Image transformation using super resolution

by

Ishitaa Sayal

B.Tech, NIIT University, India, 2018

A REPORT

submitted in partial fulfillment of the
requirements for the degree

MASTER OF SCIENCE

Department Of Computer Science
Carl R. Ice College Of Engineering

KANSAS STATE UNIVERSITY
Manhattan, Kansas

2024

Approved by:

Major Professor
William H. Hsu

Copyright

Abstract

This study aimed to evaluate the effectiveness of two super resolution imaging (SR) models based on residual , Enhanced Deep Residual Networks for Single Image Super-Resolution (EDSR) and Densely Residual Laplacian Super-Resolution (DRLN), in enhancing the quality of photographic images. Following standard practice for model validation on the SR problem, quality was measured by computing objective metrics relative to benchmark image datasets, containing full-resolution reference images. Both a standard benchmark of diverse, high-quality images (*DIV2K*) and a domain-specific image corpus (*BeeMachine*) were used for this purpose.

One specific aim of SR is to improve outdoor images of organisms in the wild, for computer vision tasks such as object detection.This was the motivation for using *BeeMachine* image corpus, consisting of crowdsourced bumble bee photographs. Photographs from this collection vary in degree of motion blurring, viewing distance, and focal distance, to a greater degree than in standard SR benchmarks. To address this challenge, two general-purpose deep learning models were applied: Enhanced Deep Super Resolution (EDSR) and Densely Residual Laplacian Networks (DRLN).

Experiments were conducted to compare EDSR and DRLM performance using (*DIV2K*) as a validation standard and *BeeMachine* to assess the effectiveness of SR on a specific ecologically-valid use case. Results of these models were examined using a set of objective metrics such as peak signal-to-noise ratio (PSNR) and structural similarity (SSIM). These results demonstrated that both models were effective at increasing image resolution. When compared to the EDSR model, the DRLN model achieved a higher PSNR and SSIM value.

Furthermore, the models were also tested on images with varying levels of degradation to evaluate their robustness. The results showed that while both models performed well on moderately degraded images, the DRLN model was better at handling highly degraded

images. This suggests that the DRLN model may be more suitable for enhancing the quality of bumble bee images in challenging field conditions where image quality is compromised.

This study thus showed that extant models (EDSR and DRLN) whose performance is competitive with the state-of-the-art on an idealized general-purpose corpus (*DIV2K*) exhibit similar performance on the *BeeMachine)* corpus. This confirms the potential of SR using these models to enhance uploaded photographs at the lower-resolution (half and quarter scale) end of the experimental range, and to thereby facilitate research on the behavior and ecology of these important pollinators. While both models performed well, the DRLN model was found to be more effective in handling highly degraded images.

Table of Contents

List of Figures

List of Tables

Acknowledgments

I would like to express my sincere gratitude to my major professor, Dr. William H. Hsu, for his guidance and support throughout my graduate studies. His mentorship and expertise have been invaluable in shaping my research, developing my critical thinking skills, and preparing me for the next steps in my academic and professional journey. I am also deeply grateful to the members of my committee, Dr. Torben Amtoft, Dr. Doina Caragea, Dr. Caterina Scoglio, for their insightful feedback and helpful suggestions. Their contributions have been instrumental in refining my research questions, improving my methodology, and ensuring the quality of my work. Without their support and encouragement, this research would not have been possible. Thank you.

Chapter 1

Introduction

1.1 Problem Statement: Super-Resolution Imaging (SR)

This work addresses the problem of *super-resolution imaging* (SR), also known simply as super-resolution, a process of improving the level of detail and quality of an image beyond that of the original. It uses advanced techniques and algorithms to enhance the details and clarity of low-resolution images, making them look sharper and more detailed. Super-resolution is used in various fields, such as medical imaging, microscopy, remote sensing, video surveillance, satellite imaging, computer vision, and astronomy.

SR Ledig et al.[1] can be specified as a machine learning task in which the goal is to improve the resolution of an image, often by a factor of four or more, while retaining as much content and detail as feasible. The ultimate product is a high-resolution version of the original image. This activity can be used for a variety of purposes, including improving image quality, boosting visual detail, and increasing computer vision algorithms' accuracy.

The objective of SR is to improve image quality at the rescaled resolution (a case of upsampling or oversampling). Objective metrics that have been derived for SR, and which are typically used in validation, therefore, measure the proportion of similarity to a reference image. The *structural similarity index measure (SSIM)*[2] measures texture similarity over a weighted neighborhood given a specified weighting kernel (e.g., Gaussian). The *peak signal-*

to-noise ratio (PSNR)[2] measures the ratio of maximum possible power of a signal (from the original image) and the power of corrupting noise (in the SR output).

SR has a wide range of practical applications. In medical imaging, it can help doctors to detect and diagnose diseases by providing more detailed and accurate images. Super-resolution technique is used to improve the quality of Computed Tomography scans (CT scans) and Magnetic Resonance Scans (MRI scans). In satellite imaging, it can be used to monitor environmental changes and track the movement of natural disasters. In Forsyth and Ponce[3]computer vision, it can enhance the quality of images and help in better and more accurate detection based on detection tasks.

Traditional and Zhang et al.[4] deep learning methods are the two basic groups of super-resolution approaches. Classical algorithms have been available for decades, but their deep learning-based rivals outperform them. As a result, most current methods rely on data-driven deep learning models to recreate the necessary information for accurate super-resolution. In recent years, the development of deep learning techniques has led to significant improvements in super-resolution. Deep learning algorithms, such as convolution neural networks (CNNs), effectively generate high-quality images with improved resolution. These algorithms can learn the underlying patterns and structures in low-resolution images and then use this information to predict high-resolution images. As a subset of machine learning, it seeks to learn the link between input and output directly from data.

1.2 Application Domain: BeeMachine

Bumble bees are a species of bee that pollinates a wide range of blooming plants, including numerous crops. They are dubbed "bumble bees" because of the way they fly: they move their wings back and forth (rather than up and down like other bees do), producing a characteristic buzzing sound. They are a species of social bees in the genus Bombus, family Apidae. Bumble bees are classified into about 250 species globally. Each bumble bee species has distinct qualities and features that set it apart from other species. Physical qualities such as coloration, size, and hair pattern, as well as behavioral traits such as nesting patterns,

foraging activity, and the sorts of flowers they visit, are examples of these characteristics.

