University of Moratuwa

Department of Electronics and Telecommunication

EN3160 - Image Processing and Machine Vision

Project Report

Lokugeegana D.L. 200356A, Pathirana K.P.T.R. 200449L

Introduction

Aiming to restore damaged picture content or complete missing information, image restoration and enhancement are important computer vision jobs. The fields of graphics and vision have shown a growing interest in these basic research areas in recent years. Not only has the number of relevant papers steadily increased, but significant advancements have also been made.

Every advancement makes it easier for humans or computers to use images to complete new tasks, with image enhancement or restoration acting as a crucial frontend. Thus, it should come as no surprise that there are an increasing number of applications in domains like electronics, medical image analysis, automotive, remote sensing, and surveillance. The advent and widespread usage of wearable and mobile gadgets present yet another opportunity for new uses and expedient techniques.

Single-image super-resolution is the process of restoring high-frequency rich information from a single low-resolution input image to a high-resolution image based on a series of previous examples of low-resolution and matching high-resolution images. There are four tracks in the task.

Track 1: "Classic bicubic" uses the traditional/classic settings seen in the literature on single-image super-resolution; in other words, the degradation operators are the ground truth high resolution image downscaled using bicubic interpolation (imresize Matlab function).

Track 2: "Realistic mild" unfavorable circumstances It makes the assumption that training pairs of low- and high-resolution photos can be used to estimate the degradation operators (such as blur kernel, decimation, and downscaling technique) that simulate the image acquisition process from a digital camera. While the randomly produced blur kernels and the pixel shifts they cause are consistent across images, the degradation operators' settings are not. Modeling the low to high image resolution mapping relation is the goal of the sizable training collection of examples of low and equivalent high resolution images.

Track 3: "Realistic difficult" unfavorable circumstances - same as previously, but with more extensive corruptions.

In order to better simulate the real world scenario, Track 4: "Realisitic wild" conditions makes the assumption that the degradation process (which mimics the process of acquiring images from digital cameras) varies from image to image. Specifically, this means that the parameters of the degradation operators change at specific intervals, and that the randomly generated blur kernels and the pixel shifts they cause also differ from image to image. This environment is the most like the real "wild" world.

Dataset

The NTIRE 2018 Challenge on Single Image Super-Resolution used the DIV2K dataset. Here is a brief description of the dataset:

Overview: The DIV2K dataset is a collection of 2K resolution high-quality images with diverse contents. It was specifically collected for the NTIRE 2017 and NTIRE 2018 Super-Resolution Challenges to encourage research on image super-resolution with more realistic degradation.

Contents: The dataset contains 1,000 images with different scenes.

Split: The images are split into 800 for training, 100 for validation, and 100 for testing.

For the NTIRE 2018 Challenge, the following configurations of the DIV2K dataset were used:

- Bicubic x8: This configuration uses bicubic downscaling by a factor of 8 to create low-resolution images.
- Realistic Mild x4: This configuration introduces realistic mild degradations to the high-resolution images before downscaling by a factor of 4. The degradations include motion blur, Poisson noise, and pixel shifting.
- Realistic Wild x4: This configuration introduces realistic wild degradations to the high-resolution images before downscaling by a factor of 4. The degradations are of different levels from image to image.

Related Work

In the domain of image super-resolution, numerous methods have been introduced and refined over time. This section highlights notable contributions in this field.

- **Super-Resolution Convolutional Neural Network (SRCNN):** SRCNN is one of the first methods to apply deep neural networks for image super-resolution. It directly learns an end-to-end mapping between the low/high-resolution images. The mapping is represented as a deep convolutional neural network (CNN) that takes the low-resolution image as the input and outputs the high-resolution one.

- **Super-Resolution Generative Adversarial Network (SRGAN):** SRGAN is a GAN-based method for single image super-resolution. It uses a perceptual loss function which consists of an adversarial loss and a content loss. The adversarial loss pushes the solution to the natural image manifold using a discriminator network that is trained to differentiate between the super-resolved images and original photo-realistic images

- **Efficient Sub-Pixel Convolutional Neural Network (ESPCN):** ESPCN introduces an efficient sub-pixel convolution layer which learns an array of upscaling filters to upscale the final low-resolution feature maps into the high-resolution output.

