 fps for scale x2, and 60 fps for full high-definition resolution.

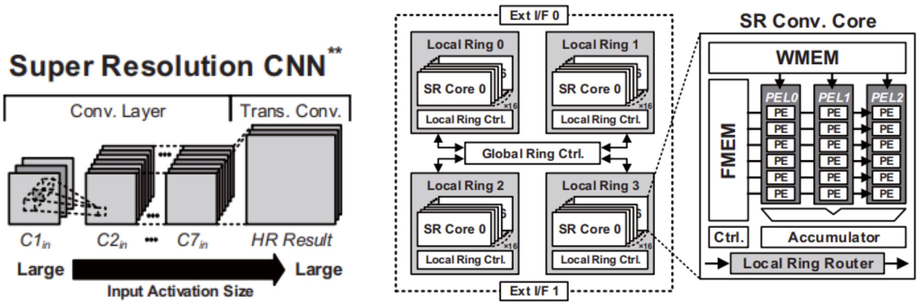


Figure 33 Block Diagram of Deep Learning Method on Hardware [50]

Computational complexities are the major prohibition for implementing SR on small edge hardware. To solve this issue, a firth method is proposed implementing real-time CNN hardware upscaling 2K FHD video frames to 4K UHD video at 60 fps using FPGA. The proposed method [51] contributes to: Implementing SR on hardware utilizing less memory by processing the input line by line and maintaining the parametric values and proposing a cascaded structure of 1D convolutions for sustaining large horizontal receptive fields while keeping smaller vertical fields. This network also reduces the number of filter parameters by introducing depthwise separable convolution with residual connections and introducing a simplified and effective quantization process to convert 32-bit FP data to fixed point data and a compression method to reduce line memory. The proposed SR-based Hardware can efficiently reconstruct 4K UHD at 60 fps and can be implemented for more than twice up-scaling or arbitrary up-scaling.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

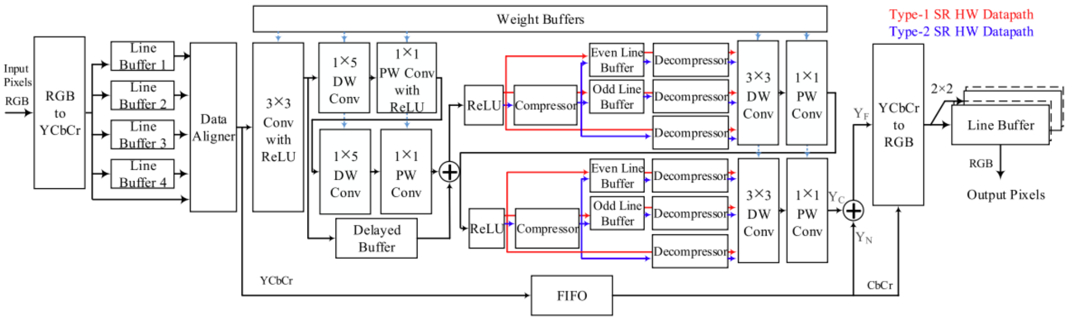


Figure 34 Block Diagram of Deep Learning Method on Hardware [51]

To decrease the computational load and enlarge the utilization of the process element (PE) of the CNN-based SR approach, Lee et al. [52] proposed a data flow for hardware-effective purposes by enlarging data reuse and equally assigning the computational workload for each layer. The proposed data flow calculated the pixel in the receptive field row-wise by arranging memory addresses with a circular shift. To reduce the expensive calculation of a single pipeline step, they utilized the partial convolution-based pipeline architecture instead of the conventional convolution, as shown in Figure 35. To support a 4K UHD video stream with 60 fps by embedding the proposed architecture on an FPGA.

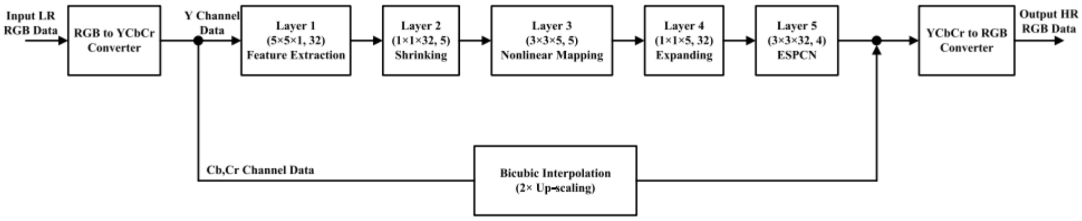

Input LR RGB Data
RGB to YCbCr Converter
Y Channel Data
Layer 1 (5×5×1, 32) Feature Extraction
Layer 2 (1×1×32, 5) Shrinking
Layer 3 (3×3×5, 5) Nonlinear Mapping
Layer 4 (1×1×5, 32) Expanding
Layer 5 (3×3×32, 4) ESPCN
YCbCr to RGB Converter
Output HR RGB Data
Cb,Cr Channel Data
Bicubic Interpolation (2× Up-scaling)

Figure 35 Block Diagram of Deep Learning Method on Hardware [52]

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter 3

Image Super-Resolution

The image Super-resolution (SR) aims to reconstruct the high-resolution (HR) image from the low-resolution (LR) input. The SR is considered an ill-posed problem. Therefore, there are many existing SR techniques to produce the HR image. However, the SR image's perceptual quality and the SR model's execution time are two critical factors in designing an effective and robust SR model for applying in real-world applications. A real-world problem in utilizing closed-circuit television (CCTV) [53] for intelligent monitoring [54] such as traffic congestion and accident [55], home management security [56], face recognition [57], and social identification [58] is appeared due to the low-resolution (LR) images produces by the CCTV camera [59].

Many research topics such as feature preservation in video coding [60, 61] and super-resolution [36, 62, 63] have been published to solve the low-resolution problem. Deep Convolutional Neural Networks (CNN) have recently been a powerful model in solving the ill-posed SR problem [30]. Dong et al. [5, 21] proposed Super-resolution using deep Convolutional Neural Networks (SRCNN) model with the architecture of a three-layer network that is linearly stacked together. In this shallow and straightforward architecture, the first layer is designed to extract the features from the LR input, the second layer is used for the non-linear mapping from low-dimension to high-dimension features, and the final layer is responsible for aggregating the previous layer’s feature maps to the final HR result. This SR network is trained in an end-to-end approach to minimize the Mean Squared Error (MSE) between the ground truth (GT) images and the reconstructed (SR) image. Following by SRCNN model, other models, including Very Deep Super-resolution [7] (VDSR) and Deeply Recursive Convolutional Network (DRCN), achieved significant improvements compared to the SRCNN and FSRCNN models by increasing the depth of their CNN architectures.

Following DRCN and VDSR models, various architecture models based on different approaches, including the residual network [31-39], Recursive Network [8,

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

23, 40], progressive reconstruction [6], dense connection network [26, 27] , multi-branch architecture [41, 64], attention-based [28, 42] mechanism have been proposed to improve the result of SR model by increasing the data flow of information in the deep learning architecture.

Inspired by Residual network [31] (ResNet) architecture, some SR models attempted to design deeper SR architecture. Based on the residual concept, Lim et al. proposed a very deep SR architecture known as Enhanced Deep Residual Network [25] (EDSR). This SR model stacks residual blocks to design deeper networks (almost 165 convolutional layers) and achieved considerable improvement compared to the earlier SR models. However, training such a deep network trainable network like EDSR [25] is challenging, and the numerous network parameters of a very deep network are an important obstacle for fast execution in real-world applications and hardware implementation.

On the other hand, some recent CNN-based models [6, 8, 39, 41, 64] utilized multi-path network architecture to solve the limitations of the deep network. In this approach, rather than using a single-path network, the multi-path of the network operates in parallel. Based on this concept, the depth of the network is decreased while the performance of the SR model is increased. According to the Multi-path structure, Hui et al. proposed Information Distillation Network (IDN) by designing the cascaded network paths operating in parallel. It means the layers of the network do not require to wait for the previous layer calculation. Additionally, the different types of the extracted features of each path are mixed, consequently improving the SR model's operation time and perceptual result. Although the SR models based on the multi-path approach get acceptable performance and execution time, their results archived low PSNR and SSIM.

Additionally, the low-level feature sharing approach is used in some SR models [34, 35, 37, 42] to enhance the low-frequency information flow in the SR network structure. This low-level feature-sharing approach attempts to transfer the low-level features of the early CNN layer to the latest layers and fuse the low and high-level features. Hence this technique improves the reconstruction quality by

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

enhancing the sharpness of the SR Image. Motivated by this technique, the SR models such as Feedback Network [34] (SRFBN), Adaptive Weighted Super-Resolution Network [35] (AWSRN), SelNet [42], and Attentive Auxiliary Features [37] (A2F) utilized the feature sharing approach. This idea demonstrates more effective enhancement in the lightweight designed SR architecture.

To deal with these issues, an efficient lightweight SR model for image enlargement has been proposed, stating the *main contributions* as follows,

1. The multi-depth cross-channel network is proposed to obtain the local pixel attention feature from a low-resolution image.
2. The proposed network also investigates the doubling stage of residual identity connection for retrieving the merged features represented by low-level features.
3. This method explores low-level feature sharing to fuse low-level feature information into unsampled features to enhance the reconstruction ability of the model.

3.1. Multi-FusNet of Cross Channel Network

The proposed network architecture is based on a multi-path residual architecture that designs a wider network rather than a deeper one. The proposed network, named Multi-FusNet of Cross Channel Network (MFCC), consists of four main modules, including feature extraction, multi residual group (Multi RG), enlargement, and low-level fusing, as shown in Figure 36 . The Multi-RG architecture has been designed by integrating the RCAN [34] with a multi-identical residual link. Due to the increase in the number of multiplicity (possible paths from the input to the output layer) in the proposed architecture, the information flow between Multi-RGs gradually increases and helps to reduce computational complexity.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

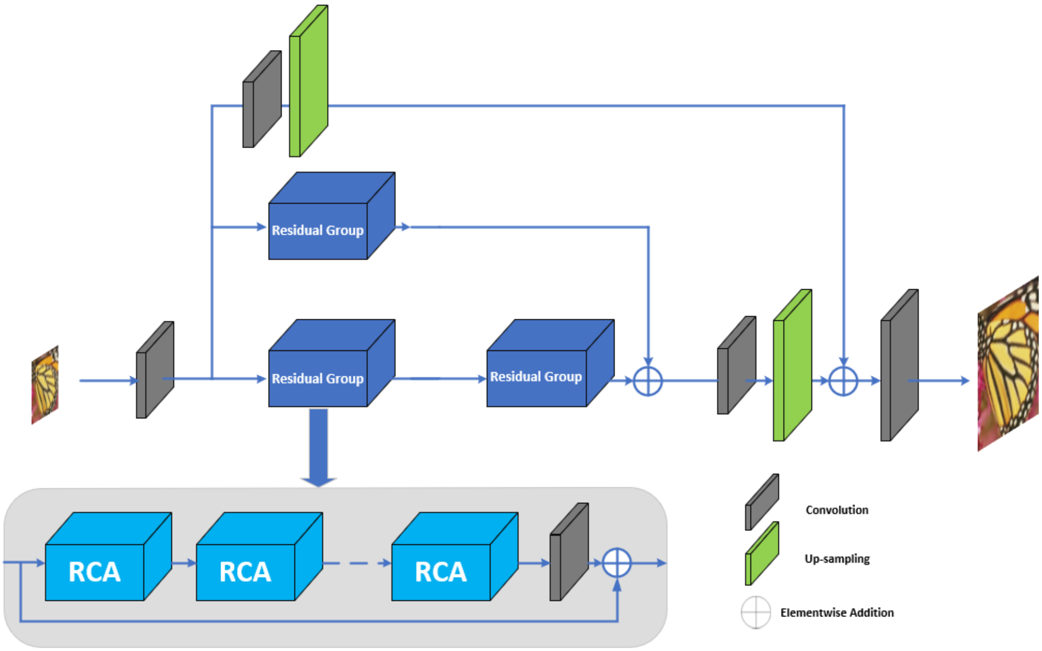


Figure 36 *Multi-FusNet of Cross Channel Network for Image Super-Resolution*

The pixel shuffle [65] technique is used as an enlargement module in stated architecture. This technique transforms the low-level feature maps into different channels, transforms the shuffling operation on features, and enlarges the features. The first convolution layer is the feature extraction module which feeds the features to the RG blocks and the fusion module. The idea behind the fusion module is to share the low-level features into the up-sampled features of the multi-RG.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

The MFCC network is designed according to the cascading approach, where three different paths constitute the proposed multi-path residual architecture, as shown in *Figure* 36. The Residual Group (RG) block consists of N stacked Residual Channel Attention (RCA) blocks and a short residual skip connection within the block.

As shown in *Figure* 36, our model designs Residual Group blocks in the first path and one Residual Group block in the second path. The third path of the model aims to share the low-level features of the earlier layer with the up-sampled features of the multi-path residual. The LR input is first fed to the convolutional layer. The output is then fed to three different paths. The kernel size of each convolutional layer in the proposed network is k×k. Due to many parameters in the RCAN model, it is challenging to use the model in real-time applications such as hardware implementation. The proposed method aims to reduce the number of residual groups (RGs) in our architecture to have a lightweight architecture.

Additionally, a multi-path residual network architecture is incorporated into our model. Since the multi-path residual architecture can improve accuracy and increase the processing speed, it exhibits more efficient performance in lightweight networks compared to the residual architecture. In other words, we design broader lightweight architectures rather than deeper models. The third improvement is to exploit the low-level features of the early CNN layer and share them with the up-sampled features of the multi-path residual network. Due to the lack of high-frequency details in the latest layer of CNN networks, the results of most SR models suffer from over-smooth degradation and show weaknesses in extracting sharp attribute details. Consequently, the resulting images are perceptually unpleasant at large scales.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

To address these limitations, a pixel shuffling fusion method is used to fuse low-level features from early layers into the up-sampled features from the latest layers of the proposed multi-path residual network. The following section describes the residual group block, multi-path residual configuration, and pixel shuffle fusion method.

