Image Super Resolution using Deep Learning Algorithms

2023S AML 3104 Neural Networks and Deep Learning

Mukul Bisht   
*AIMT   
(of Affiliation)*  
*Lambton College  
(of Affiliation)*Toronto, Canada  
[C0857928@mylambton.ca](mailto:C0857928@mylambton.ca)

Neelesh Vashist

*AIMT*

*(of Affiliation)*

*Lambton college*

*(of Affiliation)*

Toronto,Canada

[c0858518@mylambton.ca](mailto:c0858518@mylambton.ca)

Rohit Kumar

*AIMT*

*(of Affiliation)*

*Lambton college*

*(of Affiliation)*

Toronto,Canada

[c0859060@mylambton.ca](mailto:c0859060@mylambton.ca)

Saurabh Singh

*AIMT*

*(of Affiliation)*

*Lambton college*

*(of Affiliation)*

Toronto,Canada

[c0859334@mylambton.ca](mailto:c0859334@mylambton.ca)

*Abstract*

The project "Image Super-Resolution" aims to enhance the quality of low-resolution images using advanced deep learning techniques. By leveraging convolutional neural networks (CNNs) and generative adversarial networks (GANs), the project addresses the challenge of upscaling images while preserving important details. The report discusses the significance of image super-resolution in various domains, presents the methodology involving data preprocessing, network architecture, and training process. It evaluates the model's performance through quantitative and qualitative measures, showcasing its ability to generate high-quality, realistic high-resolution images from low-resolution inputs. The project demonstrates the potential of AI-driven image enhancement for practical applications in imaging technologies.

# Introduction

In the realm of modern visual content, the quest for enhancing image quality has fuelled the development of innovative solutions. This project probes into the fascinating realm of image super-resolution, a fundamental task in image processing and computer vision. Our endeavour revolves around harnessing the transformative capabilities of deep learning algorithms, specifically the Super-Resolution Convolutional Neural Network (SRCNN) and the Super-Resolution Generative Adversarial Network (SRGAN). With the proliferation of high-definition displays and the growing demand for clearer visual content, the ability to upscale low-resolution images while preserving essential details becomes crucial. This project delves into the application of deep learning algorithms, specifically the Super-Resolution Convolutional Neural Network (SRCNN) and Super-Resolution Generative Adversarial Network (SRGAN), for image super-resolution tasks.

The project aims to significantly enhance the resolution and visual fidelity of images. This report unveils the inner workings, performance, and prospects of these algorithms, offering insights into their pivotal role in elevating the realm of image enhancement.

The remainder of this report is structured as follows: Section 2 delves into Literature review, examining papers that were explored for create high resolution image from low resolution images before the advent of deep learning techniques. Section 3 details of the architecture for both the algorithms, encompassing the system architecture, Section 4 presents the experiments undertaken, including the design, dataset preparation, evaluation metrics, hyperparameter selection. Finally, Section 5 provides a conclusive summary and future work related to the image super resolution project.

Top of Form

# LITERATURE REVIEW

The landscape of image super resolution has been extensively explored through numerous studies, focusing on different approaches and techniques to provide the desired output. This section summarizes relevant papers that contribute to the understanding of these methodologies.

1. Dong et al [1] Employed deep convolutional networks to achieve high-quality image super-resolution, enhancing details and edges effectively.

2. Ledig et al [2]. introduce a GAN-based approach for photo-realistic single-image super-resolution, enhancing image quality. (Ledig, Theis, Huszár, Caballero, Cunningham, Acosta, Wang).

3. Kim et al [3] present accurate image super-resolution via deep convolutional networks, achieving state-of-the-art performance. (Kim, Lee, Lee).

4. Lim et al [4] enhance residual networks for super-resolution, delivering improved performance in single-image processing. (Lim, Son, Kim, Nah, Lee.

5. Zhang et al [5] introduce residual learning using CNNs for image denoising, achieving superior results in denoising applications. (Zhang, Zuo, Chen, Meng, Zhang).

6. Wang et al [6] propose ESRGAN, a GAN-based solution for enhanced super-resolution tasks, achieving realistic and high-quality outcomes. (Wang, Yu, Wu, Gu, Liu, Dong).

7. Shi et al [7] present an efficient sub-pixel CNN for real-time image and video super-resolution, optimizing image quality. (Shi, Caballero, Huszár, Totz, Aitken, Bishop, Wang)

Researchers Dong, Ledig, Kim, Lim, Zhang, Wang, and Shi have contributed to image super-resolution advancements. Their work includes deep convolutional networks, GAN-based approaches, enhanced residual networks, efficient sub-pixel CNNs, and residual learning. Their methods improve image quality and achieve state-of-the-art performance in various single-image and video super-resolution tasks.