While bumble bee species can vary significantly in their physical and behavioral traits, they are generally similar in their basic biology and ecology. All bumble bees are social insects that live in colonies with a queen, workers, and males.

Bumble bees are crucial in pollinating various plants, including fruits, vegetables, and wildflowers. They are particularly effective at pollinating crops such as tomatoes, peppers, strawberries, blueberries, and cranberries, as well as many kinds of wildflowers.

In addition to their role as pollinators, bumble bees also play an important ecological role. They are a food source for many other animals, including birds and small mammals. They also help to control populations of other insects, such as aphids, which can be harmful to plants.

Unfortunately, many species of bumble bees are declining in number due to habitat loss, pesticide use, and climate change. This is a cause for concern, as bumble bees play an essential role in pollination and the broader ecosystem. Efforts to conserve and protect bumble bee populations are therefore essential not just for the bees themselves but for the plants and animals that rely on them. This has led to an urgent need to understand the causes of bee decline and establish conservation measures to maintain biodiversity and ensure a continuous supply of pollination services.

One of the most important but challenging aspects of bee research is accurately identifying individual bee species. This is crucial to determine the number of species and populations in a region and to track changes in bee populations over time. Traditional methods of bee identification involve careful observation and manual classification, which is time-consuming and requires extensive training.

To address this challenge, the BeeMachine group, a collaboration between the Spiesman[5] lab and Laboratory for Knowledge Discovery in Databases[6] (KDD Lab) at Kansas State University, is working towards the automatic species-level identification of bees from photos using deep learning classification models. Deep learning is a subset of machine learning that uses artificial neural networks to analyze and learn from large amounts of data. The goal is to develop models that can accurately identify different bee species from images, allowing

researchers to monitor bee populations more efficiently and accurately.

To automate the process of bee identification, the BeeMachine group has already developed a deep learning classification model[7]Spiesman et al. and is now investigating image alteration techniques to improve the identification process and obtain more accurate results. Image cropping, color adjustments, noise reduction, image sharpening, translation, and rotation are all image transformation techniques. Cropping an image is the process of eliminating a segment of an image; it can be beneficial for removing undesired aspects of an image or focusing on a specific portion of an image. Color adjustments or modification refers to the process of altering the colors or contrast of an image. It can be used to adjust color balance or to set a mood. A noise reduction technique is utilized to reduce noise from an image. It can help improve image quality in low-light situations and reduce compression artifacts. Image sharpening is the process of enhancing an image's edges and features. It can help to improve an image's clarity and sharpness. Translation entails moving a picture horizontally or vertically. It can be used to align photos or to create panoramas. In this study, the main focus will be on the image alteration technique "super-resolution" to improve bumble bee detection and identification. Investigate and conduct experiments on two of the most generalizable super-resolution models, Lim et al.[8]EDSR and Anwar and Barnes[9]DRLN.

As previously stated, Super-Resolution refers to boosting an image's resolution. Assume an image has a resolution (i.e., image dimensions) of 64x64 pixels and is super-resolved to 256x256 pixels. The procedure is termed *4x upsampling* in this example because the spatial dimensions (height and width of the image) are upscaled four times.

A Low Resolution (LR) image can now be mathematically modeled from a High Resolution (HR) image using a degradation function, delta, and a noise component.

$$y = D(x) + n \tag{1.1}$$

where $\mathbf{x}$ represents the high-resolution image $\mathbf{y}$ represents the low-resolution image $\mathcal{D}$ represents the degradation function, which includes blurring and down-sampling $\mathbf{n}$ represents

the noise component

Models in supervised learning approaches try to purposely degrade the image using the above equation and use this degraded image as inputs to the deep model, while the original images serve as HR ground truth. This pushes the model to learn a mapping between a low-resolution image and its high-resolution equivalent, which can subsequently be used to super-resolve any new image encountered during testing.

1.3 Application Use Cases of SR in Science and Medicine

In the case of Spiesman et al.[7] BeeMachine, this will help in enhancing the native resolution of uploaded images, leading to revelation of fine details and textures of the bumble bees for example their hair color pattern which is a widely used characteristics for bee identification. Another important bumble bee feature is the bee's wing venation pattern. the veins pattern in their wings which is different for different species. These details will help in improving model's overall accuracy and performance since it will be able to classify these bumble bees into difference species more accurately.

Another example where super resolution can be beneficial is Georgescu et al.[10]. the authors have developed a SR model that takes in images of MRI and CT scans perform super resolution. These newly generated high resolution images will help the doctors to more accurately and precisely diagnose the problem or disease the patient is suffering from and decide on the treatment.

1.4 Single Image Vs. Multi Image Super Resolution

Two types of Image SR methods exist depending on the input information available. In Single-Image SR, only one LR image is available, which needs to be mapped to its high-resolution counterpart. In Multi-Image SR, however, multiple LR images of the same scene or object are available, all of which are used to map to a single HR image.

In Single-Image SR methods, since the amount of input information available is low,

fake patterns may emerge in the reconstructed HR image, which has no discernible link to the context of the original image. This can create ambiguity which in turn can lead to the misguidance of the final decision-makers (scientists or doctors), especially in delicate application areas like bio-imaging.

Intuitively, we can conclude that Multi-Image SR produces better performance since it has more information to work with. However, the computational cost in such cases is also increased several-fold, making it infeasible in many application scenarios where there is a substantial resource constraint. Also, obtaining multiple LR images of the same object is tedious and impractical. Thus, Single-Image SR is more closely related to the real world, although it is a more challenging problem statement.