3.1.1. Residual Group:

In this section, the Residual Group (RG), which is the robust feature extraction of low-resolution, is presented. The proposed Residual Group is constructed by applying identity residual connection at the edge of *N* sequences of the statistical

Channel Attention network in [28], where $N>0$. The output of RG, O_{RG}, can be defined as in Eq. (7).

$$O_{RG} = Y_{RG} + W_{RG}(F_{RCA_N}) \tag{7}$$

Where Y represents the input of RG. W_{RG} denotes weight parameter in Residual Group block, and $F_{RCA_N}$is the channel-wise feature from Residual Channel Attention block (RCA). The channel-wise feature from RCA can be determined by Eq. (7) and (8).

$$F_{RCA_N} = {H^N}_{RCA}\left(F_{RCA_{N-1}}\right) \tag{8}$$

.

$$F_{RCA_0} = {H^0}_{RCA}(I_{RCA}) \tag{9}$$

Where F_{RCA_N} and $F_{RCA_{N-1}}$ denote the output of N^{th} and $(N-1)^{th}$ Residual Channel Attention blocks, respectively. H_{RCA} shows the corresponding operation function of RCA. F_{RCA_0} is the output of the first Residual Channel Attention block. Y_{RG} shows the input of the first Residual Channel Attention block. The Channel Attention (CA) mechanism is a technique that uses the interdependencies among the feature channels. The CA technique leads to more focus on informative features in the SR model and consequently improves the image reconstruction capability of the model. More details of the CA mechanism and the corresponding operation function in the Residual Channel Attention block can be found in RCAN [28].

3.1.2. Multipath Residual:

As demonstrated in *Figure* 36, the combination of Residual Group blocks under the multipath-residual architecture [66] is used in our model. Based on multi-path residual evidence [66, 67], a wider residual architecture significantly improves the accuracy and computation speed of the model compared to a deeper residual architecture. These improvements are related to raising the multiplicity in the wider residual network. The multiplicity implies the number of possible paths from the input layer to the output layer. A sequence of two Residual Group blocks is utilized in the first path of our model, and one Residual Group block is employed in the second path. Eq.4 defines the multi-path output of our model.

$$O_{MR} = O^{2}{}_{RG} + O^{1}{}_{RG} \tag{10}$$

Where $O2_{RG}$ and $O1_{RG}$ denote the output of two Residual Group blocks in the first path and output of a Residual Group block in the second path, respectively. O_{MR} denotes the output of the proposed multi-path residual architecture.

3.1.3. Pixel Shuffle Fusion

The low-level feature-sharing approach is employed to improve the reconstructed results' sharpness. Since the low-level features of the early layer contain more high-frequency information, sharing them improves the challenging weakness of the SR model to recover the sharper attributes of the lines and edge. At the same time, it preserves our model against over-smoothing degradation. Our model utilizes the pixel shuffle fusion approach to bypass the low-frequency feature of the early layer of the SR network to the up-sampled features.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

The proposed model employs the pixel shuffle [65] technique to up-sample the image. Based on the features sharing concept, the features of the early layer also are up-sampled by the pixel shuffle model and fuse with up-sampled features of the multi-path residual network, as demonstrated in *Figure* 36.

$$PS_{Fus} = PS_1\big(W_{MR}(O_{MR})\big) + PS_2(W_p(F_L)) \tag{11}$$

Where, PS_1 and PS_2 represent the pixel shuffle up-sampling on a multi-path residual network and pixel shuffle up-sampling on low-level features sharing. W_{MR} and W_p demonstrates convolution operations of the multi-path residual output and up-sampled low-level features. The pixel shuffle mathematically can be shown in (12).

$$PS(U)_{i,j,c} = U_{[{}^{i}/_{a}],[{}^{j}/_{a}]},C \cdot a \cdot mod(i,a) + C \cdot a \cdot mod(i,a) + c \tag{12}$$

Where $PS(U)$ is output, a demonstrates scale factor, $i, j,$ show pixel coordinates. $\boldsymbol{C}$ is the position of the channel. To modernize the final SR image, the up-sampled high-frequency details of the early layer are fused with the up-sampled features of the multi-path residual model.

$$SR = W_{SR}(PS_{Fus}) \tag{13}$$

Where PS_{Fus} denotes the fusing pixel shuffle result, W_{SR} defines the last convolution operation to produce the final SR result. Utilizing the proposed fused approach improves the capability of our model to recover the sharp attributes of images and improves the perceptual quality of results by preventing an over-smoothing problem.

$$L_1(\emptyset) = \frac{1}{n \times m} \sum_{i=1}^{n} \sum_{j=1}^{m} \|SR(i,j) - y(i,j)\| \tag{14}$$

Where SR and y demonstrate the result and the reference image, n and m are parameters related to the training dataset.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

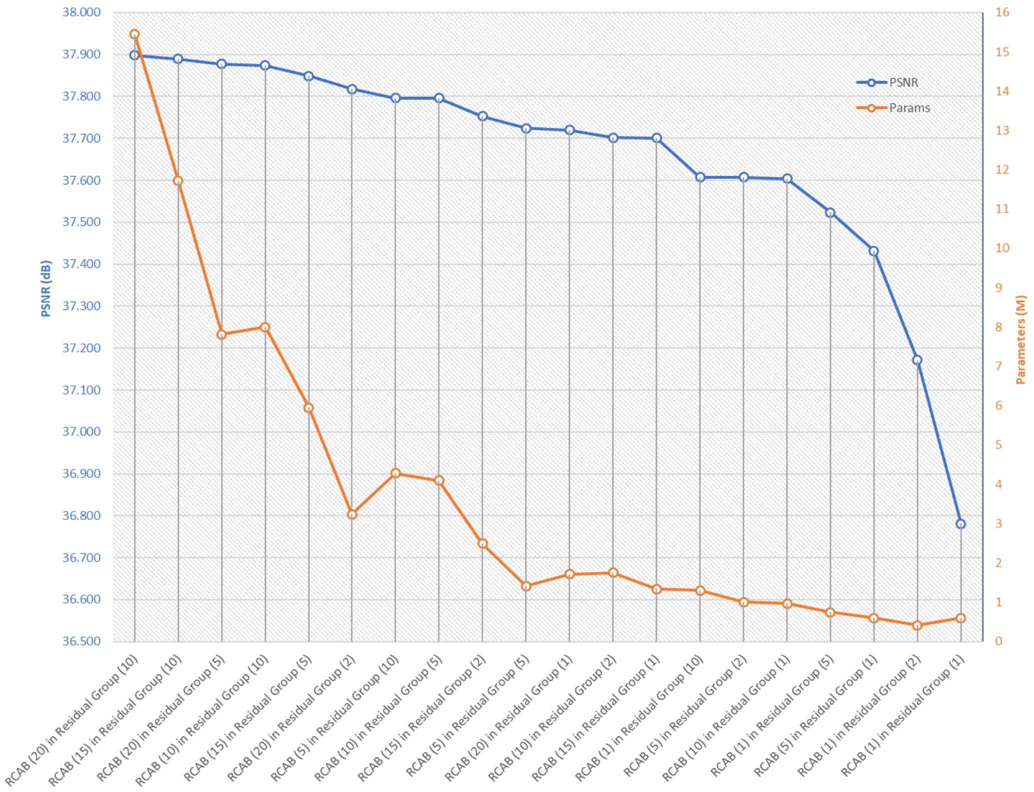


Figure 37 PSNR vs. Parameter

We have varied the change of RCAB and RG from (20,10) to (1,1). We noticed that the parameter's curve would be saturated from RCAB 5, RG 5 to RCAB 1, RG1. It means that the parameter will not be able to reduce much by decreasing both layers RCAB and RG under five layers. Our target is to obtain the minimal parameter and obtain high PSNR. In terms of this characteristic, we consider the curve's median the optimal point we could use as the selected parameters. We have computed the median of the existing data. We found that the suitable number of RCAB and RG is RCAB 20 RG 1. Then, we select RCAB 20 and RG 1 as our final architecture.

Table 1 BASED LINE MODEL AT 50 EPOCH ON SET5 (2x)

Model	PSNR	Params
RCAB (20) in Residual Group (10)	37.90	15,444,667
RCAB (20) in Residual Group (5)	37.88	7,816,427
RCAB (20) in Residual Group (2)	37.82	3,239,483
RCAB (20) in Residual Group (1)	37.72	1,713,835
RCAB (15) in Residual Group (10)	37.89	11,722,867
RCAB (15) in Residual Group (5)	37.85	5,955,527
RCAB (15) in Residual Group (2)	37.75	2,495,123
RCAB (15) in Residual Group (1)	37.70	1,341,655
RCAB (10) in Residual Group (10)	37.87	8,001,067
RCAB (10) in Residual Group (5)	37.80	4,094,627
RCAB (10) in Residual Group (2)	37.70	1,750,763
RCAB (10) in Residual Group (1)	37.60	969,475
RCAB (5) in Residual Group (10)	37.80	4,279,267
RCAB (5) in Residual Group (5)	37.72	1,415,511
RCAB (5) in Residual Group (2)	37.61	1,006,403
RCAB (5) in Residual Group (1)	37.43	597,295
RCAB (2) in Residual Group (10)	37.70	2,046,187
RCAB (2) in Residual Group (5)	37.58	1,117,187
RCAB (2) in Residual Group (2)	37.42	559,787
RCAB (2) in Residual Group (1)	37.15	373,987
RCAB (1) in Residual Group (10)	37.61	1,301,827
RCAB (1) in Residual Group (5)	37.52	745,007
RCAB (1) in Residual Group (2)	37.17	410,915
RCAB (1) in Residual Group (1)	36.78	597,295

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

In addition, by checking the effectiveness of RCAB and RG, we found that if we reduce the layer of RCAB, the performance of PSNR is dropped faster than decreasing the layer of RG Moreover, by reducing the layer of RCAB, the performance of the parameter is dropped less than decreasing the layer of RG Based on this consideration, we chose the RCAB as 20, and RG as 1.

Multi-Residual Networks [68] work residual blocks by expanding residual information to a broader network model. The network with the design of a residual network is called a multi-Residual network. This network can improve the processing speed by removing some parts of the sequential block from a parallel block. It is also noted that the results show better accuracy.

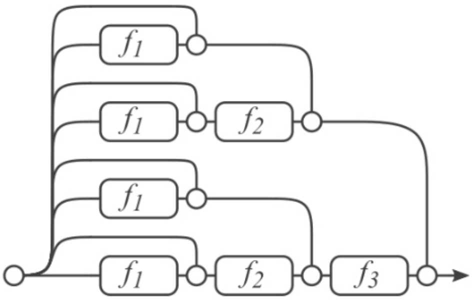


Figure 38 Multi-Residual Networks [68] Structure

Multi Residual Based on the experimental results shown, reducing the number of RCABs considerably affects the quality. Since the PSNR value of the baseline model RCAN (RCAB (20) in Residual Group (10)) is 37.513 dB, the goal of the proposed method is to keep that quality in terms of PSNR and reduce the number of parameters and processing time. An experiment is conducted by training the model with 50 epochs using the DIV2K dataset and testing the model with Set5 in scale 2. According to the observation, RG has a negligible impact on quality. Therefore, only one RG block is decided to be used to reduce the number of parameters. On the other hand, if the number of RCABs is reduced, there is a considerable drop in PSNR value. Based on the experimental results, (RCAB (20) in Residual Group (1) is chosen as the final design.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Table 2 THE EXTENSION OF RESIDUAL IN BASED LINE MODEL

Method	Params	PSNR (dB)	Best Epoch	All Epoch
RCAN (Re-Train)	15,444,667	38.25	524	600
2 Multi Residual Network of branch RCAB (20) in Residual Group (1)	1,713,835	38.06	556	600
RCAB (20) in 2 Residual Network of Residual branch Group (1)	1,713,835	38.14	576	600
2 Multi Residual Network of branch RCAB (10) in Residual Group (2)	1,750,763	38.07	543	600
RCAB (20) in 2 Residual Network of Residual branch Group (1) - Upscale	1,861,547	38.16	417	600

shows the simulation results of several configurations of RCAB and RG with 600 epochs. It is noticed that both the PSNR value and some parameters of (2 Multi Residual Network of branch RCAB (20) in Residual Group (1)) is almost the same as that of (RCAB (20) in 2 Residual Network of Residual branch Group (1)). Another configuration is adding an upscale layer to the network in a multi-path fashion. The edge of the image can be seen more clearly by using this configuration. Moreover, this configuration can achieve a stable model of about 100 epochs faster than other configurations. Hence, this configuration is chosen to be used in the proposed architecture even though the number of parameters and PSNR values are not the best.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.2. Dataset Image

3.2.1. Training Dataset image

3.2.1.1. DIV2K [69]

An image database was used in the NTIRE (New Trends in Image Restoration and Enhancement workshop) competition. The images are at 2K resolution, divided into a database for training 800 models and reviewing and testing 100 models.