# Methodologies

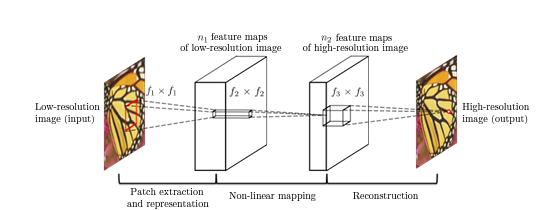
## 3.1 Introduction to Methodologies

In this section, we provide an overview of the methodologies employed in our experiment to design and implement the image super resolution project Our approach combines convolutional neural network along with generative adversarial network. The algorithm architecture and layering are depicted in the architecture diagram (Figure 1)

## 3.2 System Architecture Diagram for SRCNN

The SRCNN is a pioneering deep learning model designed for single-image super-resolution. It was introduced in the paper "Image Super-Resolution Using Deep Convolutional Networks" by Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang in 2016.

SRCNN is a deep learning-based approach designed to enhance the resolution of images. It uses a convolutional neural network to learn the mapping between low-resolution (LR) and high-resolution (HR) image patches. The network is trained on a dataset of LR-HR image pairs to learn the inherent relationships between low and high-resolution features



[Figure 1: SRCNN Architecture Diagram]

## 3.2.1 SRCNN Architecture Details

The SRCNN architecture consists of three main layers: the patch extraction layer, the non-linear mapping layer, and the reconstruction layer.

**Patch Extraction Layer:**

This layer takes the LR input image and divides it into small overlapping patches. Its function is to extract small patches from the low-resolution input image. These patches are fed into the subsequent layers for processing.

**Non-linear mapping:**

Conv1: The first convolutional layer learns low-level features from the LR patches. The ReLU activation function introduces non-linearity, enhancing the features captured by the filters.

Conv2: Deeper layers like Conv2 learn higher-level features and more complex patterns from the features obtained in Conv1.

Conv3: Conv3 fuses and transforms the features learned in Conv2, preparing them for reconstruction.

The activation function used is ReLU (Rectified Linear Unit). ReLU introduces non-linearity and helps the network learn complex relationships between features.

**Reconstruction:**

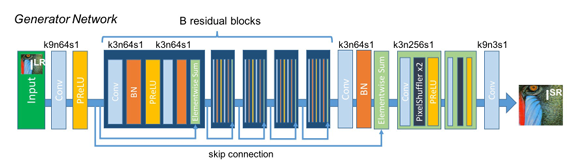
Deconv: The deconvolutional layer processes the transformed features from Conv3 and reconstructs high-resolution patches. Since this layer uses a linear activation function, it doesn't introduce non-linearity to the reconstructed image.

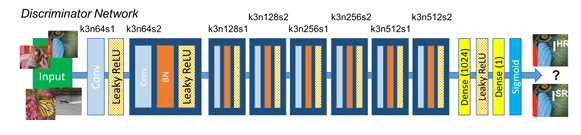
**Up sampling**

Before feeding the low-resolution patch into the network, bicubic interpolation is often applied to upscale the image to the desired target resolution. This upsampling helps the network focus on learning the residual details between low and high-resolution images.

## 3.3 System Architecture Diagram for SRGAN

SRGAN is a deep learning model designed for single-image super-resolution using generative adversarial networks. It was introduced in the paper "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network" by Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, et al., in 2017.





[Figure 2: SRGAN Architecture Diagram]

## 3.3.1 SRGAN Architecture Details

The SRGAN (Super-Resolution Generative Adversarial Network) architecture consists of two main parts: the Generator and the Discriminator. The Generator takes a low-resolution image and generates a high-resolution version by extracting features, enhancing details through residual blocks, and rearranging pixels for upscaling. The Discriminator evaluates the realism of the generated image compared to real images.

**Generator Block:**

The generator takes a low-resolution image as input, extract features from LR input and applies parametric rectified Linear unit (PreLU) activation function. It has convolutional layer to reduce the no of channels to down sample features.

Several residual blocks enhance features while maintaining information. PreLU introduces non-linearity within each residual block.

**Discriminator block:**

The Discriminator Network in SRGAN evaluates the authenticity of generated high-resolution images. It consists of convolutional layers that extract features from images and fully connected layers that combine these features to determine realism.

The output layer produces a probability score indicating whether an image is real or fake. During training, the Discriminator learns to distinguish between real and generated images. This adversarial process creates a feedback loop with the Generator, pushing it to create more authentic high-resolution images. The Discriminator's role is pivotal in training the Generator to produce visually convincing results.