In this report, I will experiment with single image super-resolution models, Enhanced Deep Residual Single Image Super Resolution Model (EDSR), and Densely Residual Laplacian Super Resolution model (DRLN). EDSR uses deep residual networks with skip connections to address the challenge of super-resolving images with finer details. The model takes a single low-resolution image as input and produces a corresponding high-resolution image as output. The network is trained on pairs of high-resolution and low-resolution images to learn a mapping between the two image domains. DRLN is a deep learning-based image super-resolution model that uses a combination of residual learning and dense connections to improve the performance of image super-resolution. It is based on a deep convolutional neural network (CNN) with densely connected residual blocks. The residual blocks contain several convolutional layers, followed by a shortcut connection that skips the convolutional layers and directly connects the input to the output. The dense connections are used to connect each layer to every other layer in a feed-forward manner, allowing information to flow more efficiently through the network.

1.5 Organization of This Report

Chapter 1: Introduction

- Problem Statement

- Overview of super resolution and it's different types

In this chapter, I will provide a description of the problem statement and motivation for the study. I will then present some information about super-resolution. Finally, I will outline the report structure, highlighting the key topics and themes covered in each chapter.

Chapter 2: Literature Review

- Summary of the relevant literature and theoretical frameworks
- Synopsis of the models used for this survey
- sneak peek into other approaches

In Chapter 2, I will review the relevant literature and theoretical frameworks that inform the study. I also provide a synopsis of the models that will be used for this project and also provide a sneak peek into other approaches apart from the standard ones.

Chapter 3: Methodology

- Description and explanation of the parameters used for training and testing the model.
- Description of the evaluation metrics used and the reason for choosing them

Chapter 3 will describe the hyper-parameters used for training and testing the models and the configuration settings used. I will also discuss the evaluation models used and the reason for choosing them over other existing metrics.

Chapter 4: Experiments

- Details about the data set used and
- Presentation and interpretation of the findings

Chapter 4 will present details about the data set used in this project. I also will present and interpret the findings of the study.

Chapter 5: Summary and Future Work

- Summary of the key findings and contributions

In the final chapter, I will summarize the key findings and contributions of the study. I will also provide suggestions for future work.

Chapter 2

Related Work

2.1 Existing approaches to basic single-image SR

Traditional image super-resolution methods have given way to learning-based systems. The goal is to build a high-resolution image from a low-resolution RGB image. Filtering-based techniques such as bicubic upsampling and edge-preserving filtering are examples of traditional approaches. These filtering methods typically produce an incredibly smooth texture in the final high-resolution image. Several approaches look for similar patches in a training dataset or in the image itself using patch matching.

The tremendous capabilities of deep neural networks have recently resulted in remarkable advancements in SR. Various CNN architectures have been researched for SR since Dong et al.[11] originally presented a deep learning-based SR technique. Kim et al.[12]. pioneered the use of the residual network to train considerably deeper network designs and got improved results. They demonstrated, in particular, that skip connection and recursive convolution reduce the load of carrying identity information in the super-resolution network.

Dong et al.[11] suggested groundbreaking work in super-resolution by creating SRCNN, a fully convolutional network consisting of three convolution layers followed by ReLU. SRCNN accepts input in the form of bicubic interpolated pictures, which reduces high-frequency frequencies and necessitates additional computation. To reduce these additional computations,

Dong et al.[11] introduced FRSCNN. It takes the original low-resolution image as an input and employs deconvolution to upsample the features to the desired dimensions before the final objective function.

Deep neural networks trained with a variety of losses have recently been applied to super-resolution. Many new techniques for super-resolution are based on generative adversarial networks. Ledig et al.[1] SRGAN is a method for super-resolution photos that uses a GAN Creswell et al.[13] to generate high-resolution images. SRGAN's loss combines a deep feature-matching loss and an adversarial loss. Lai et al.[14] propose the Laplacian Pyramid Super Resolution Network for predicting the residual of high-frequency details in a coarse-to-fine image pyramid. Wang et al. proposed ESRGAN[15], which improves image super-resolution by incorporating a Relativistic GAN that calculates how much one image is more realistic than another. Wang et al. also explore class-conditioned picture super-resolution and propose SFT-GAN[16], which is trained using a GAN loss and a perceptual loss.

In this report, I have picked two super-resolution models that are not the current state-of-the art (SOTA) baseline but achieve performance that is similar to SOTA baseline models and are generic models that can run on any type of image, regardless of image type.

2.2 Enhanced Deep Residual Super Resolution Network (EDSR)

EDSR is a super-resolution model designed to generate high-quality, high-resolution images from low-resolution inputs. It is a model based on the residual network (ResNet) architecture, which is a type of deep neural network that uses skip connections to allow the model to learn residual functions. The EDSR model extends the ResNet architecture by using deeper networks with smaller filter sizes, which allows for better feature extraction and higher-quality super-resolution images.

The EDSR model consists of a series of residual blocks as shown in Figure 2.1, each including a rectified linear unit (ReLU) activation function layer sandwiched between two