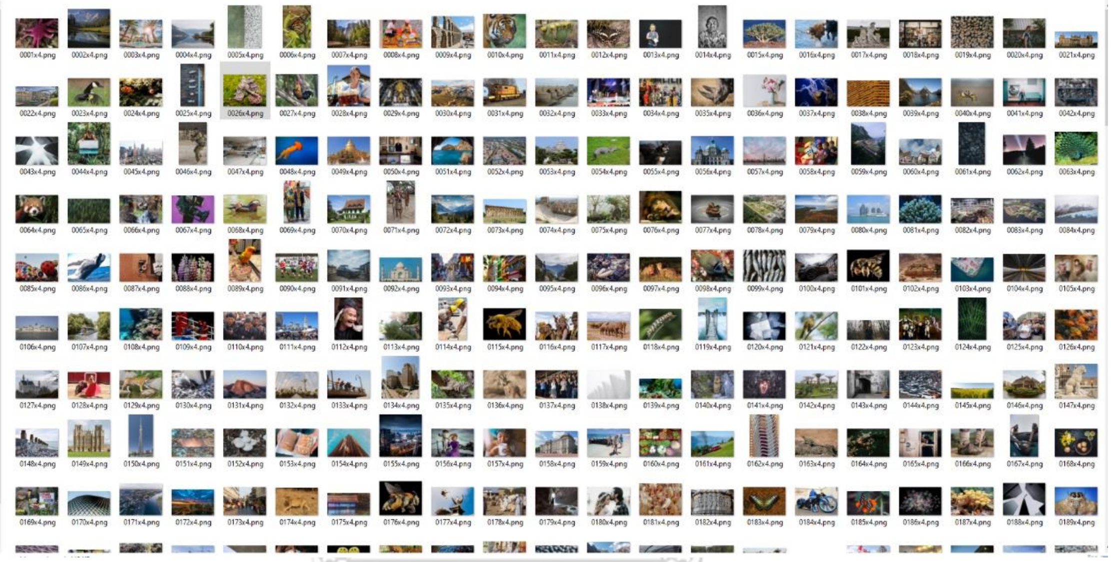


Figure 39 An example of an image in DIV2K image

3.2.2. Testing Dataset Image

3.2.2.1. Set5 [70]

A primary image database is used in research in the field of ultra-high resolution image recovery, which consisted of 5 test images, namely a child's face, bird, butterfly, head, and woman, at a high-resolution size between 250x361 pixels and 720x576.

Figure 40 An example of an image in Set5

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

3.2.2.2. Set14 [71]

There will be more images than in the Set5 image database. However, the image resolution will be low, between 250x361 pixels and 720x576 pixels.

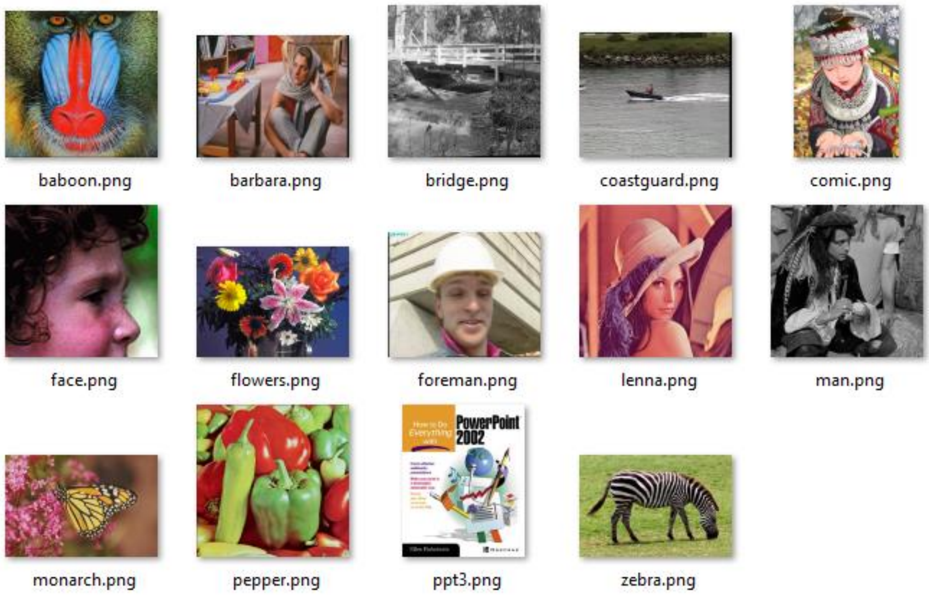


Figure 41 An example of an image in Set14

3.2.2.3. BSD100 [72]

A primary image database contains 100 test images with resolutions ranging from 321x421 pixels to 481x321 pixels, spanning various images from nature to objects such as airplanes, people, food, and others.

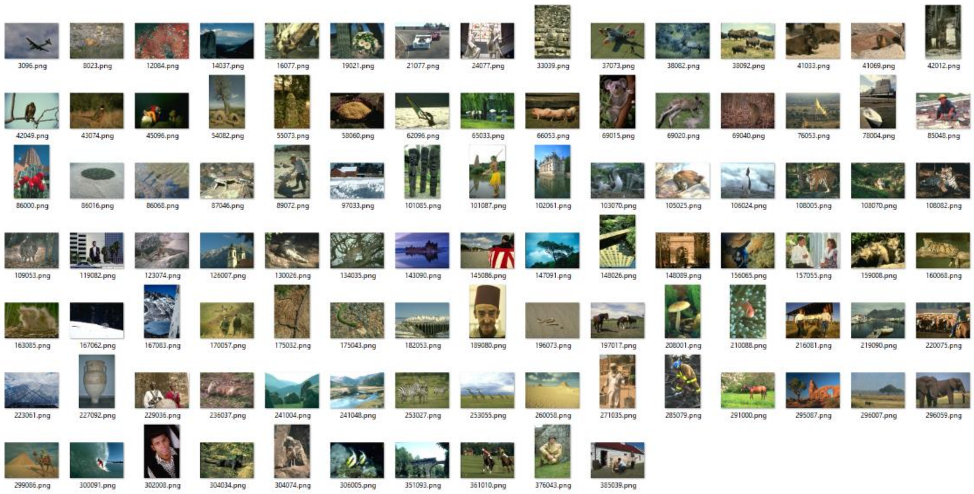


Figure 42 An example of an image in BSD100

3.2.2.4. Urban100 [19]

It has the same images as BSD100 but focuses on photographing buildings. with resolutions ranging from 567x1024 pixel to 1280x963 pixel

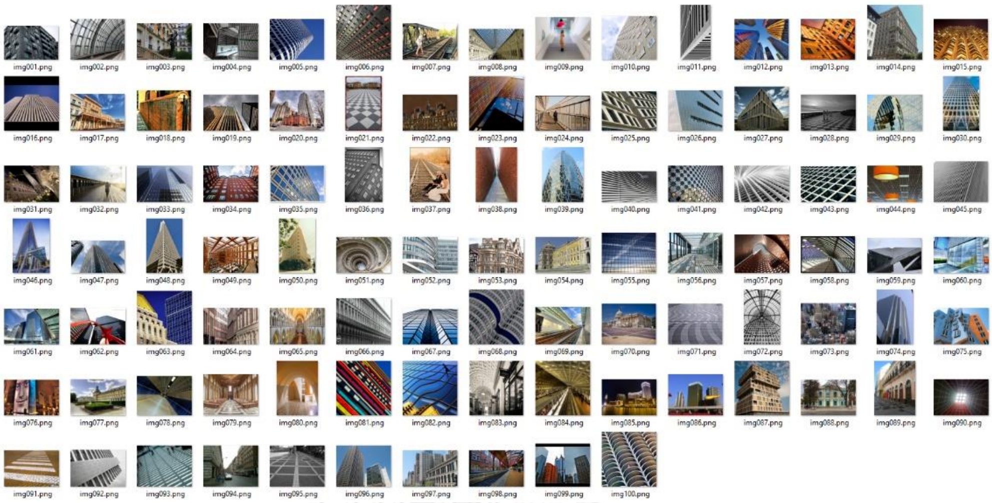


Figure 43 An example of an image in Urban100

3.2.2.5. Manga109 [73]

It is the latest database used to test ultra-high resolution image recovery algorithms. The image data set consisted of 109 images for testing from various comic books Made by Japanese artists between 1970 and 2010 with a resolution of 742x1072 pixels to 827x1170 pixels.

จุฬาลงกรณ์มหาวิทยาลัย

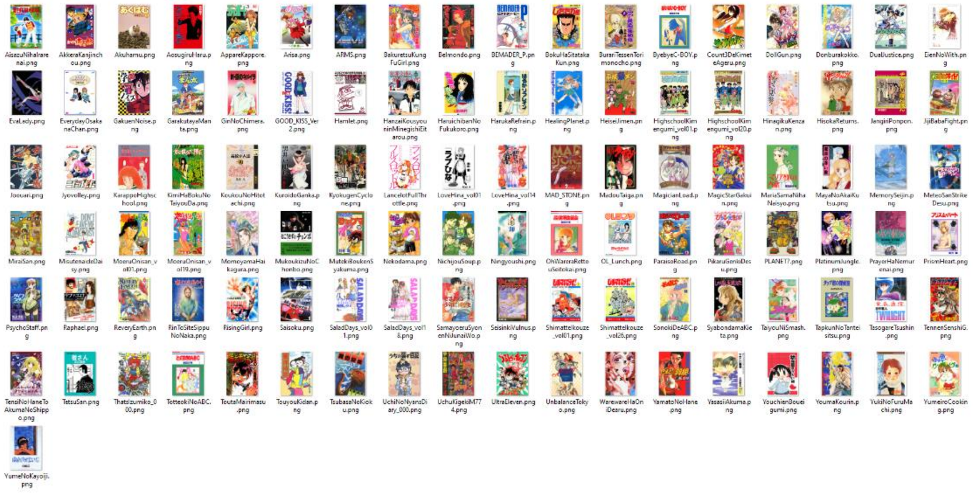
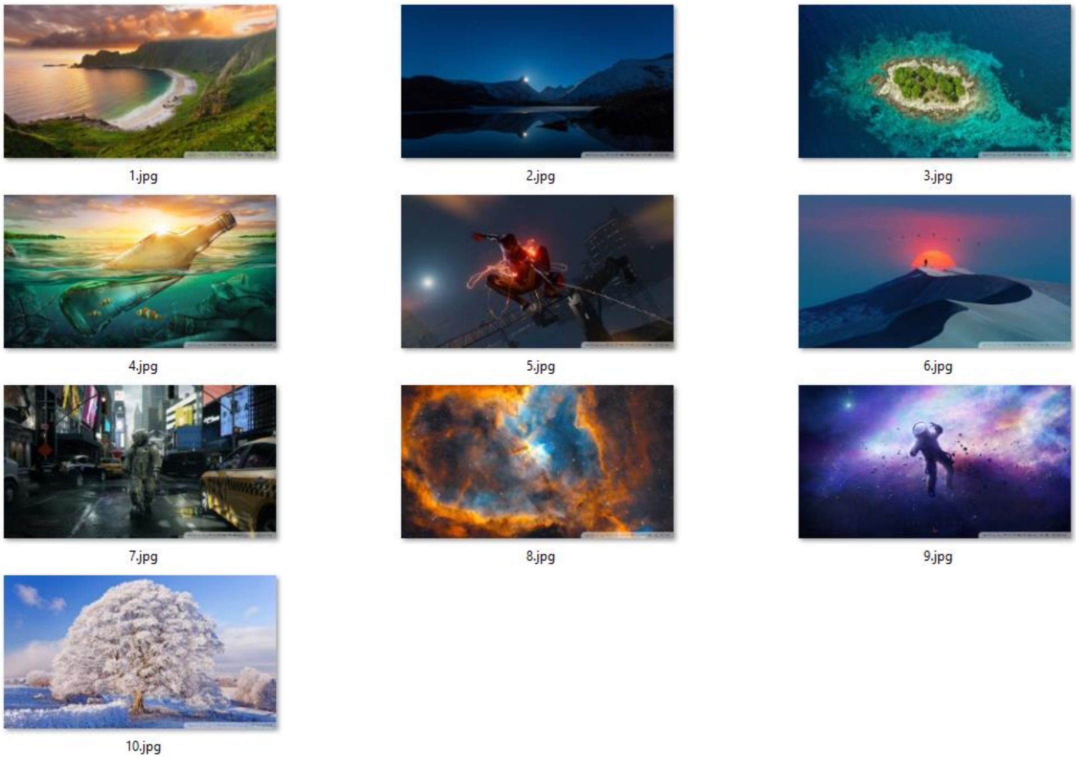


Figure 44 An example of an image in Manga109

3.2.3. Testing Dataset Image of HW

The Internet source dataset demonstrates the proposed algorithm's real-time implementation. The selected image dataset includes ten wallpaper images with a resolution of 640x360, which are reasonably selected for testing in the proposed hardware.



จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Figure 45 Example of an image in Testing Dataset Image of HW

3.3. Hardware

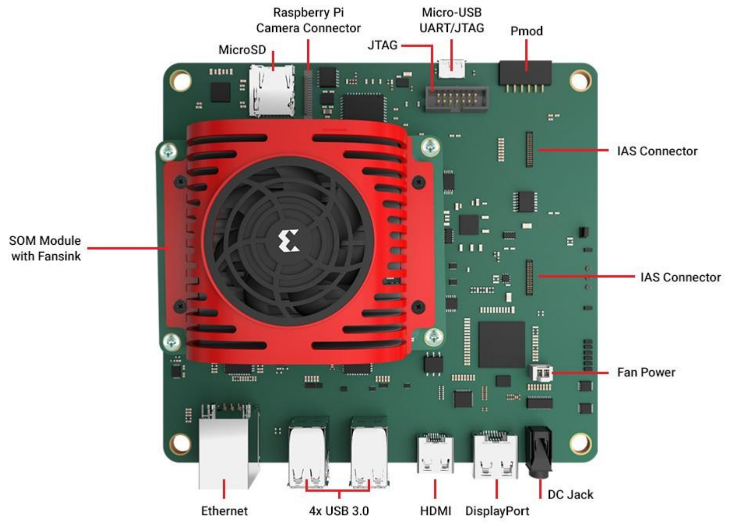
3.3.1. System on Chip (SoC) FPGA

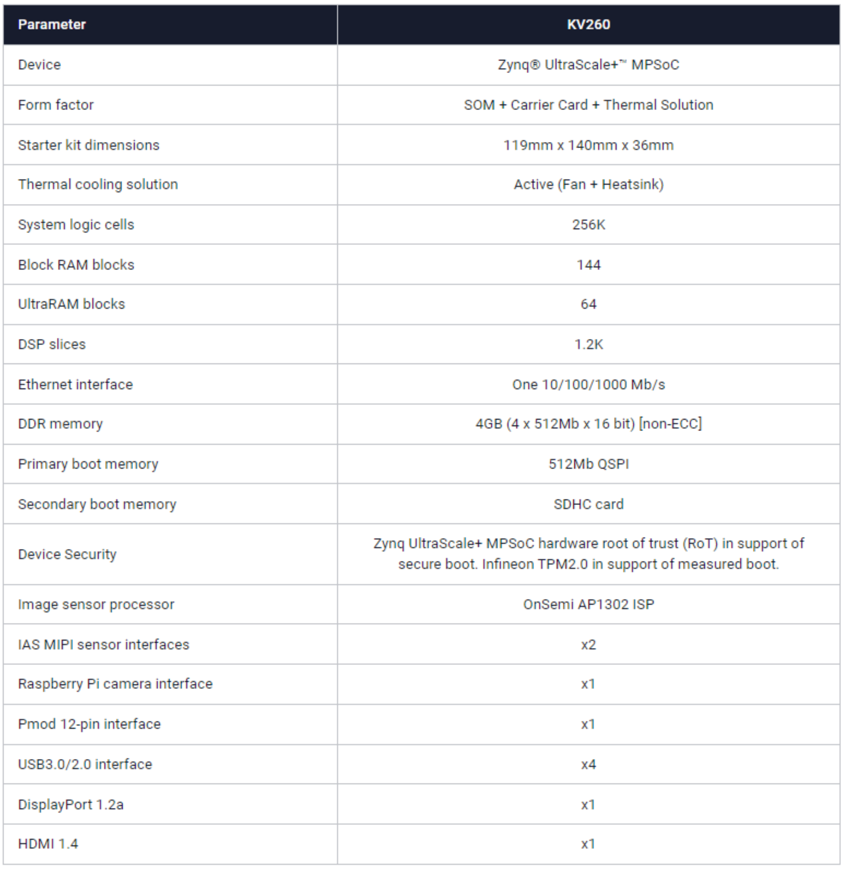
The central processing unit (CPU) is typically designed from fix gate and core. The processing is based on task and fixed arithmetic engines. Also, the input and output port is fixed. The integrated circuit is also known as System on Chip (SOC) [74]. The circuit consists of memory, CPU, and input/output port with low power consumption. The objective of combining SoC with FPGAS is to use some library or processing from CPU with Pararell processing of FPGA for fast processing. FPGA Field Programmable

Gate Array is programmable logic interconnected digital circuits. They are used because of their higher efficiency and multitasking capacity, which is even better than many microcontrollers and GPUs. FPGAs are easier to manufacture and are more adaptable than Application Specific ICs (ASIC). Mostly used CNN or DNN-based algorithms require efficient hardware architectures. The main drawback in GPUS hardware is high power consumption, and for ASIC, it is less adaptable. FPGAs have been employed to address these specific issues, capable of processing billions of 32-bit Floating Point logical and arithmetic operations with fast speed and better accuracy. To increase the processing of convolution operations. Block Random Access Memories are used. The following hardware and framework requirements have been stated for accelerating the hardware solution of ML/DL algorithms and showing the platform's flexibility in incorporating FPGA on adaptive SoC with ARM-based embedded CPU.

We use SoC, which is the combination of CPU and FPGA. For edge-based AI applications, the framework is targeted on Xilinx Kria SoM, which embeds extra port connectivity giving high computation performance based on ZYNQ Ultrascale and MPSoC. For workflow interface, deploying a pre-trained model to Kria SoM is like using a GPU with CUDA and DNN Kit. FPGA was used for workflow demonstration with performance collecting metrics amended by Kria SoM. This demonstration showed good computation evaluation in a low power budget. It is a qualitative analysis to check the abilities of an adaptable platform.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University





Parameter	KV260
Device	Zynq® UltraScale+™ MPSoC
Form factor	SOM + Carrier Card + Thermal Solution
Starter kit dimensions	119mm x 140mm x 36mm
Thermal cooling solution	Active (Fan + Heatsink)
System logic cells	256K
Block RAM blocks	144
UltraRAM blocks	64
DSP slices	1.2K
Ethernet interface	One 10/100/1000 Mb/s
DDR memory	4GB (4 x 512Mb x 16 bit) [non-ECC]
Primary boot memory	512Mb QSPI
Secondary boot memory	SDHC card
Device Security	Zynq UltraScale+ MPSoC hardware root of trust (RoT) in support of secure boot. Infineon TPM2.0 in support of measured boot.
Image sensor processor	OnSemi AP1302 ISP
IAS MIPI sensor interfaces	x2
Raspberry Pi camera interface	x1
Pmod 12-pin interface	x1
USB3.0/2.0 interface	x4
DisplayPort 1.2a	x1
HDMI 1.4	x1

Figure 46 Kria KV260 specifications

3.3.2. Hardware Design Flow

We used Vitis AI to support the latest framework, TensorFlow, and PyTorch for AI. It is an open-source platform that compiles the flow of the model from cloud to edge. Using a Vitis-based AI environment accelerates the interpretation of AI on the Xilinx hardware platform working for both cloud and edge. It is very efficient and user-friendly. Using this platform can develop many deep learning algorithms. Vitis Artificial Intelligence design is implemented with the below-mentioned stature:

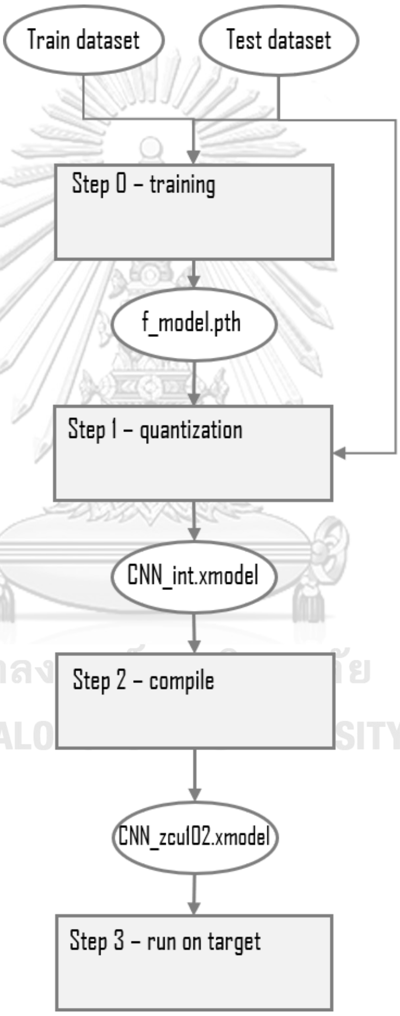


Figure 47 Hardware Design Flow

The proposed SR model is initially trained with the GPU's training image dataset. To implement our trained model in hardware (DPU), we quantize it. We convert the floating-point checkpoint into a fixed-point integer checkpoint in the quantization procedure. Once the quantization procedure is done, the quantized model can be found in the DPU module. Then the compile operation is applied to the quantized model. The compile.sh shell script is used to compile our quantized model. It creates a .xmodel file that contains the instructions and data. These instructions and data make our model ready to execute on the DPU.

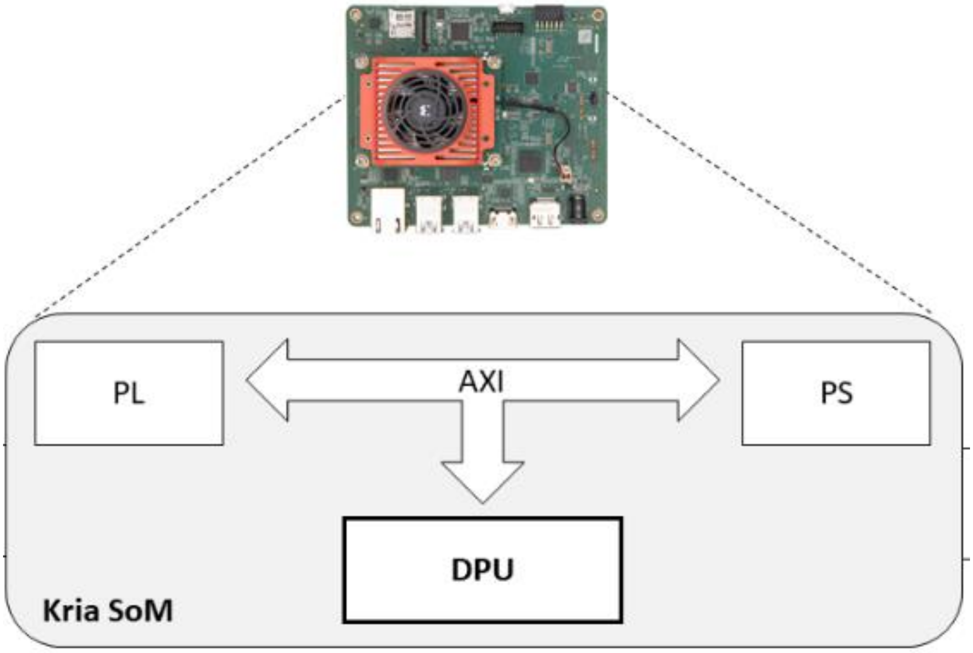


Figure 48 Kria KV260 Model

For further compaction of model design and higher performance, quantization is used. A quantized model utilizes 8-bit integer tensors instead of 32-bit floating operations. 8-bit Integer quantization is advantageous over 32-bit FLOPs because it reduces the model size and memory bandwidth requirement by up to 4 times. It also provides 2-4 times faster calculation speed than FP32 tensors. The INT8 quantized tensors are particularly used to accelerate the inference and forward calculative operations. The specialty of PyTorch is that it uses FP32 for training and then converts to INT8. All the calculations (forward or reverse) are performed through a fake quantization module using floating-point numbers, and then the trained FP trained model is converted to lower accuracy.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

At this level, the low accuracy quantized tensor directly builds the model. This raises a question about what a quantized tensor is and how to generate a quantized tensor in PyTorch. A quantized tensor is a tensor that can store quantized data (i.e., INT8/UNIT8/INT32 data) and quantization parameters (i.e., scaling zeros and quantized zeros). The first thing to quantify a tensor in PyTorch is to represent the quantized data as a tensor. This generated quantized tensor can be serially and parallelly connected to perform multiple operations. Two types of processes are used to quantize the tensor, i.e., per-tensor asymmetric linear quantization and per-channel asymmetric linear quantization. In per-tensor asymmetric linear quantization, all the tensors are scaled similarly. In contrast, in per-channel asymmetric linear quantization, each slice of channel dimension (for a particular dimension) of the tensor uses different scaling and offset to reduce the error in the quantization process. This process of quantization is represented through a mathematical mapping equation stated as:

Quantization aware training: This type of quantization is used for acquiring more accuracy where the computations are performed with FP32 and then converted to INT8 quantization. Thus, the above processes compact the model design, achieve higher performance and make the model suitable for mobile edge devices.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

3.4. Performance Evaluate

3.4.1. Peak Signal to Noise Ratio: PSNR

We use the measurement Root Mean Square Error: RMSE: A small value indicates high accuracy.

$$RMSE = \sqrt{\sum \frac{(I-T)^2}{N}} \tag{15}$$

By I is the original image.

T is an image obtained from various methods.

N is the image size (width x length).

The measure of the ratio of the signal to noise (Peak Signal to Noise Ratio: PSNR), which is higher, means that the picture has less noise. It can be calculated as follows:

$$PSNR = 10\, log_{10} \left[\frac{(2^{n}-1)^2}{RMSE}\right] \tag{16}$$

where n is the number of bits used to represent the color value in each image point.

3.4.2. Structural similarity (SSIM)

จุฬาลงกรณ์มหาวิทยาลัย

There are three terms to define the SSIM index: luminance, contrast, and structural need to do a multiplicative combination.

$$SSIM(x,y) = [l(x,y)]^{\alpha} \cdot [c(x,y)]^{\beta} . [s(x,y)]^{\gamma} \tag{17}$$

Where,

$$\begin{gathered} l(x,y) = \frac{2\mu_x\mu_y + C_1}{\mu_x^2 + \mu_y^2 + C_1} \\ c(x,y) = \frac{2\sigma_x\sigma_y + C_2}{\sigma_x^2 + \sigma_y^2 + C_2} \\ s(x,y) = \frac{\sigma_{xy} + C_3}{\sigma_x\sigma_y + C_3} \end{gathered} \tag{18}$$

Where $\mu_x, \mu_y, \sigma_x, \sigma_{y,} \sigma_{xy}$ are the local means, standard deviations, and cross-covariance for images *x, y*. If $\alpha = \beta = \gamma = 1$, and $C_3 = \frac{C_2}{2}$

$$SSIM(x, y) = \frac{(2\mu_x\mu_y + C_1)(2\sigma_{xy} + C_2)}{(\mu_x^2 + \mu_y^2 + C_1)(\sigma_x^2 + \sigma_y^2 + C_2)} \tag{19}$$

3.4.3. Frame Per Second (FPS)

A simple and effective performance metric of a deep learning-based SR algorithm is to measure the number of image frames per second, called frame per second (FPS), as Equation (20) is depicted below.

$$FPS = \frac{Totlal\ Image}{Time_{Start} - Time_{End}} \tag{21}$$

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter 4
Experimental Results and Discussions

4.1. Experimental Setup

For the training set of our experiments, we utilize the DIV2K [69] dataset, including 800 images. For the testing set, five standard benchmark datasets such as Set5 [69], Set14 [71], B100 [19], Urban100, and Manga109 are reasonably considered. For degradation models, we apply Bicubic (BI) in our experiments. Y- PSNR, and Y-SSIM are efficiently presented for the performance metrics to evaluate the SR results. For data augmentation, the 800 training images are randomly applied with three rotations such as 90, 180, 120 degrees, and horizontal flipping. We extract 16 in 48×48 test LR color patches to get the input for each training batch. We train our model with an ADAM optimizer, and the parameter values of β_1, β_2, and ϵ are 0.9, 0.999, and 10^{-8}, respectively. Initially, we set the learning rate as 10^{-4}. Then, for every 2×10^{-5} iterations, we reduce the learning rate to half. Our efficient models are implemented by effectively utilizing PyTorch on Titan Xp GPU.