Method

Method Section: Enhanced Deep Super-Resolution (EDSR)

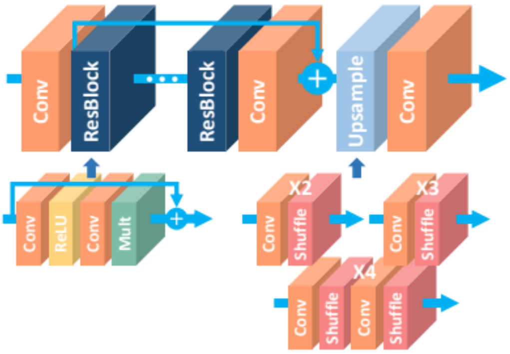
In this project, the chosen method for image super-resolution is the Enhanced Deep Super-Resolution (EDSR) model. EDSR represents a significant advancement in the realm of super-resolution techniques and was developed as an improvement over its predecessor, SRResNet.

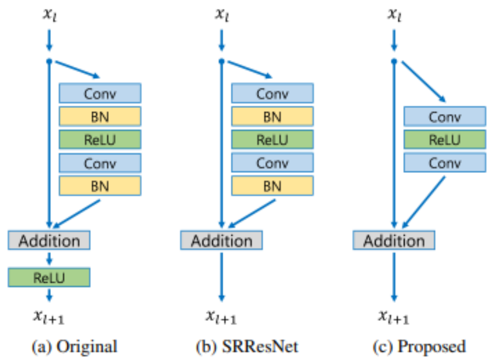
Model Architecture

EDSR builds upon the foundations of SRResNet, which successfully addressed challenges related to processing time and memory consumption in super-resolution tasks. However, it was observed that the ResNet architecture employed in SRResNet, while effective for image classification, was not optimal for super-resolution tasks. As a result, EDSR takes a more deliberate approach to create an optimized model for super-resolution.

Optimization by Module Removal

One of the key optimizations in EDSR involves the removal of certain modules from the ResNet architecture. For instance, BatchNormalization, a commonly used module in deep neural networks, is excluded from the EDSR model. This decision is rooted in research findings, which indicated that BatchNormalization can limit the range flexibility of the model.





<u>Training</u>

In this project, the training process for the super-resolution model involved two configurations: "bicubic x8" and "wild x4." These configurations presented distinct challenges and necessitated specific adjustments in the training procedure.

Bicubic x8 Configuration

The "bicubic x8" configuration was geared toward achieving an eightfold increase in image resolution. This setting leveraged bicubic interpolation to downscale the high-resolution images by a factor of 8. The training data consisted of low-resolution images derived from this downscaled high-resolution set.

Wild x4 Configuration and Noise Reduction

The "wild x4" configuration introduced a real-world scenario, including variations in degradation processes such as noise. To address the presence of noise in this scenario, a noise reduction step was integrated into the training process. This noise reduction step aimed to mitigate the impact of noise while preserving as much image detail as possible.

The noise reduction process involved reducing each image by a factor of 2 while applying a median filter. This filter operated on a 2x2 pixel grid, replacing the central pixel value with the median value of the four pixels. This step successfully reduced noise, but it also halved the image resolution. Consequently, to maintain the target super-resolution factor of 4x, the same model used for "x8 bicubic" was repurposed and further trained in this context.

Model Configuration

The super-resolution model was meticulously crafted with 64 filters and included 16 residual blocks. This architecture was designed to capture intricate image details and patterns, facilitating high-quality super-resolution.

Optimization and Learning Rate Schedule

The model was optimized using the Adam optimizer, and a dynamic learning rate schedule was employed. This schedule-initiated training with a specific learning rate, and after a certain number of training steps, the learning rate was adjusted. This adaptive approach to learning rate scheduling facilitated improved convergence and enhanced model performance.

Loss Function

The loss function utilized during training was the mean absolute error (MAE), commonly known as L1 Loss. MAE quantifies the dissimilarity between the super-resolved output and the corresponding ground truth high-resolution images, serving as a guide for the model to minimize errors.

Image Augmentation

To enhance the model's robustness and ability to handle a wide range of image variations, image augmentation techniques were employed during training. Random cropping, flipping, and rotation were applied to the input data, ensuring that the model learned to adapt to diverse image conditions.