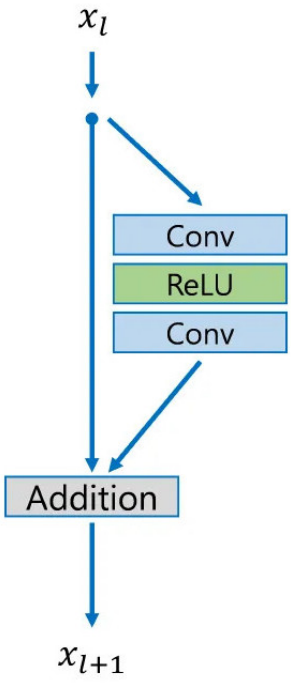


Figure 2.1: Residual bock Structure of EDSR model

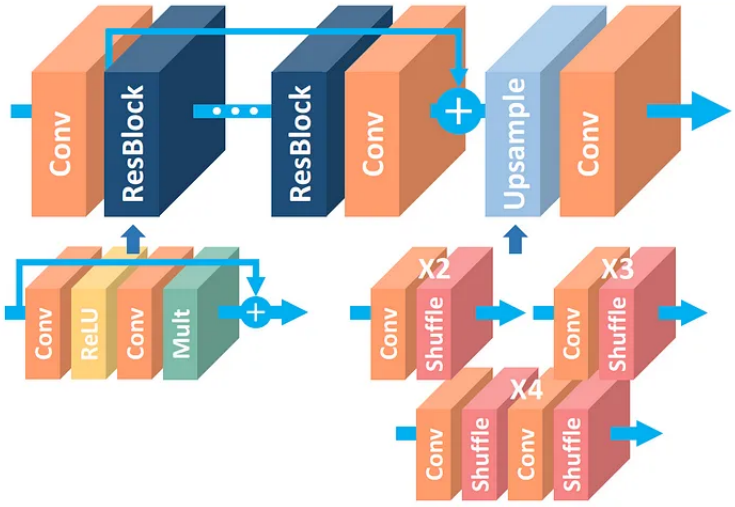


Figure 2.2: EDSR model Architecture

convolutional layers. The model also includes skip connections between the input and output of each residual block, which allows for the direct propagation of low-level features through the network. Figure 2.2 shows the architecture of EDSR model, which is comprised of a convolutional layer, residual blocks and an upsample layer which is modified based on the scale used to super-resolute the image.

The paper's authors also propose several modifications to the EDSR model to improve its performance further. For example, they introduce a new residual scaling technique that involves multiplying the output of each residual block by a learnable scalar, which helps to prevent the model from overfitting to the training data. They also introduce a new training strategy that involves using a high-pass filter to remove the low-frequency components of the input images during training, which helps the model focus on learning the high-frequency details.

2.3 Densely Residual Laplacian Super-Resolution (DRLN)

The DRLN model is a deep learning model for SR image reconstruction. It uses the dense connections between residual blocks to access previously computed features. A Laplacian pyramid attention module is also included in the model to weigh the features on several scales and according to their importance. The model proposed to address the issue of the state-of-the-art method model the features equally or on a limited scale, ignoring the abundant rich frequency representation at other scales, resulting in a lack of discriminative learning capabilities and capability across channels, ultimately limiting the ability of convolutional neural networks. The model architecture leverages cascade over residual on the residual architecture, which can aid in deep network training. Different connection types and cascading over residual on residual help in bypassing enough low-frequency information to learn more accurate representations.

The model consists of four main components: feature extraction, cascading residual on the residual, upsampling, and reconstruction. The feature extraction component uses a single convolutional layer to extract primitive features from the low-resolution input. The extracted features are then passed on to the cascading residual on the residual component. Cascading Residual on the Residual structure Figure 2.3 has a hierarchical architecture and is made up of cascading blocks. Each block is made up of dense residual Laplacian modules, which each have a densely connected residual unit, a compression unit, and a Laplacian pyramid attention unit. The output features from this component are novel and contribute

to the high performance of the DRLN model.

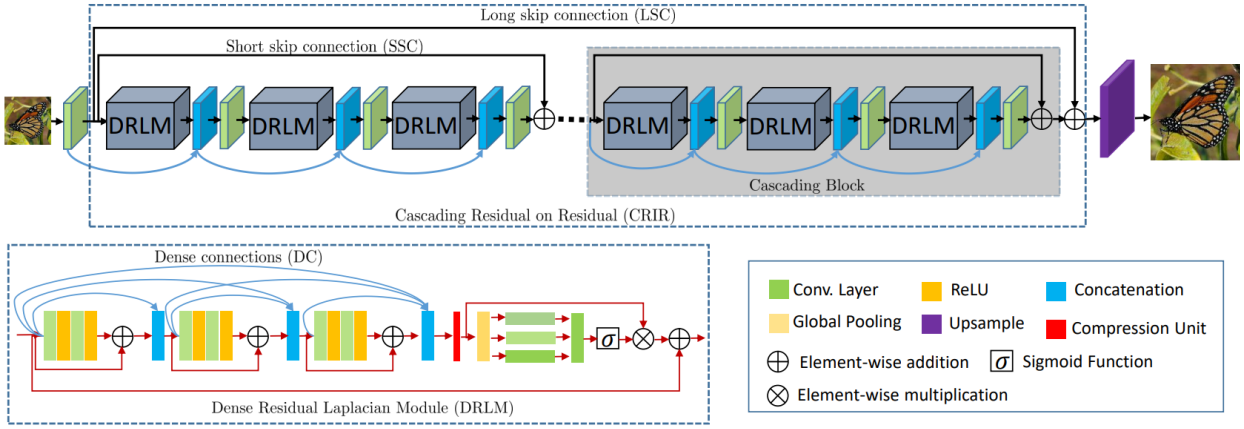


Figure 2.3: DRLN model architecture including details of Cascading Residual on Residual Blocks

The extracted deep features from the cascaded residual on the residual component are then upscaled through an upsampling component using the ESPCN method. Finally, the upscaled features are passed through a reconstruction component, which uses a single convolutional layer to predict the super-resolved RGB color channels as an output.

2.4 Additional Approches

Most existing super-resolution models use a synthetic low-resolution RGB image (typically downsampled from a high-resolution image) as input. Zhang et al. [17] proposed a model Zoom To Lean, Learn To Zoom that obtains authentic low-resolution images taken with shorter focal lengths and uses optically zoomed images as ground truth. Zoom To Lean, Learn To Zoom (Z2LL2Z) model is an image super-resolution approach that combines deep learning with a zooming mechanism to improve the quality of super-resolution images. The proposed model is a two-stage process that involves training a deep convolutional neural network (CNN) to generate high-resolution images from low-resolution inputs and then using the trained CNN to zoom in on the high-resolution image and generate even higher-resolution images. The model operates on raw sensor data and proposes a method to super-resolve images by jointly optimizing for the camera image processing pipeline and super-resolution

from raw sensor data. It uses SR-RAW to train a deep network with a novel contextual bilateral loss that is robust to mild misalignment between input and output images.

The current state-of-the-art (SOTA) baseline for raw sensor data is Z2LL2Z. Despite being SOTA, it cannot be utilized to do experiments with our data since, as previously stated, it inputs raw sensor data while the data sets used in this research are compressed images (.png images). Despite the various experiments performed in the publication, Zhang et al.[17] claim that the proposed model performed best with raw data rather than compressed data.