4.2. Result Analysis of Network

To analyze the performance of our model, several state-of-the-art approaches are chosen. Table 5 describes the parameters for each SR model and Y-PSNR and Y-SSIM values for five datasets with upscale factors 2x, 3x, and 4x, respectively. According to Table 3, we can conclude that our model can achieve higher performance with a reasonable number of parameters than others. For Set 5, Set 14, B100, Urban100, and Mango109, our model improves 0.05 dB, 0.07 dB, 0.02 dB, 0.16 dB, and 0.16 dB over the second-best PSNR model called AWSRN, A2F-L, AWSRN, AWSRN, and A2F-L, respectively, under a comparable number of parameters. As we can see from Tables 5 and 6, our model can achieve a big PSNR improvement of 0.3

dB and 0.26 dB for upscale factors 3x and 4x over AWSRN and A2F-L on the Manga109 data set, respectively.

Table 3 RESULTS QUANTITATIVE OF MODEL X2

Model	Param	Set5	Set14	B100	Urban100	Manga109
FSRCNN [22]	12K	37.00/0.9558	32.63/0.9088	31.53/0.8920	29.88/0.9020	36.67/0.9694
SRCNN [21]	57K	36.66/0.9542	32.42/0.9063	31.36/0.8879	29.50/0.8946	35.74/0.9661
DRRN [40]	297K	37.74/0.9591	33.23/0.9136	32.05/0.8973	31.23/0.9188	37.92/0.9760
A2F-SD [37]	313K	37.91/0.9602	33.45/0.9164	32.08/0.8986	31.79/0.9246	38.52/0.9767
A2F-S [37]	320k	37.79/0.9597	33.32/0.9152	31.99/0.8972	31.44/0.9211	38.11/0.9757
FALSR-B [39]	326K	37.61/0.9585	33.29/0.9143	31.97/0.8967	31.28/0.9191	38.20/0.9762
AWSRN-SD [35]	348K	37.86/0.9600	33.41/0.9161	32.07/0.8984	31.67/0.9237	-
AWSRN-S [35]	397K	37.75/0.9596	33.31/0.9151	32.00/0.8974	31.39/0.9207	37.90/0.9755
FALSR-C [39]	408K	37.66/0.9586	33.26/0.9140	31.96/0.8965	31.24/0.9187	-
CARN-M [32]	412K	37.53/0.9583	33.26/0.9141	31.92/0.8960	31.23/0.9193	-
SRFBN-S [34]	483K	37.78/0.9597	33.35/0.9156	32.00/0.8970	31.41/0.9207	38.06/0.9757
IDN [41]	552K	37.83/0.9600	33.30/0.9148	32.08/0.8985	31.27/0.9196	-
VDSR [7]	665K	37.53/0.9587	33.03/0.9124	31.90/0.8960	30.76/0.9140	37.22/0.9729
MemNet [23]	677K	37.78/0.9597	33.28/0.9142	32.08/0.8978	31.31/0.9195	37.27/0.9740
IMDN [64]	694K	38.00/0.9605	33.63/0.9177	32.19/0.8996	32.17/0.9283	38.88/0.9774
LapSRN [6]	813K	37.52/0.9590	33.08/0.9130	31.80/0.8950	30.41/0.9100	-
SelNet [42]	974K	37.89/0.9598	33.61/0.9160	32.08/0.8984	-	-
A2F-M [37]	999K	38.04/0.9607	33.67/0.9184	32.18/0.8996	32.27/0.9294	38.87/0.9774
FALSR-A [39]	1021K	37.82/0.9595	33.55/0.9168	32.12/0.8987	31.93/0.9256	-
MoreMNAS-A [38]	1039K	37.63/0.9584	33.23/0.9138	31.95/0.8961	31.24/0.9187	-
AWSRN-M [35]	1063K	38.04/0.9605	33.66/0.9181	32.21/0.9000	32.23/0.9294	38.66/0.9772
A2F-L [37]	1363K	38.09/0.9607	33.78/0.9192	32.23/0.9002	32.46/0.9313	38.95/0.9772
AWSRN [35]	1397K	38.11/0.9608	33.78/0.9189	32.26/0.9006	32.49/0.9316	38.87/0.9776
SRMDNF [24]	1513K	37.79/0.9600	33.32/0.9150	32.05/0.8980	31.33/0.9200	-
CARN [32]	1592K	37.76/0.9590	33.52/0.9166	32.09/0.8978	31.92/0.9256	-
DRCN [8]	1774K	37.63/0.9588	33.04/0.9118	31.85/0.8942	30.75/0.9133	37.63/0.9723
MFCC (Ours)	1861K	38.16/0.9611	33.85/0.9195	32.28/0.9010	32.65/0.9331	39.11/0.9780
MSRN [33]	5930K	38.08/0.9607	33.70/0.9186	32.23/0.9002	32.29/0.9303	38.69/0.9772

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Table 4 RESULTS QUANTITATIVE OF MODEL X3

Model	Param	Set5	Set14	B100	Urban100	Manga109
FSRCNN [22]	12K	33.16/0.9140	29.43/0.8242	28.53/0.7910	26.43/0.8080	30.98/0.9212
SRCNN [21]	57K	32.75/0.9090	29.28/0.8209	28.41/0.7863	26.24/0.7989	30.59/0.9107
DRRN [40]	297K	34.03/0.9244	29.96/0.8349	28.95/0.8004	27.53/0.8378	32.74/0.9390
A2F-SD [37]	316K	34.23/0.9259	30.22/0.8395	29.01/0.8028	27.91/0.8465	33.29/0.9424
A2F-S [37]	324K	34.06/0.9241	30.08/0.8370	28.92/0.8006	27.57/0.8392	32.86/0.9394
AWSRN-SD [35]	388K	34.18/0.9273	30.21/0.8398	28.99/0.8027	27.80/0.8444	33.13/0.9416
CARN-M [32]	412K	33.99/0.9236	30.08/0.8367	28.91/0.8000	27.55/0.8385	-
AWSRN-S [35]	477K	34.02/0.9240	30.09/0.8376	28.92/0.8009	27.57/0.8391	32.82/0.9393
SRFBN-S [34]	483K	34.20/0.9255	30.10/0.8372	28.96/0.8010	27.66/0.8415	33.02/0.9404
IDN [41]	552K	34.11/0.9253	29.99/0.8354	28.95/0.8013	27.42/0.8359	-
VDSR [7]	665K	33.66/0.9213	29.77/0.8314	28.82/0.7976	27.14/0.8279	32.01/0.9310
MemNet [23]	677K	34.09/0.9248	30.00/0.8350	28.96/0.8001	27.56/0.8376	-
IMDN [64]	703K	34.36/0.9270	30.32/0.8417	29.09/0.8046	28.17/0.8519	33.61/0.9445
A2F-M [37]	1003K	34.50/0.9278	30.39/0.8427	29.11/0.8054	28.28/0.8546	33.66/0.9453
AWSRN-M [35]	1143K	34.42/0.9275	30.32/0.8419	29.13/0.8059	28.26/0.8545	33.64/0.9450
SelNet [42]	1159K	34.27/0.9257	30.30/0.8399	28.97/0.8025	-	-
A2F-L [37]	1367K	34.54/0.9283	30.41/0.8436	29.14/0.8062	28.40/0.8574	33.83/0.9463
AWSRN [35]	1476K	34.52/0.9281	30.38/0.8426	29.16/0.8069	28.42/0.8580	33.85/0.9463
SRMDNF [24]	1530K	34.12/0.9250	30.04/0.8370	28.97/0.8030	27.57/0.8400	-
CARN [32]	1592K	34.29/0.9255	30.29/0.8407	29.06/0.8034	28.06/0.8493	-
DRCN [8]	1774K	33.82/0.9226	29.76/0.8311	28.80/0.7963	27.15/0.8276	32.31/0.9328
MFCC (Ours)	1862K	34.67/0.9940	30.51/0.8456	29.22/0.8080	28.64/0.8616	34.15/0.9478
MSRN [33]	6114K	34.46/0.9278	30.41/0.8437	29.15/0.8064	28.33/0.8561	33.67/0.9456

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Table 5 RESULTS QUANTITATIVE OF MODEL X4

Model	Param	Set5	Set14	B100	Urban100	Manga109
FSRCNN [22]	12K	30.71/0.8657	27.59/0.7535	26.98/0.7150	24.62/0.7280	27.90/0.8517
SRCNN [21]	57K	30.48/0.8628	27.49/0.7503	26.90/0.7101	24.52/0.7221	27.66/0.8505
DRRN [40]	297K	31.68/0.8888	28.21/0.7720	27.38/0.7284	25.44/0.7638	29.46/0.8960
A2F-SD [37]	320K	32.06/0.8928	28.47/0.7790	27.48/0.7373	25.80/0.7767	30.16/0.9038
A2F-S [37]	331K	31.87/0.8900	28.36/0.7760	27.41/0.7305	25.58/0.7685	29.77/0.8987
AWSRN-SD [35]	412K	31.92/0.8903	28.42/0.7762	27.44/0.7304	25.62/0.7694	-
FSRCNN [22]	444K	31.98/0.8921	28.46/0.7786	27.48/0.7368	25.74/0.7746	30.09/0.9024
SRFBN-S [34]	483K	31.98/0.8923	28.45/0.7779	27.44/0.7313	25.71/0.7719	29.91/0.9008
IDN [41]	552K	31.82/0.8903	28.25/0.7730	27.41/0.7297	25.41/0.7632	-
AWSRN-S [35]	588K	31.77/0.8893	28.35/0.7761	27.41/0.7304	25.56/0.7678	29.74/0.8982
VDSR [7]	665K	31.35/0.8838	28.01/0.7674	27.29/0.7251	25.18/0.7524	28.83/0.8809
MemNet [23]	677K	31.74/0.8893	28.26/0.7723	27.40/0.7281	25.50/0.7630	-
IMDN [64]	715K	32.21/0.8948	28.58/0.7811	27.56/0.7353	26.04/0.7838	30.45/0.9075
LapSRN [6]	813K	31.54/0.8850	28.19/0.7720	27.32/0.7280	25.21/0.7560	29.09/0.8845
A2F-M [37]	1010K	32.28/0.8955	28.62/0.7828	27.58/0.7364	26.17/0.7892	30.57/0.9100
AWSRN-M [35]	1254K	32.21/0.8954	28.65/0.7832	27.60/0.7368	26.15/0.7884	30.56/0.9093
A2F-L [37]	1374K	32.32/0.8964	28.67/0.7839	27.62/0.7379	26.32/0.7931	30.72/0.9115
SelNet [42]	1417K	32.00/0.8931	28.49/0.7783	27.44/0.7325	-	-
SRMDNF [24]	1555K	31.96/0.8930	28.35/0.7770	27.49/0.7340	25.68/0.7730	-
AWSRN [35]	1587K	32.27/0.8960	28.69/0.7843	27.64/0.7385	26.29/0.7930	30.72/0.9109
CARN [32]	1592K	32.13/0.8937	28.60/0.7806	27.58/0.7349	26.07/0.7837	-
DRCN [8]	1774K	31.53/0.8854	28.02/0.7670	27.23/0.7233	25.14/0.7510	28.98/0.8816
MFCC (Ours)	2157K	32.42/0.8973	28.73/0.7849	27.670.7399	26.48/0.7977	30.98/0.9131
SRResNet [29]	2015K	32.02/0.8934	28.35/0.7770	27.53/0.7337	26.05/0.7819	-
MSRN [33]	6078K	32.26/0.8960	28.63/0.7836	27.61/0.7380	26.22/0.7911	30.57/0.9103

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

For the most challenging upscale factor 8x, our model can continue as the best PSNR SR model, as shown in Table 6 and can improve a significant PSNR improvement of 0.91 dB with a comparable trade-off over the state-of-the-art model called EDSR.