Batch Size and Steps Per Epoch

The training process was executed with a batch size of 16, facilitating the concurrent processing of multiple data samples. Each epoch consisted of 200 steps, providing comprehensive exposure to the training data and enabling the model to learn from a diverse set of image examples.

The model underwent 200 epochs of training on both tracks, ensuring comprehensive learning. Additionally, Kaggle's GPU access was harnessed to expedite the training process, significantly enhancing computational efficiency.

Result

For bicubic x8,

Mean PSNR for Validation Dataset: 24.775328

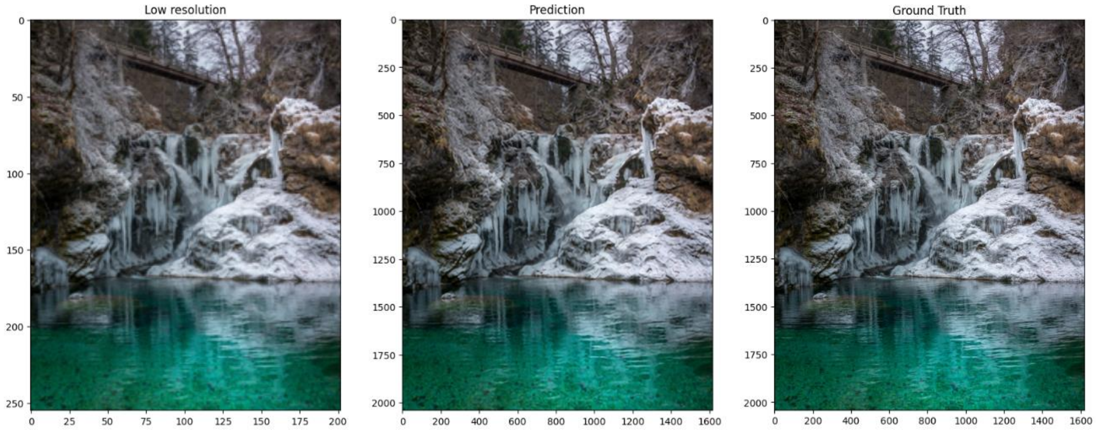
Mean SSIM for Validation Dataset: 0.6442471