Chapter 3

Methodology

3.1 Experiments

For comparison identical experiments have been performed on EDSR and DRLN models. Both models are evaluated using a set of objective matrices such as Peak Signal To Noise Ratio (PSNR) and the Structural Similarity Index (SSIM). For this study, I have fine-tuned both models instead of training them from scratch. The results are computed for both x2 and x4 models.

The resulting super-resolution images are useful inputs for specialized models that are intended for image identification or classification, which can improve the efficacy and accuracy of the models. Georgescu et al. [10], one example uses EDSR model for the refinement of MRI or X-ray images utilizing super-resolution, which helps in the more exact and accurate identification of disorders. Furthermore, super-resolution models can help YOLOv7 (You Only Look Once) by Wang et al. [18], since these improved images can help these systems in detection and classification abilities.

3.1.1 Training Setting

Increasing the number of parameters is the simplest technique to improve the performance of the network model. Model performance in the convolutional neural network can be improved

by stacking several layers or increasing the number of filters. In the convolution neural network model, the performance can be improved by either stacking up many of the layers or increasing the number of feature channels. General CNN architecture with depth or in other words the number of layers B and width or number of feature channels F occupies roughly $O(BF)$ memory with $O(BF^2)$ parameters. Therefore increasing the value of F can maximize the model capacity given limited computational resources. Increasing the number of feature maps after a certain level would mathematically make the training procedure unstable. To solve this issue EDSR model introduces a residual scaling layer at the end of the last conventional layer in a residual block.

For this study, there are two configurations that can be used to fine-tune the EDSR model. In the first setting, we can fine-tune the model with a number of layers as 32 and the number of feature channels as 256, in this case since there are high number of feature channels we can set the residual scale as 0.1 as mentioned in[8]. The other set of configurations that can be used is the number of layers as 16 and the number of channels as 64 with the residual scale set to 1. For comparison of both the models, I have used the second set of configurations as the DRLN model has a single configuration that has a number of feature channels 64 and a number of layers 16.

3.1.2 Hyper-parameters Used

Keeping this in mind in EDSR's CNN architecture, the number of layers is set as 16 and the number of feature channels is set to 64 as DRLN works with the same configuration and it will be a fair comparison if both are compared on the same level.

The model employs a 48x48 RGB input patch size from a low-resolution image with equivalent high-resolution patches. All the images are preprocessed by subtracting the mean RGB value of the associated data set. The training data is augmented by applying random horizontal flips and 90-degree rotations.

The Model is trained with ADAM optimizer by setting the value of $\beta 1 = 0.9, \beta 2 = 0.999, \epsilon = 10^{-8}$, and the minimum batch size used is 16. The learning rate is initialized as

10^{-4} and halved at every 2×10^5 minibatch updates.

3.2 Evaluation

As mentioned earlier the models are evaluated with the help of PSNR and SSIM metrics.

PSNR: PSNR is defined as the ratio of a signal's maximum achievable power to the power of corrupting noise that influences the fidelity of its representation. PSNR is frequently used to regulate the quality of digital signal transmission. In the case of images, each pixel can be thought of as a component of an 8-bit RGB signal.

$$\text{PSNR} = 10.\log_{10}\left(\frac{\text{MAX}_I^2}{\text{MSE}}\right) \tag{3.1}$$

where MAX_I is the minimum valid value for pixels and MSE is the mean square error between the high-resolution image and the super-resolved image. A logarithmic scale is used because the pixel values have a very wide dynamic range, and PSNR is a pixel by pixel comparison over the entire image.

SSIM: It is designed based on three factors — correlation, luminance distortion, and contrast distortion. This index is calculated on various windows of the image instead of a direct pixel-by-pixel comparison. If x, y are windows of size NxN in images:

$$\text{SSIM}(x, y) = \frac{(2\mu_x\mu_y + c_1)(2\sigma_{xy} + c_2)}{(\mu_x^2 + \mu_y^2 + c_1)(\sigma_x^2 + \sigma_y^2 + c_2)} \tag{3.2}$$

where μ_x is the average of x; μ_y is the average of y; σ_x^2 is the variance of x; σ_y^2 is the variance of y; σ_{xy} is the covariance of x and y; $c_1 = (k_1 L)^2$, $c_2 = (k_2 L)^2$ two variables to stabilize the division with a weaker denominator; L is the dynamic range of pixel-values $k_1 = 0.01, k_2 = 0.03$ by default

When comparing two super-resolution models, such as EDSR and DRLN, the choice of evaluation metrics depends on the specific use case and the requirements of the application. While Learned Perceptual Image Patch Similarity (LPIPS) is a state-of-the-art metric that

takes into account perceptual differences between images, it can be computationally expensive and may not be necessary for all applications. PSNR and SSIM are simpler metrics that are well-established in the image processing community and are widely understood and easy to interpret, making them a good choice for comparing different models. Therefore, the choice to use only PSNR and SSIM scores for evaluating both models has been made based on a trade-off between evaluation accuracy and computational cost, as well as the need for a standard and widely understood evaluation metric for comparison purposes.

Chapter 4

Experiments

4.1 Data and Benchmarks

In this study, I am fine-tuning the EDSR and DRLN models using the DIV2K and Bee-Machine data sets. The DIV2K dataset includes a variety of photos, including those of a building, the natural world, animals, people, etc., whereas the bee machine dataset only includes photographs of bumble bees from a variety of species, including Bombus Affins, Bombus Appositus, Bombus Bifarius, and others. Both datasets include roughly 1K high-resolution images and their counterpart low-resolution bicubic images.

4.1.1 DIV2K

Agustsson and Timofte[19] DIV2K is a common SISR dataset, which is also a benchmark for both the models, i.e. EDSR and DRLN. It consists of a total of 1,000 images, divided into 800 for training, 100 for validation, and 100 for testing. It was gathered for the NTIRE2017 and NTIRE2018 Super-Resolution Challenges to stimulate research towards image super-resolution with more realistic deterioration. This dataset includes LR photos with various forms of degradations. In addition to the typical bicubic downsampling, many forms of degradations are taken into account when generating LR images for different tracks of the challenges. The data collection includes high-resolution training and validation data as

well as bicubic low-resolution images. The low-resolution images are downsized versions of the high-quality images by factors of 2, 3, and 4. The same structure is employed for bicubic test images.