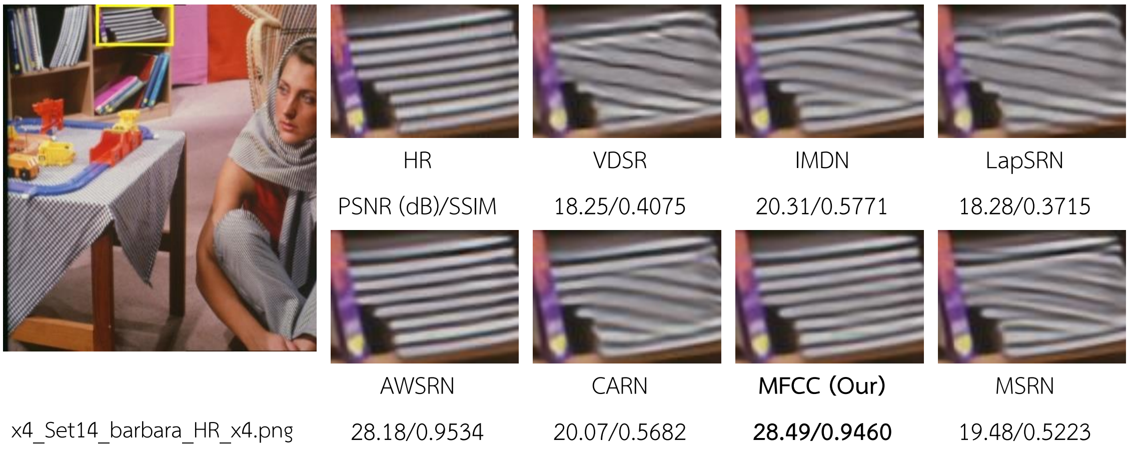
Table 6 RESULTS QUANTITATIVE OF MODEL X8

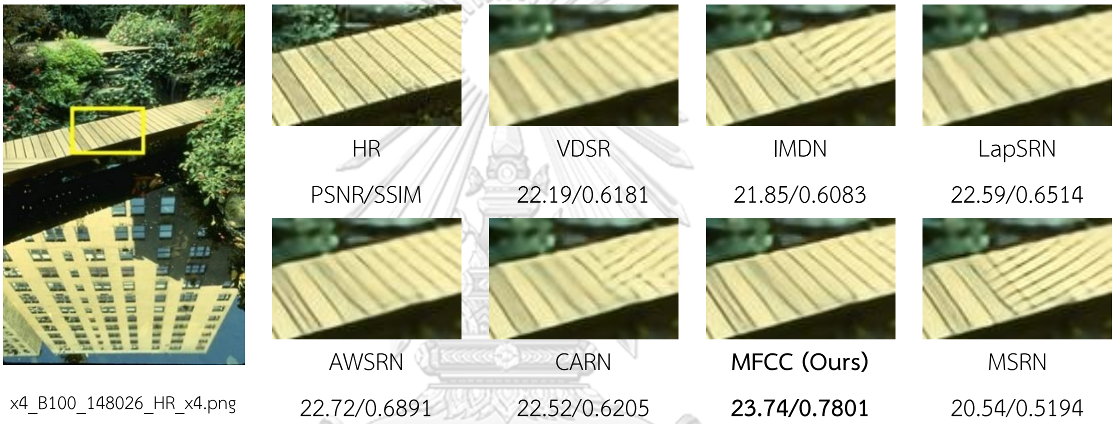
Model	Param	Set5	Set14	B100	Urban100	Manga109
FSRCNN [22]	12K	25.42/0.6440	25.34/0.5482	24.21/0.5112	21.32/0.5090	22.39/0.6357
SRCNN [21]	57K	25.34/0.6471	23.86/0.5443	24.14/0.5043	21.29/0.5133	22.46/0.6606
VDSR [7]	665K	25.73/0.6743	23.20/0.5110	24.34/0.5169	21.48/0.5229	22.73/0.6688
LapSRN [6]	813K	26.15/0.7028	24.45/0.5792	24.54/0.5293	21.81/0.5555	23.39/0.7068
DRCN [8]	1,774K	25.93/0.6740	24.25/0.5510	24.49/0.5170	21.71/0.5230	23.20/0.6690
AWSRN [35]	2,348K	26.97/0.7747	24.99/0.6414	24.80/0.5967	22.45/0.6174	24.60/0.7782
MFCC (Ours)	2,453K	27.07/0.7762	25.01/0.6412	24.84/0.5980	22.54/0.6196	24.63/0.7791
MSRN [33]	6,226K	26.59/0.7254	24.88/0.5961	24.70/0.5410	22.37/0.5977	24.28/0.7517

In addition, to the objective performance of our model, we reasonably describe the subjective quality of our model. As we can see in Figure 49, the visual quality of the SR image of our model is clear and sharp enough to follow the visual system of humans. As displayed in this figure, all models suffer over-smoothing degradation or cannot reconstruct the tiny lines. In contrast, our model outperforms compared to the other models. In addition, the highest PSNR and SSIM belong to our model.

In contrast to our result, other models produced blur results and showed weakness in reconstructing a sharp image. The PSNR and SSIM of our results are significantly better than other models. *Figure 51* demonstrates the visual comparisons belonging to the Urban100 dataset. According to this figure, our MFCC result successfully generates the SR image similar to the GT image. In contrast, the other models' results show weakness in reconstructing the lines identical to the reference image. The best PSNR and SSIM belong to our model.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University





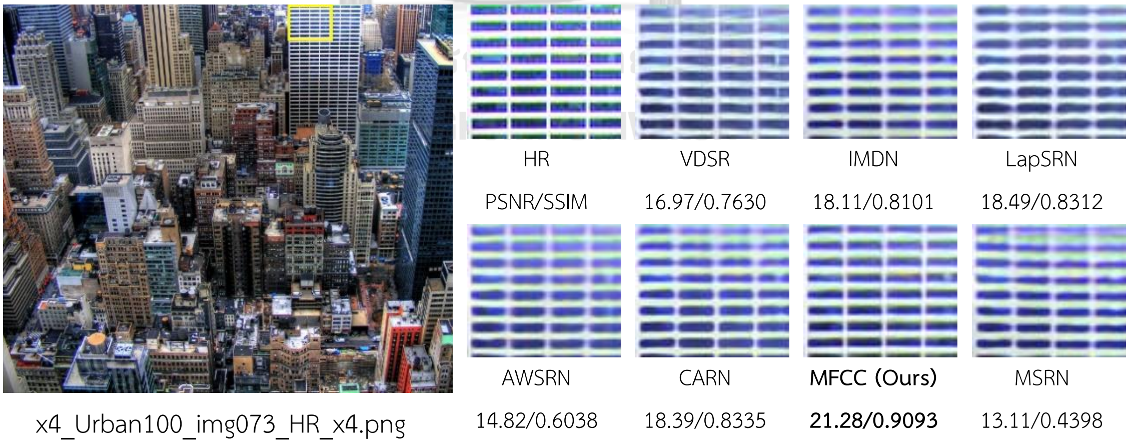


Figure 49 Qualitative comparison over datasets for scale x4. The red rectangle indicates the area of interest for zooming.

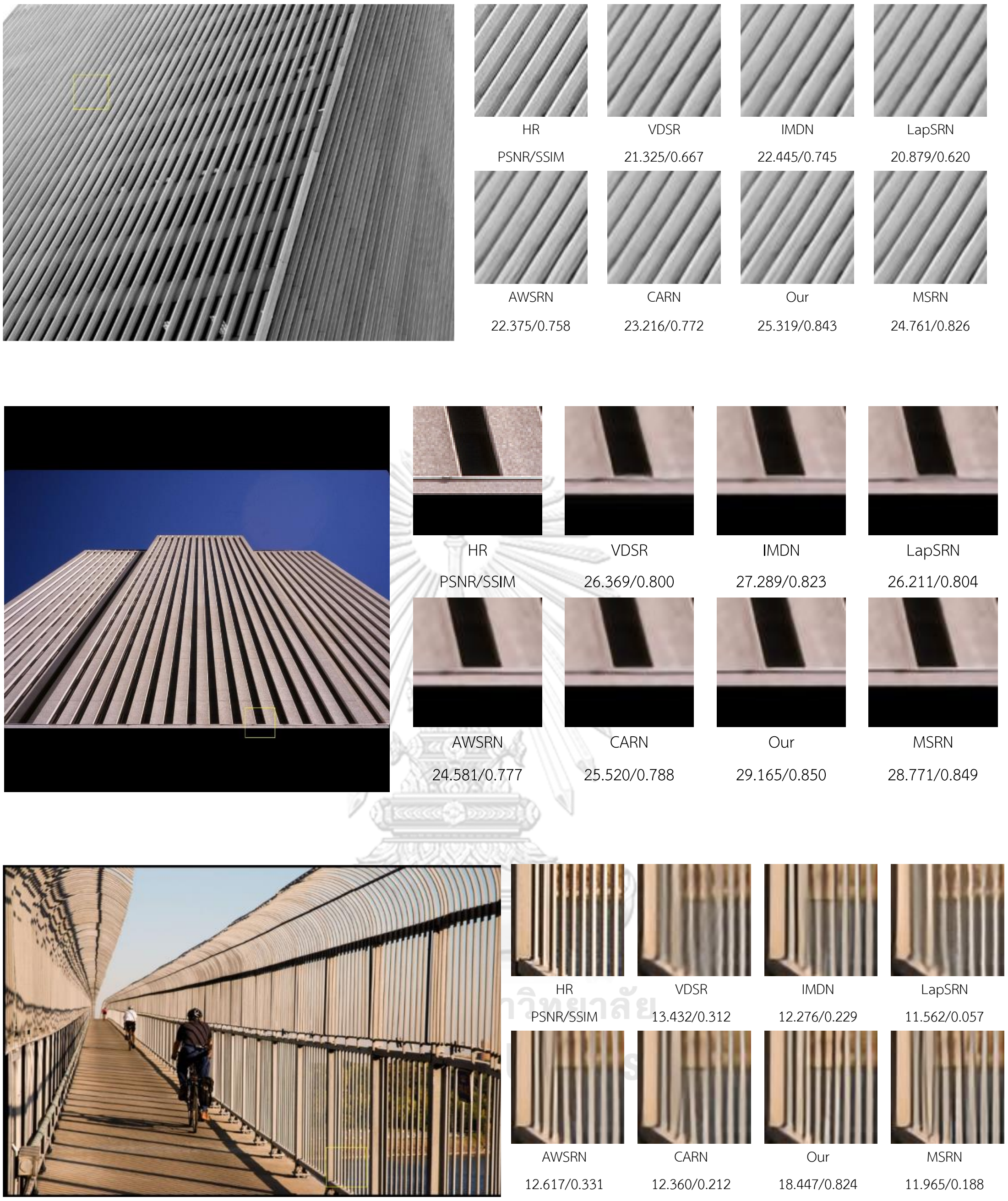
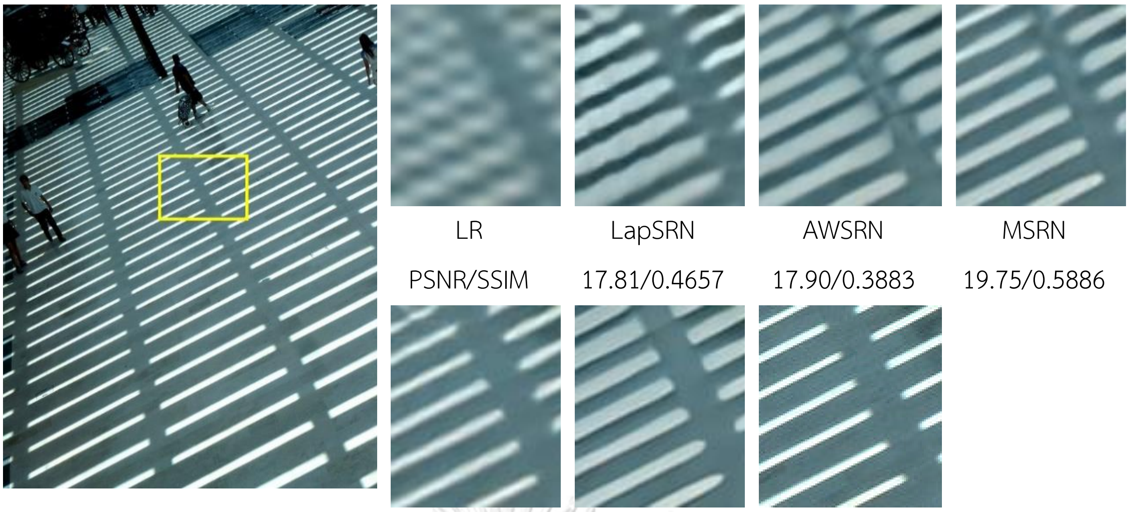


Figure 50 Qualitative comparison over datasets for scale ×4. The red rectangle indicates the area of interest for zooming.