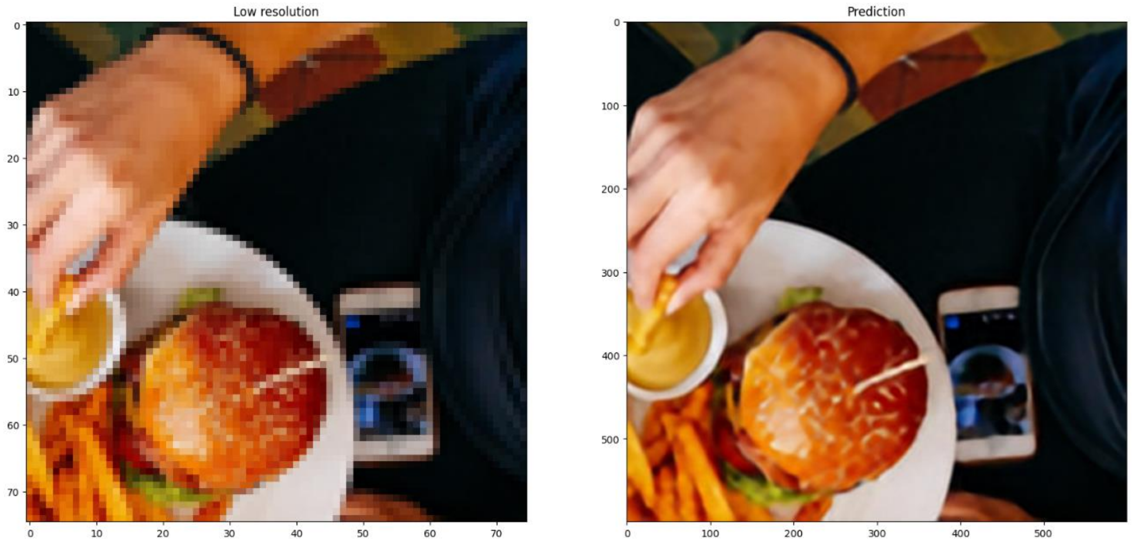
For wild x4,

Mean PSNR for Validation Dataset: 18.751143

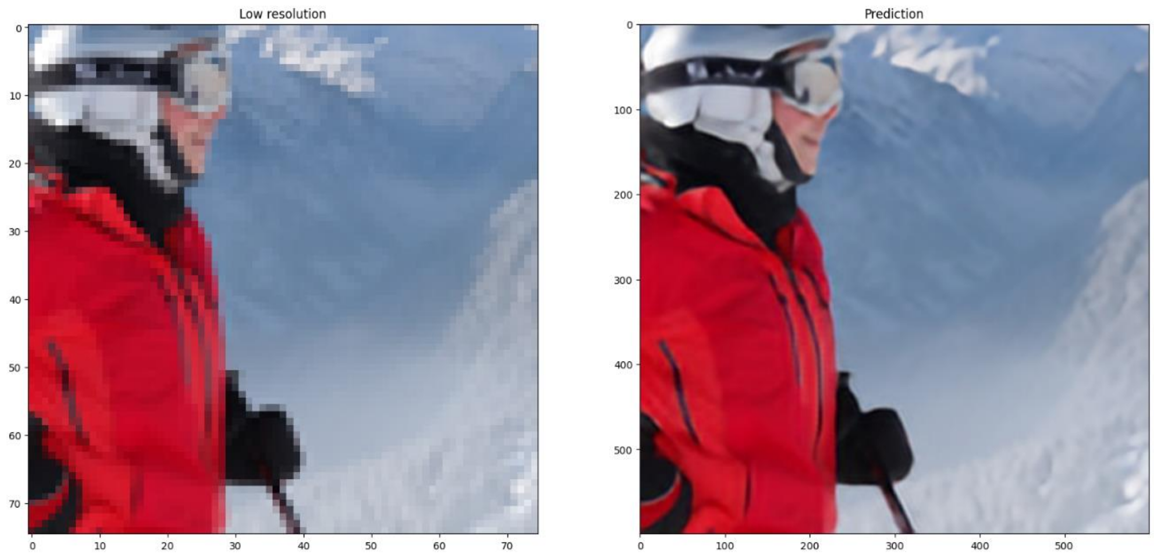
Mean SSIM for Validation Dataset: 0.5073049