4.1.2 Bee Machine

Bumblebee data, Bee Machine is used as the test data set and both models are fine-tuned with it before being compared to see which is more effective for enhancing bee identification. It includes various bee photos taken around the United States and Canada. These images were gathered from Bumble Bee Watch, iNaturalist, and BugGuide. data set contains approximately 1,20,000 images belonging to 42 different species. For this study, I took a very small portion of the data, a size similar to DIV2K. The data set for this study similar to DIV2K consists of 800 training images, 100 validation images, and 100 testing images.

In contrast to DIV2K, the bee machine data set does not provide bicubic low-resolution images. To obtain bicubic low-resolution images of factors 2x, 3x, and 4x, I created a method that generates the desired images when high-resolution images are supplied as input. It uses bicubic interpolation to downsample the images and obtain bicubic bumblebee images.

4.2 Result

I have provided the quantitative results of the two selected models EDSR and DRLN, on two data sets, First the common single image super-resolution data set, DIV2K and second on the bumble bee data set, the bee machine data. Table 4.1 shows the PSNR and SSIM values for both the models on two scales x2 and x4. Table 4.1 shows that for both data sets, the DRLN model beats the EDSR model in terms of PSNR and SSIM scores for two different scales x2 and x4.

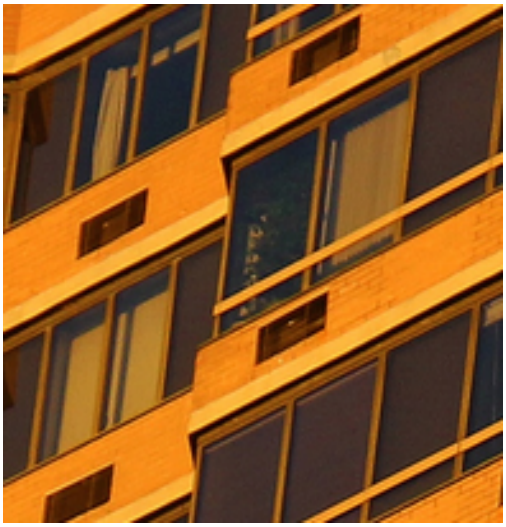
We can also observe the marginal difference between both models. Looking at the PSNR and SSIM values for DIV2K for x2 we can see that there is +1-2% difference in PSNR and +1% in SSIM score similarly for x4 there is +1% or less margin in PSNR score and +0.2%

or less in SSIM score. Looking at Bee Machine data for x2 there is +1% or less margin in PSNR score +0.1%in SSIM score, whereas for x4 there is +1% margin in PSNR and +0.5% in SSIM score. looking at the margin differences, we can see that there for DIV2K x2 margin is higher than x4, whereas for Bee Machine data x2 margin is lower than x4, which shows that it is possible that the high-quality images in the DIV2K dataset allow the model to extract more fine-grained details at lower magnification levels and on the other hand, the lower-quality images in Bee Machine may not have enough details to extract at lower magnification levels. Figure 4.1 and 4.2 shows the super-resoluted images produced by the EDSR and DRLN model on DIV2K and Bee Machine Data. We can clearly observe that DRLN outperforms EDSR model as images super-resolved with DRLN are clearer as compared to images produced by EDSR.

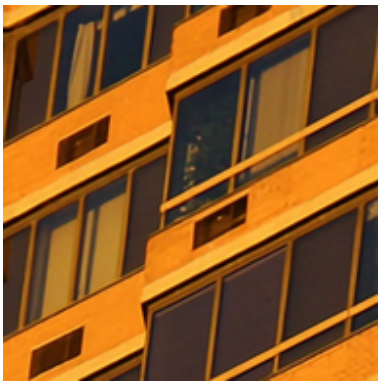
A larger margin difference is detected when comparing the performance of super-resolution models EDSR and DRLN on DIV2K data utilizing peak signal-to-noise ratio (PSNR) and structural similarity index (SSIM) scores on both full-resolution and downsampled validation pictures. On downsampled validation images, the EDSR Model returns a PSNR score of 33.708 and an SSIM score of 0.914. PSNR has a marginal difference of +2% or more and SSIM has +4% when compared to EDSR values at full resolution. Similarly, DRLN on downsampled validation images yields a PSNR score of 34.591 and an SSIM score of 0.966, with a +2% difference in PSNR and a +3% difference in SSIM when compared to full resolution. A larger difference may imply that the algorithm is more effective at restoring high-frequency information.

Dataset	Model	x2		x4	
		PSNR	SSIM	PSNR	SSIM
DIV2K	EDSR	35.695	0.964	31.271	0.899
	DRLN	**37.291**	**0.970**	**31.925**	**0.915**
Bee Machine	EDSR	46.683	0.984	42.139	0.946
	DRLN	**48.756**	**0.998**	**43.156**	**0.993**

Table 4.1: PSNR and SSIM values for EDSR and DRLN Models

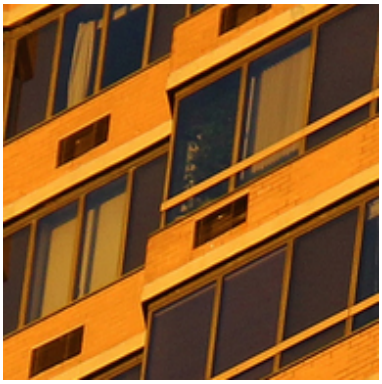


(a) DIV2K original Image

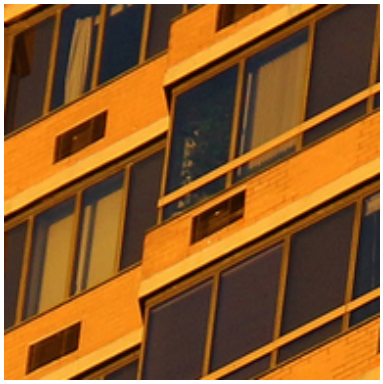


(b) EDSR X2

(c) EDSR X4



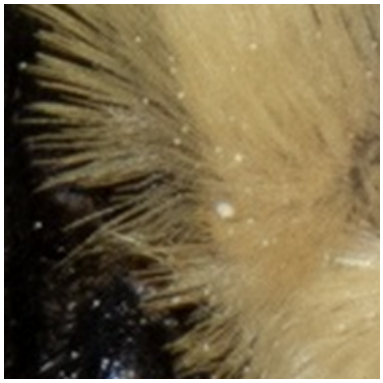
(d) DRLN X2



(e) DRLN X4

Figure 4.1: The images above are super-resolution images for the DIV2K data set for both the EDSR and DRLN models at scales X2 and X4.