LR PSNR/SSIM

LapSRN 17.81/0.4657

AWSRN 17.90/0.3883

MSRN 19.75/0.5886

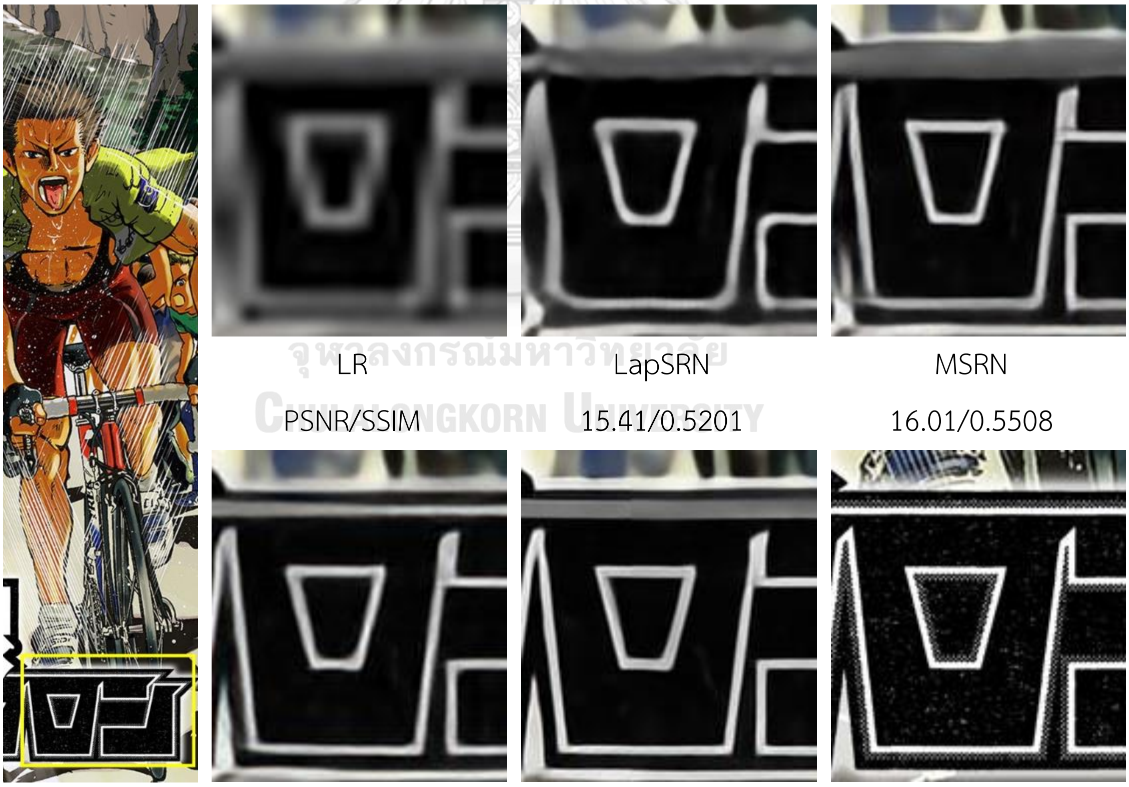
MFCC (Ours) **23.90/0.8060**

RCAN 19.13/0.5807

HR

Urban100_img093_HR_x8.png

Figure 51 Qualitative comparison of img093 belongs to Urban100 dataset for scale x8. The yellow rectangle indicates the area of interest for zooming



LR PSNR/SSIM

LapSRN 15.41/0.5201

MSRN 16.01/0.5508

MFCC (Ours) **18.09/0.6485**

RCAN 17.05/0.6203

HR

Manga109_KyokugenCyclone_HR_x8.png

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Figure 52 Qualitative comparison of an image belongs to Manga109 dataset for scale x8. The yellow rectangle indicates the area of interest for zooming.

Figure 52 compares the results of an image belonging to Manga109 datasets. The PSNR and SSIM of our proposed model are the highest. According to this figure, other models could not reconstruct the parallel lines located at the top of the selected patch and merge the lines. In contrast, our model shows a robust ability to produce tiny edges at this scale.

4.3. Result Analysis of Hardware

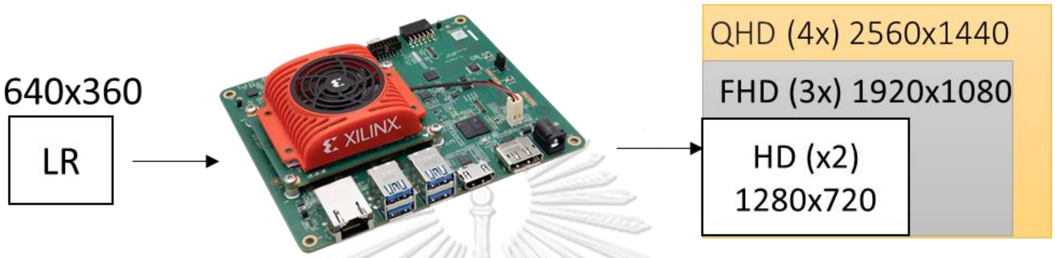


Figure 53 Target Resolution

Figure 53 shows three target resolution of the input LR image with 640x360 that the proposed hardware can support. They are x2, x3, and x4 scale factors with HD (1280x720), FHD (1920x1080), and QHD (2560x1440) FPS output size, respectively. The dataset for the proposed hardware can be found in *Figure 45*.

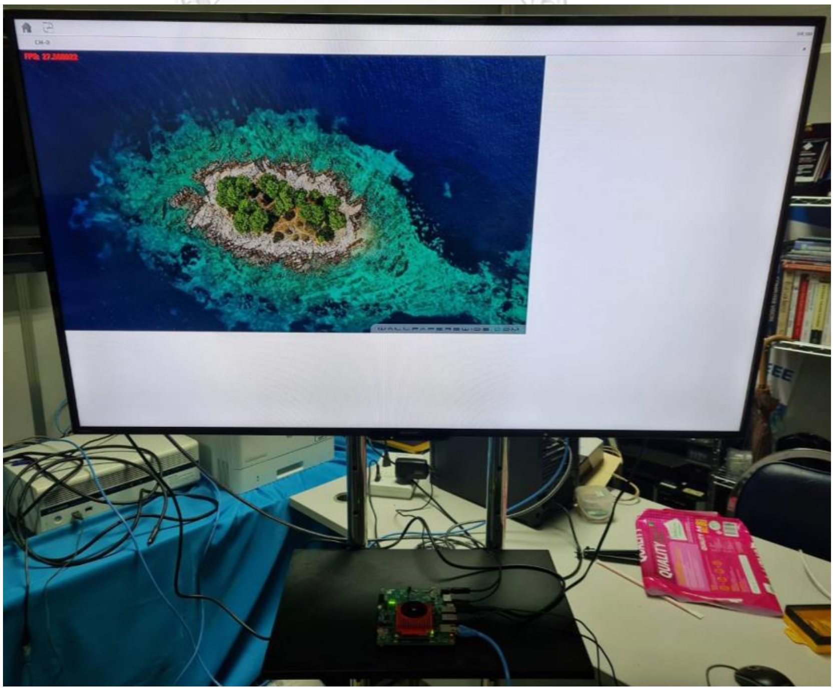


Figure 54 Result of KV260 processing super resolution method MFCC 2x

Figure 54 shows that the frame rate decreases from 45 fps to 30 fps due to the display process of the proposed hardware. Due to the utilizing computer vision tasks in a wide range of applications, single image super resolution is crucial in the computer vision task. Because of the constant increase in the network depth of super-resolution models recently, it has led to a massive amount of computation and memory utilization. On the other hand, the depth of these super-resolution models achieves a small increase in the results' performance. To design a more effective SR model. The experimental results on four benchmark datasets demonstrates outperformance of our model in terms of speed and accuracy, compared to the other state-of-the-art methods. Besides, this model effectively decreasing the number parameters and computation time. Moreover, the progressive reconstruction approach improves result at higher scale and simultaneously provides a variety of scales in a single model. This model was evaluated on the benchmark image datasets and achieved efficient performance over the other state-of-the art models.

Table 7 Hardware implementations average PSNR in with a scaling factor 2 of Set5

Scale (FPS Output Size)	x2 (1280x720)			x3 (1920x1080)			x4 (2560x1440)		
Methods	Param(k)	PSNR (dB)	FPS	Param(k)	PSNR (dB)	FPS	Param(k)	PSNR (dB)	FPS
RCAN (RG1, RCAB8) – Thread1	87	37.38	18	98	33.85	13	96	31.80	9
RCAN (RG1, RCAB8) – Thread2			19			14			10
MFCC (Our) – Thread1	20	37.20	39	32	33.77	21	29	31.55	11
MFCC (Our) – Thread2			45			24			13

- FPS Calculate from Input Image size is 640x360

Table 7 shows the number of parameters and average PSNR of the hardware implementation on scale 2, 3, and 8 of Set 5 for RCAN and MFCC in Tread1 and Tread2. The number of parameters for RCAN are 87, 98, and 96 while the number of parameters for MFCC are 20, 32, and 29 for x2, x3, and x4, respectively. Table 7 shows that MFCC consumes less parameters than RCAN about 77%, 67%, 70% for x2, x3, and x4, respectively under a comparable PSNR. In addition, MFCC can achieve a higher frame rate than RCAN in Tread1 and Trade 2 for all scale factors.

Table 8 provides the important comparison of hardware implementation of the proposed algorithm compared with start-of-the-art algorithms. According to Table 8, it can be found that the supporting scale factors and number of target resolution of the proposed algorithm are more than others. Consequently, our algorithm is more suitable for several scenarios under a minimum hardware resource.

Table 8 Hardware implementations comparison

<table>
<tr><th colspan="2">SR Methods</th><th>Sharp Filters Lagrange [44]</th><th>ANR [47]</th><th colspan="2">Edge Orientation Learn Linear Mappings [48]</th><th>FSRCNN [50]</th><th>FSRCNN-s [51]</th><th>FSRCNN-s [52]</th><th>Ours</th></tr>
<tr><td colspan="2">Implementation</td><td>0.13 um CMOS</td><td>90 mm CMOS</td><td>Xilinx XCKU040</td><td>0.13 um CMOS</td><td>65 nm CMOS</td><td>Xilinx XCKU040</td><td>Xilinx XCVU9P</td><td>Xilinx KV260</td></tr>
<tr><td colspan="2">HW Resources</td><td>5.1K</td><td>1.985K</td><td>Slice LUTS : 3,395
Slice Regs : 1,952
DSP Blocks : 108</td><td>159K</td><td>-</td><td>Slice LUTS : 151K
Slice Regs : 121K
DSP Blocks: 1,920</td><td>Slice LUTs : 94K
Slice Regs : 19K
DSP Blocks: 2,146</td><td>Slice LUTs : 51K
Slice Regs : 98K
DSP Blocks: 710</td></tr>
<tr><td colspan="2">Memory Size (KB)</td><td>-</td><td>235</td><td colspan="2">92</td><td>572</td><td>194</td><td>53</td><td>255</td></tr>
<tr><td colspan="2">Max. Frequency (MHz)</td><td>431</td><td>124.4</td><td>150</td><td>220</td><td>200</td><td>150</td><td>200</td><td>300</td></tr>
<tr><td colspan="2">Supported Scale</td><td>x2, x3</td><td>x2</td><td colspan="2">x2</td><td>x2, x4</td><td>x2</td><td>x2</td><td>x2, x3, x4</td></tr>
<tr><td rowspan="2">PSNR (dB)*</td><td>Baseline</td><td>-</td><td>34.00</td><td colspan="2">-</td><td>-</td><td>36.66</td><td>36.49</td><td>37.50</td></tr>
<tr><td>HW Results</td><td>-</td><td>33.83</td><td colspan="2">34.78</td><td>33.12</td><td>36.51</td><td>36.42</td><td>37.20</td></tr>
<tr><td colspan="2">Target Resolution</td><td>4K UHD (30 fps)</td><td>FHD (60 fps)</td><td colspan="2">4K UHD (60 fps)</td><td>FHD (25, 60 fps)</td><td>4K UHD (60 fps)</td><td>4K UHD (60 fps)</td><td>HD,FHD,QHD (45, 24, 13) fps</td></tr>
</table>

*Set 5 x2

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

Chapter 5

Conclusion and Future Works

This research proposed a lightweight single-image super-resolution model based on constructing Residual Group blocks on a multi-path residual architecture (MFCC). Utilizing a multi-path residual network increases the efficiency of the proposed lightweight model. Additionally, we addressed the lack of low-frequency details problem by employing the pixel shuffle fusion method. Based on this approach, the low-frequency details of the early layer are up-sampled and bypassed to the up-sampled features of the multi-path residual network. These layers' high and low-frequency information are fused and improve the proposed model's line and edge reconstructing capability. The experimental results on five benchmark datasets demonstrate our lightweight MFCC model outperforms other state-of-the-art models, especially on a scale of ×8. For future work, the proposed model will be tuned to achieve the optimum parameters and then implemented on the FPGAs to support real-world applications.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

REFERENCES

[1] M. Islam, V. Asari, M. Islam, and M. Karim, "Super-resolution enhancement technique for low resolution video," *IEEE Transactions on Consumer Electronics,* vol. 56, pp. 919-924, 2010.

[2] S. C. Park, M. K. Park, and M. G. Kang, "Super-resolution image reconstruction: A technical overview," *IEEE Signal Processing Magazine,* vol. 20, pp. 21-36, 2003.

[3] H. S. Hou and H. C. Andrews, "Cubic Splines for Image Interpolation and Digital Filtering," *IEEE Transactions on Acoustics, Speech, and Signal Processing,* vol. ASSP-26, pp. 508-517, 1978.

[4] J. Yang, J. Wright, T. Huang, and Y. Ma, "Image super-resolution as sparse representation of raw image patches," in *2008 IEEE Conference on Computer Vision and Pattern Recognition*, ed: IEEE, 2008, pp. 1-8.

[5] C. Dong, C. C. Loy, K. He, and X. Tang, "Learning a Deep Convolutional Network for Image Super-Resolution," in *Computer Vision–ECCV 2014* vol. 8689, ed, 2014, pp. 184-199.

[6] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian Pyramid Networks for Fast and Accurate Super-Resolution," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ed: IEEE, 2017, pp. 5835-5843.

[7] J. Kim, J. K. Lee, and K. M. Lee, "Accurate Image Super-Resolution Using Very Deep Convolutional Networks," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)* vol. 38, ed: IEEE, 2016, pp. 1646-1654.

[8] J. Kim, J. K. Lee, and K. M. Lee, "Deeply-Recursive Convolutional Network for Image Super-Resolution," in *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ed: IEEE, 2016, pp. 1637-1645.

[9] W. B. Lim, M. K. Park, and M. G. Kang, "Spatially adaptive regularized iterative high-resolution image reconstruction algorithm," in *Visual Communications and Image Processing 2001*, 2000, vol. 4310, pp. 10-21: International Society for Optics and Photonics.

[10] H. Stark and P. Oskoui, "High-resolution image recovery from image-plane arrays,

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

using convex projections," *JOSA A,* vol. 6, no. 11, pp. 1715-1726, 1989.