Bicubic X8



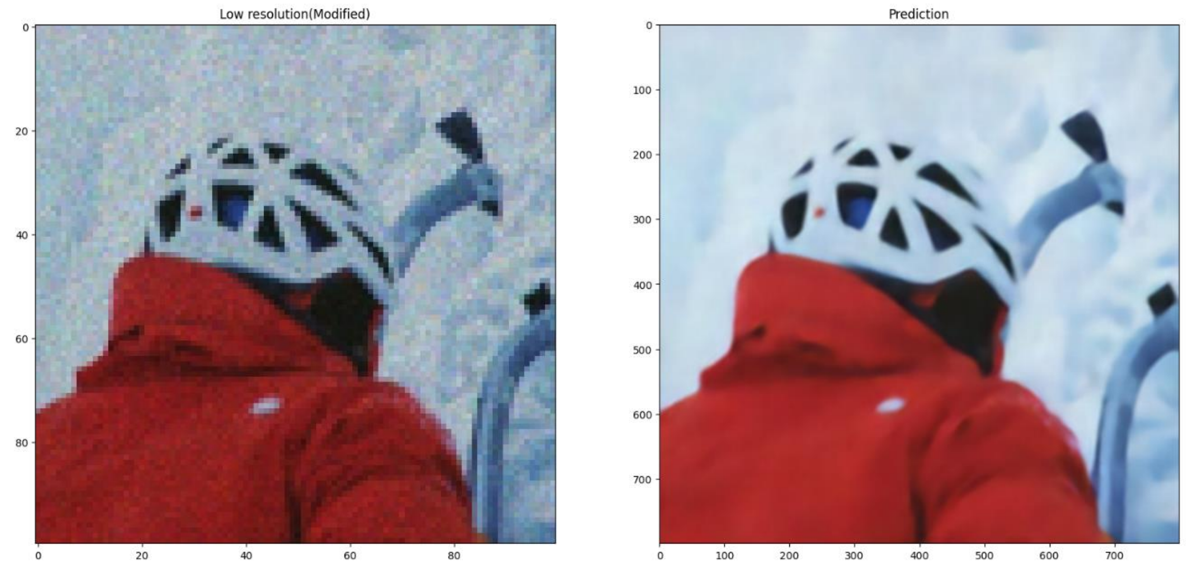


PSNR: 24.102085
SSIM: 0.66654354

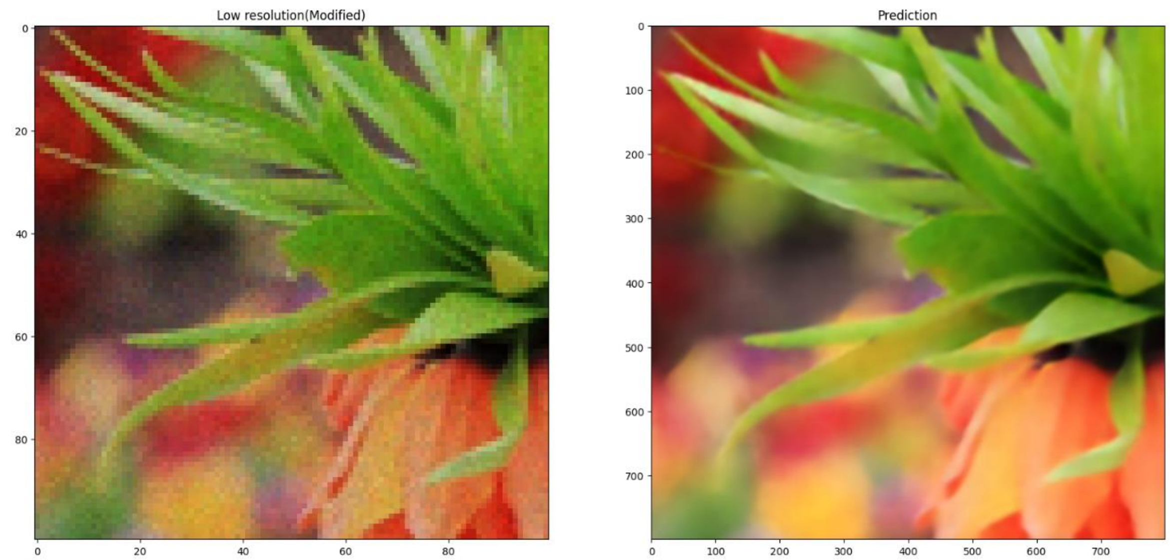


PSNR: 28.248075
SSIM: 0.8546037

Wild



PSNR: 23.8381
SSIM: 0.8653485



PSNR: 21.950336
SSIM: 0.7661794

Discussion

There are 2 methods of evaluation, PSNR and SSIM.

PSNR (Peak Signal-to-Noise Ratio):

PSNR is a measure of the quality of a reconstructed or compressed image in comparison to the original, where higher values indicate better quality.

It is calculated using the following formula:

PSNR = 10 * log10((MAX^2) / MSE)

MAX is the maximum possible pixel value in the image (usually 255 for 8-bit images).

MSE (Mean Squared Error) is the average of the squared differences between the corresponding pixels of the original and distorted images.

PSNR is typically expressed in decibels (dB).

A higher PSNR value suggests a smaller difference between the original and distorted image, indicating better image quality.

SSIM (Structural Similarity Index):

SSIM = (luminance similarity) * (contrast similarity) * (structure similarity)SSIM is designed to better align with human perception, making it a more robust metric for assessing image quality.

It takes into account not only pixel-level differences but also structural and perceptual information in the images.

In the Wild section mean PSNR is relatively lower than bicubic x8, that is because wild has more noise which will be result in more harder to achieve the ground truth.

GitHub link - https://github.com/Dulan24/Super-Resolution-Project/tree/main

References

NTIRE 2018 Challenge on Single Image Super-Resolution: Methods and Results (thecvf.com)

https://openaccess.thecvf.com/content_cvpr_2017_workshops/w12/papers/Lim_Enhanced_Deep_Residual_CVPR_2017_paper.pdf