(a) Bee Machine Original



(b) Bee Machine Original Cropped

(c) EDSR X2

(d) EDSR X4

(e) DRLN X2

(f) DRLN X4

Figure 4.2: The images above are super-resolution images for the Bee Machine data set for both the EDSR and DRLN models at scales X2 and X4.

To Validate the effectiveness of super resolution on BeeMachine dataset I performed a an experiment where I ran super resolution on BeeMachine images on factor of x2 and x4 and compared the BeeMachine classifier's performance using Mean Average Precision. The mean average Precision score shows that the Super Resolution can be a good choice as a data

augmentation technique as Mean Average precision score as shown in 4.2 for images scales at x4 and x2 magnification gives better results as compared to images with super resolution. From 4.2 we can also see that Mean Average Precision of x4 magnification scale is better than x2 magnification scale.

Table 4.2: Mean Average Precision for Original, x2 Scaled, and x4 Scaled Images

	Original	**x2**	**x4**
Mean Avg Precision	0.66	0.81	0.82

Chapter 5

Summary and Future Work

5.1 Summary

This work has evaluated EDSR and DRLN, two extant models whose performance is competitive with the state-of-the-art (SOTA) on *DIV2K*. This was done bu first repeating experiments and reproducing trends in PSNR and SSIM on DIV2K, then confirming these trends, and finally and examining downstream mAP results on the computer vision task of object detection and classification using *BeeMachine* data.

Bumble bees are vital pollinators that play an important role in environmental health. Because of its widespread decline in recent years, it is critical to investigate and document its behavior and ecology. To identify bee species, the bee-machine group has already constructed a deep learning classification model. Performing super-resolution on the images is one of the ways to improve the identifying process. I conducted a survey study on two generalized super-resolution models, EDSR and DRLN, using two data sets, DIV2K and Bee machine, in this report I fine-tuned both models and compared the outcomes using DIV2K and bee machine data. Both models performed well in the super-resolution test; however, DRLN beats EDSR in terms of PSNR and SSIM score, as shown in table‘4.1. The trials used two data sets to confirm the models' universality as well as the results, implying that given any data set, DRLN will most likely outperform EDSR. The results also reveal that DRLN

outperforms EDSR on any given scale, whether 2x or 4x.

According to the results of this survey, DRLN models outperform EDSR models for any given dataset and scale factor. Based on the results, we can conclude that DRLN is a suitable fit for the bee machine project to perform super-resolution and help improve the bee identification process.

We cam also conclude from the brief experiment performed using super-resolved images on the BeeMachine classifier that the Super Resolution technique can be a good choice as a data augmentation technique that can be used to improve the overall performance and accuracy of the deep learning model developed by the KDD lab students.

5.2 Future Work

5.2.1 Image Quality Assessment (IQA) with SR for Image Selection and Critiquing

Super-resolution can be used to improve the resolution of images, which can be particularly useful for object detection and classification tasks where fine details need to be accurately identified and analyzed. This, in turn, can be used to select which uploaded images are to be retained by predicting expected gain or to *critique*[20] what aspects of an image need to be improved using explainable AI techniques to relate deficiencies in the performance of computer vision models on a task to parts of an image (e.g., segmentation masks).

In this context, Hosu et al.[21] Image Quality Assessment (IQA) becomes a critical factor. IQA is the process of evaluating and forecasting image quality utilizing a large, diverse sample of real-world photos and human assessment scores. This technique aids in the development of computational models that accurately reflect human perceptions of image quality, hence addressing the demand for automated and scalable image quality evaluation tools. However, traditional image quality assessment (IQA) metrics like PSNR and SSIM may not be sufficient to capture the improvements in object detection and classification tasks when super-resolution is used. IQA offers essential metrics for evaluating and enhancing Super-

Resolution (SR) techniques, ensuring they effectively improve image detail and quality in line with human visual perception and application requirements.

One way to improve IQA for object detection and classification tasks with SR is to use task-specific evaluation metrics in self-supervised learning. For example, in the case of bumblebee images, one could use the downstream average precision (AP) and mean average precision (mAP) as evaluation metrics, which are commonly used in object detection and classification tasks. These metrics take into account both the accuracy and completeness of object detection and classification results, which are important factors for real-world applications.

Another way to improve IQA for object detection and classification tasks when super-resolution is used is to use perceptual metrics that are specifically designed to capture the improvements in human perception. For example, the Learned Perceptual Image Patch Similarity (LPIPS) metric considers perceptual differences between images, which can be particularly useful for object detection and classification tasks where human perception plays a critical role. However, as mentioned earlier, LPIPS can be computationally expensive and may not be necessary for all applications.

Overall, improving IQA for object detection and classification tasks when super-resolution is used requires careful consideration of the specific use case and the requirements of the application, as well as the use of task-specific and perceptual metrics that are designed to capture the improvements in human perception and accuracy of object detection and classification results.

One of the crucial steps in future work is the Image Quality assessment of the resulting images using the Keep & Toss Experiment designed using the Yolov7 model. It will help assess the performance of the super-resolution model, evaluating how well the super-resolution process improves the visibility and distinguishability of bees in images, which is critical for applications in classifying different bee species.