[11] T. Komatsu, K. Aizawa, T. Igarashi, and T. Saito, "Signal-processing based method for acquiring very high resolution images with multiple cameras and its theoretical analysis," *IEE Proceedings I (Communications, Speech and Vision),* vol. 140, no. 1, pp. 19-25, 1993.

[12] S. Baker and T. Kanade, "Hallucinating faces," in *Automatic Face and Gesture Recognition, 2000. Proceedings. Fourth IEEE International Conference on*, 2000, pp. 83-88: IEEE.

[13] C. Liu, H.-Y. Shum, and C.-S. Zhang, "A two-step approach to hallucinating faces: global parametric model and local nonparametric model," in *Computer Vision and Pattern Recognition, 2001. CVPR 2001. Proceedings of the 2001 IEEE Computer Society Conference on*, 2001, vol. 1, pp. I-192-I-198 vol. 1: IEEE.

[14] X. Wang and X. Tang, "Hallucinating face by eigentransformation," *Systems, Man, and Cybernetics, Part C: Applications and Reviews, IEEE Transactions on,* vol. 35, no. 3, pp. 425-434, 2005.

[15] C. Liu, H.-Y. Shum, and W. T. Freeman, "Face hallucination: Theory and practice," *International Journal of Computer Vision,* vol. 75, no. 1, pp. 115-134, 2007.

[16] Y. Hu, T. Shen, and K. M. Lam, "Region-based Eigentransformation for face image hallucination," in *Circuits and Systems, 2009. ISCAS 2009. IEEE International Symposium on*, 2009, pp. 1421-1424: IEEE.

[17] J. Yang, J. Wright, T. S. Huang, and Y. Ma, "Image super-resolution via sparse representation.," *IEEE transactions on image processing : a publication of the IEEE Signal Processing Society,* vol. 19, pp. 2861-73, 2010.

[18] R. Timofte, V. De, and L. V. Gool, "Anchored Neighborhood Regression for Fast Example-Based Super-Resolution," in *2013 IEEE International Conference on Computer Vision* vol. 9006, ed: IEEE, 2013, pp. 1920-1927.

[19] J.-B. Huang, A. Singh, and N. Ahuja, "Single image super-resolution from transformed self-exemplars," in *2015 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, ed: IEEE, 2015, pp. 5197-5206.

[20] S. Anwar, S. Khan, and N. J. A. C. S. Barnes, "A deep journey into super-resolution: A survey," vol. 53, no. 3, pp. 1-34, 2020.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[21] 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,* vol. 38, 2016.

[22] C. Dong, C. C. Loy, and X. Tang, "Accelerating the super-resolution convolutional neural network," in *European Conference on Computer Vision*, 2016, pp. 391-407: Springer.

[23] Y. Tai, J. Yang, X. Liu, and C. Xu, "Memnet: A persistent memory network for image restoration," in *Proceedings of the IEEE international conference on computer vision*, 2017, pp. 4539-4547.

[24] K. Zhang, W. Zuo, and L. Zhang, "Learning a single convolutional super-resolution network for multiple degradations," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2018, pp. 3262-3271.

[25] B. Lim, S. Son, H. Kim, S. Nah, and K. M. Lee, "Enhanced Deep Residual Networks for Single Image Super-Resolution," *IEEE Computer Society Conference on Computer Vision and Pattern Recognition Workshops,* vol. 2017-July, pp. 1132-1140, 2017.

[26] 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*, 2018, pp. 2472-2481.

[27] M. Haris, G. Shakhnarovich, and N. Ukita, "Deep back-projection networks for super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 1664-1673.

[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)*, 2018, pp. 286-301.

[29] C. Ledig *et al.*, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* ed: IEEE, 2017, pp. 105-114.

[30] Z. Wang, J. Chen, and S. C. H. Hoi, "Deep Learning for Image Super-resolution: A Survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence,* pp. 1-1, 2020.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

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

[32] N. Ahn, B. Kang, and K.-A. Sohn, "Fast, accurate, and lightweight super-resolution with cascading residual network," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 252-268.

[33] J. Li, F. Fang, K. Mei, and G. Zhang, "Multi-scale residual network for image super-resolution," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 517-532.

[34] 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*, 2019, pp. 3867-3876.

[35] C. Wang, Z. Li, and J. J. a. p. a. Shi, "Lightweight image super-resolution with adaptive weighted learning network," 2019.

[36] W. Muhammad and S. Aramvith, "Multi-scale inception based super-resolution using deep learning approach," *Electronics,* vol. 8, no. 8, p. 892, 2019.

[37] X. Wang, Q. Wang, Y. Zhao, J. Yan, L. Fan, and L. Chen, "Lightweight single-image super-resolution network with attentive auxiliary feature learning," in *Proceedings of the Asian conference on computer vision*, 2020.

[38] X. Chu, B. Zhang, and R. Xu, "Multi-objective reinforced evolution in mobile neural architecture search," in *European Conference on Computer Vision*, 2020, pp. 99-113: Springer.

[39] X. Chu, B. Zhang, H. Ma, R. Xu, and Q. Li, "Fast, accurate and lightweight super-resolution with neural architecture search," in *2020 25th International Conference on Pattern Recognition (ICPR)*, 2021, pp. 59-64: IEEE.

[40] Y. Tai, J. Yang, and X. Liu, "Image super-resolution via deep recursive residual network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2017, pp. 3147-3155.

[41] Z. Hui, X. Wang, and X. Gao, "Fast and accurate single image super-resolution via information distillation network," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 723-731.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[42] J.-S. Choi and M. Kim, "A deep convolutional neural network with selection units for super-resolution," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 154-160.

[43] O. Bowen and C.-S. Bouganis, "Real-time image super resolution using an FPGA," in *2008 International Conference on Field Programmable Logic and Applications*, ed: IEEE, 2008, pp. 89-94.

[44] J. Lee and I.-C. Park, "High-Performance Low-Area Video Up-Scaling Architecture for 4-K UHD Video," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 64, no. 4, pp. 437-441, 2017.

[45] K. Seyid, S. Blanc, and Y. Leblebici, "Hardware implementation of real-time multiple frame super-resolution," *IEEE/IFIP International Conference on VLSI and System-on-Chip, VLSI-SoC,* vol. 2015-Octob, pp. 219-224, 2015.

[46] R. Redlich, L. Araneda, A. Saavedra, and M. Figueroa, "An Embedded Hardware Architecture for Real-Time Super-Resolution in Infrared Cameras," *Proceedings - 19th Euromicro Conference on Digital System Design, DSD 2016,* pp. 184-191, 2016.

[47] M. C. Yang, K. L. Liu, and S. Y. Chien, "A Real-Time FHD Learning-Based Super-Resolution System without a Frame Buffer," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 7747, pp. 1-5, 2017.

[48] Y. Kim, J.-S. Choi, and M. Kim, "2X Super-Resolution Hardware using Edge-Orientation-based Linear Mapping for Real-Time 4K UHD 60 fps Video Applications," *IEEE Transactions on Circuits and Systems II: Express Briefs,* vol. 7747, pp. 1-1, 2018.

[49] K. Sun *et al.*, "An FPGA-based residual recurrent neural network for real-time video super-resolution," 2021.

[50] J. Lee, D. Shin, J. Lee, J. Lee, S. Kang, and H.-J. Yoo, "A full HD 60 fps CNN super resolution processor with selective caching based layer fusion for mobile devices," in *2019 Symposium on VLSI Circuits*, 2019, pp. C302-C303: IEEE.

[51] Y. Kim, J.-S. Choi, M. J. I. T. o. C. Kim, and S. f. V. Technology, "A real-time convolutional neural network for super-resolution on FPGA with applications to 4K UHD 60 fps video services," vol. 29, no. 8, pp. 2521-2534, 2018.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[52] S. Lee, S. Joo, H. K. Ahn, and S.-O. J. I. A. Jung, "CNN acceleration with hardware-efficient dataflow for super-resolution," vol. 8, pp. 187754-187765, 2020.

[53] J. Trimek, "Public confidence in CCTV and fear of crime in Bangkok, Thailand," *International journal of criminal justice sciences,* vol. 11, no. 1, p. 17, 2016.

[54] S. Taweesaengsakulthai, S. Laochankham, P. Kamnuansilpa, and S. Wongthanavasu, "Thailand smart cities: what is the path to success?," *Asian Politics & Policy,* vol. 11, no. 1, pp. 144-156, 2019.

[55] V. Puncreobutr, P. Trinokorn, U. Dhanesschaiyakupta, J. Pusapukdepob, and M. A. C. Pa-Alisbo, "The Design of Real Estate Housing Project in Bangkok and the Metropolitan, Thailand for the Year 2021," *i-Manager's Journal on Management,* vol. 15, no. 1, p. 19, 2020.

[56] C. Sirirattanapol, M. Nagai, A. Witayangkurn, S. Pravinvongvuth, and M. Ekpanyapong, "Bangkok CCTV image through a road environment extraction system using multi-label convolutional neural network classification," *ISPRS International Journal of Geo-Information,* vol. 8, no. 3, p. 128, 2019.

[57] S. Saypadith and S. Aramvith, "Real-time multiple face recognition using deep learning on embedded GPU system," pp. 1318-1324: IEEE.

[58] T. Ganokratanaa, S. Aramvith, and N. Sebe, "Unsupervised anomaly detection and localization based on deep spatiotemporal translation network," *IEEE Access,* vol. 8, pp. 50312-50329, 2020.

[59] S. Aramvith, S. Pumrin, T. Chalidabhongse, and S. Siddhichai, "Video processing and analysis for surveillance applications," in *2009 International Symposium on Intelligent Signal Processing and Communication Systems (ISPACS)*, 2009, pp. 607-610: IEEE.

[60] S. Chen, H. M. Maung, and S. Aramvith, "Improving feature preservation in high efficiency video coding standard," in *2016 Asia-Pacific Signal and Information Processing Association Annual Summit and Conference (APSIPA)*, 2016, pp. 1-5: IEEE.

[61] S. Chen, S. Aramvith, and Y. Miyanaga, "Encoder Control Enhancement in HEVC Based on R-Lambda Coefficient Distribution," in *2019 International Symposium on Multimedia and Communication Technology (ISMAC)*, 2019, pp. 1-4: IEEE.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

[62] W. Ruangsang and S. Aramvith, "Super-resolution for hd to 4k using analysis k-svd dictionary and adaptive elastic-net," in *Digital Signal Processing (DSP), 2015 IEEE International Conference on*, 2015, pp. 1076-1080: IEEE.

[63] W. Ruangsang and S. Aramvith, "Efficient super-resolution algorithm using overlapping bicubic interpolation," in *2017 IEEE 6th Global Conference on Consumer Electronics (GCCE)*, 2017, pp. 1-2: IEEE.

[64] Z. Hui, X. Gao, Y. Yang, and X. Wang, "Lightweight image super-resolution with information multi-distillation network," in *Proceedings of the 27th acm international conference on multimedia*, 2019, pp. 2024-2032.

[65] W. Shi *et al.*, "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR),* pp. 1874-1883, 2016.

[66] A. Veit, M. J. Wilber, and S. J. A. i. n. i. p. s. Belongie, "Residual networks behave like ensembles of relatively shallow networks," vol. 29, 2016.

[67] Z. Meng, L. Li, X. Tang, Z. Feng, L. Jiao, and M. J. R. S. Liang, "Multipath residual network for spectral-spatial hyperspectral image classification," vol. 11, no. 16, p. 1896, 2019.

[68] M. Abdi and S. Nahavandi, "Multi-residual networks: Improving the speed and accuracy of residual networks," *arXiv preprint arXiv:1609.05672,* 2016.

[69] E. Agustsson and R. Timofte, "Ntire 2017 challenge on single image super-resolution: Dataset and study," in *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, 2017, pp. 126-135.

[70] M. Bevilacqua, A. Roumy, C. Guillemot, and M. L. Alberi-Morel, "Low-complexity single-image super-resolution based on nonnegative neighbor embedding," 2012.

[71] R. Zeyde, M. Elad, and M. Protter, "On single image scale-up using sparse-representations," in *International conference on curves and surfaces*, 2010, pp. 711-730: Springer.

[72] 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*

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

Conference on Computer Vision. ICCV 2001, 2001, vol. 2, pp. 416-423: IEEE.

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

[74] W. Badawy and G. A. Julien, *System-on-chip for Real-time Applications*. Springer Science & Business Media, 2012.

จุฬาลงกรณ์มหาวิทยาลัย
Chulalongkorn University

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY

VITA

NAME	Watchara Ruangsang
DATE OF BIRTH	12 March 1988
PLACE OF BIRTH	Bangkok
INSTITUTIONS ATTENDED	Chulalongkorn University (2015-2022) Chulalongkorn University (2011-2015) King Mongkut's University of Technology North Bangkok (2008-2011)
HOME ADDRESS	14/5 M.4, Tha TalatSam, Phran District, Nakhon Pathom, 73210
PUBLICATION	W. Ruangsang and S. Aramvith, "Efficient Super-Resolution Algorithm Using Overlapping Bicubic Interpolation," 2017 IEEE 6th Global Conference on Consumer Electronics (GCCE), Japan, 2017. Watchara Ruangsang, Supavadee Aramvith, Takao Onoye, “Line Buffer Image for Fast Super Resolution Based Convolution Neural Network,” IEEE ISCAS 2018 LATE BREAKING NEWS, Italy, 2018
AWARD RECEIVED	Gold Medal award from the Innovation week IWA 2020, Rabat, Morroco, on December 19, 2020. Silver Medal Award from The International exhibition of Inventions Geneva on March 22, 2021.

จุฬาลงกรณ์มหาวิทยาลัย
CHULALONGKORN UNIVERSITY