5.2.2 Improving Images using SR in Reinforcement Learning

Another important aspect that can handled or worked upon can be implementingBulitko et al.[22] control policies for image transformations to improve image quality (either using SSIM, PSNR, and similar metrics or using IQA as above). These rules entail predicting and modifying processing strategies in response to incoming data streams, which can considerably increase the efficiency and effectiveness of super-resolution activities. In practice, this means that the models would react to the data they are currently processing while also making intelligent predictions about future data features. This approach may lead to more adaptive, faster, and more accurate super-resolution, especially in dynamic contexts where data attributes change rapidly. Integrating these policies will be helpful in situations where quick and precise upscaling is crucial; for example in the case of bee identification, sometimes, to identify similar species, precise scaling will be crucial to identify correctly. The primary difficulty would be to create algorithms that can effectively predict the type of incoming data and alter the model's parameters in real-time while maintaining overall performance.

Bibliography

[1] C. Ledig, L. Theis, F. Huszar, J. Caballero, A. Cunningham, A. Acosta, A. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi. Photo-realistic single image super-resolution using a generative adversarial network. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*. IEEE Computer Society, 2017. doi: 10.1109/CVPR.2017.19. URL https://doi.ieeecomputersociety.org/10.1109/CVPR.2017.19.

[2] Structural Similarity Index, Peak Signal-to-Noise Ratio. https://juliaimages.org/latest/examples/image_quality_and_benchmarks/structural_similarity_index/. Accessed: 2024-04-10.

[3] D. Forsyth and J. Ponce. *Computer Vision: A Modern Approach*. Always learning. Pearson, 2012. ISBN 9780136085928. URL https://books.google.com/books?id=gM63QQAACAAJ.

[4] Aston Zhang, Zachary C. Lipton, Mu Li, and Alexander J. Smola. *Dive into Deep Learning*. Cambridge University Press, 2023. https://D2L.ai.

[5] Brian Spiesman. Pollinator Ecology Lab at KState. https://spiesmanecology.com. Accessed: 2023-03-15.

[6] Laboratory for Knowledge Discovery in Databases. Laboratory for knowledge discovery in databases. www.kddresearch.org. Accessed: 2023-03-10.

[7] Brian J. Spiesman, Claudio Gratton, Rich Hatfield, William H. Hsu, Sarina J. Jepsen, Brian P. McCornack, Krushi Patel, and Guanghui Wang. Assessing the potential for deep learning and computer vision to identify bumble bee species from images. *Scientific Reports*, 11, 2021. URL https://api.semanticscholar.org/CorpusID:233182227.

[8] Bee Lim, Sanghyun Son, Heewon Kim, Seungjun Nah, and Kyoung Mu Lee. Enhanced deep residual networks for single image super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1132–1140, 2017. doi: 10.1109/CVPRW.2017.151.

[9] Saeed Anwar and Nick Barnes. Densely residual laplacian super-resolution. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 44(3), 2022. doi: 10.1109/TPAMI.2020.3021088.

[10] M. Georgescu, R. Ionescu, A. Miron, O. Savencu, N. Ristea, N. Verga, and F. Khan. Multimodal multi-head convolutional attention with various kernel sizes for medical image super-resolution. In *2023 IEEE/CVF Winter Conference on Applications of Computer Vision (WACV)*, 2023. URL https://arxiv.org/abs/2204.04218.

[11] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *European Conference on Computer Vision*, 2014. URL https://api.semanticscholar.org/CorpusID:18874645.

[12] Jiwon Kim, Jung Kwon Lee, and Kyoung Mu Lee. Accurate image super-resolution using very deep convolutional networks. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 1646–1654, 2016. doi: 10.1109/CVPR.2016.182.

[13] Antonia Creswell, Tom White, Vincent Dumoulin, Kai Arulkumaran, Biswa Sengupta, and Anil A. Bharath. Generative adversarial networks: An overview. *IEEE Signal Processing Magazine*, 35(1), 2018. doi: 10.1109/MSP.2017.2765202.

[14] Wei-Sheng Lai, Jia-Bin Huang, Narendra Ahuja, and Ming-Hsuan Yang. Deep laplacian pyramid networks for fast and accurate super-resolution. In *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2017. doi: 10.1109/CVPR.2017.618.

[15] Xintao Wang, Ke Yu, Shixiang Wu, Jinjin Gu, Yihao Liu, Chao Dong, Yu Qiao, and Chen Change Loy. Esrgan: Enhanced super-resolution generative adversarial networks.

In Laura Leal-Taixé and Stefan Roth, editors, *Computer Vision – ECCV 2018 Workshops*, 2019. ISBN 978-3-030-11021-5.

[16] Xintao Wang, K. Yu, Chao Dong, and Chen Change Loy. Recovering realistic texture in image super-resolution by deep spatial feature transform. *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 2018. URL https://api.semanticscholar.org/CorpusID:4710407.

[17] Xuaner Zhang, Qifeng Chen, Ren Ng, and Vladlen Koltun. Zoom to learn, learn to zoom. In *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. doi: 10.1109/CVPR.2019.00388.

[18] Chien-Yao Wang, Alexey Bochkovskiy, and Hong-Yuan Mark Liao. YOLOv7: Trainable bag-of-freebies sets new state-of-the-art for real-time object detectors. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2023. doi: 10.1109/CVPR52729.2023.00721.

[19] Eirikur Agustsson and Radu Timofte. NTIRE 2017 challenge on single image super-resolution: Dataset and study. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, 2017. doi: 10.1109/CVPRW.2017.150.

[20] Bas H.M. van der Velden, Hugo J. Kuijf, Kenneth G.A. Gilhuijs, and Max A. Viergever. Explainable artificial intelligence (xai) in deep learning-based medical image analysis. *Medical Image Analysis*, 79, 2022. ISSN 1361-8415. doi: 10.1016/j.media.2022.102470. URL http://dx.doi.org/10.1016/j.media.2022.102470.

[21] Vlad Hosu, Hanhe Lin, Tamás Szirányi, and Dietmar Saupe. Koniq-10k: An ecologically valid database for deep learning of blind image quality assessment. *IEEE Transactions on Image Processing*, 29:1–1, 01 2020. doi: 10.1109/TIP.2020.2967829.

[22] Vadim Bulitko, Ilya Levner, and Russell Greiner. Real-time lookahead control policies. 2002. URL https://api.semanticscholar.org/CorpusID:8319899.