**Explanation of Layers:**

* Feature Extraction: The initial layers capture essential features from the low-resolution input image.
* Shrinking: Reducing the number of channels helps downsample features.
* Non-linear Mapping: Residual blocks enhance features while mitigating vanishing gradient problems.
* Expanding: Deconvolutional layers and Pixel Shuffle rearrange pixels for upscaling.
* Output Layer: Produces the high-resolution output image.

# Experiment

In this section, we present the details of our model implementation, including the design of the experiments, dataset preparation, evaluation metrics, and the results along with their analysis.

## 4.1 SRCNN Model Implementation

* The DIV2K dataset is a widely used dataset in the field of image super-resolution, designed specifically for training and evaluating deep learning models that aim to enhance the resolution of images. "DIV2K" stands for "Dataset for Image and Video Super-Resolution: High-Definition."
* The DIV2K dataset contains 900 high-quality RGB images of different size for training and validation[8].
* Machine Configuration: The experiments were executed on a system with Intel Core i5 5200U 2,2 GHz and 16 gb ram
* Implementation Details: The experiments were implemented Keras and Tensorflow libraries.

## 4.1.1 SRCNN Model Training

SRCNN is trained using a dataset of paired low-resolution and high-resolution images. The network's goal is to minimize the difference between the network's output and the corresponding ground truth high-resolution images. Mean Squared Error (MSE) is commonly used as the loss function during training.

* Hyperparameter Tuning: we conducted experiment with ADAM optimizer, ReLU as activation function with learning rate of 0.001 and MSE as loss function. Total no of parameters for the model training were 85889

## 4.1.2 SRCNN Image quality assessment

We employed a set of evaluation metrics to assess the image quality generated from model

* **MSE**: The difference between the Pixel value of one image and the corresponding Pixel value of the other image. The MSE measures the average of the square of the errors
* **PEAK SIGNAL TO NOISE RATION (PSNR):** PSNR is used to calculate the ratio between the maximum possible signal power and the power of the distorting noise.
* **Structure Similarity Index Method (SSIM):** It measures similarity between two images

## 4.1.3 SRCNN Results and Analysis

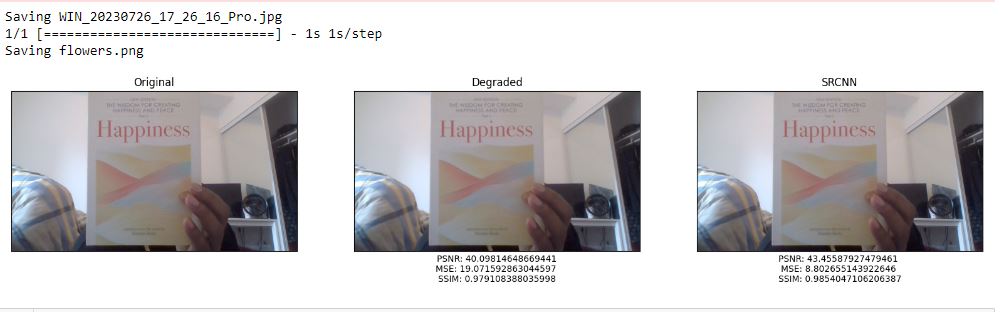
Model performed well on trained dataset and test datasett.

Below is the screenshot of the output where Mean square Error has decreased drastically while the Peak signal to noise ratio has improved marginally for SRCNN model compared to degraded image



[Figure 3: Original, degraded and SRCNN generated image from DIV2K dataset]

Model was also tested on customized dataset to see the model performance on MSE, PSNR and SSIM



[Figure 4: Original, degraded and SRCNN generated image from Custom dataset]

We tested the model on custom dataset, that were generated from computer webcam and below are the results that were achieved from the model.

* **PSNR:** Peak signal to noise ratio improved in SRCNN generated image (43.45) compared to the Degraded image (40.09)
* **MSME:** Mean Square Error decreased drastically from 19.07 for degraded image to SRCNN image (8.80)
* **SSIM**: Structural similarity improves slightly for SRCNN image (0.98) compared to Degraded image (0.97)

## 4.2 SRGAN Model Implementation

* We used the MIRFLICKR dataset having 25000 images and 3gb in size. Images were in .PNG format [9]
* All images resized to 128x128 to represent HR and 32x32 to represent LR.
* The generator and discriminator are jointly trained using adversarial loss (binary\_crossentropy and content loss (MSE - Mean Squared Error) to achieve photo-realistic single-image super-resolution.
* Machine Configuration: The experiments were executed on a system with Intel Core i5 5200U 2,2 GHz and 16 gb ram
* Implementation Details: The experiments were implemented Keras and Tensorflow libraries.

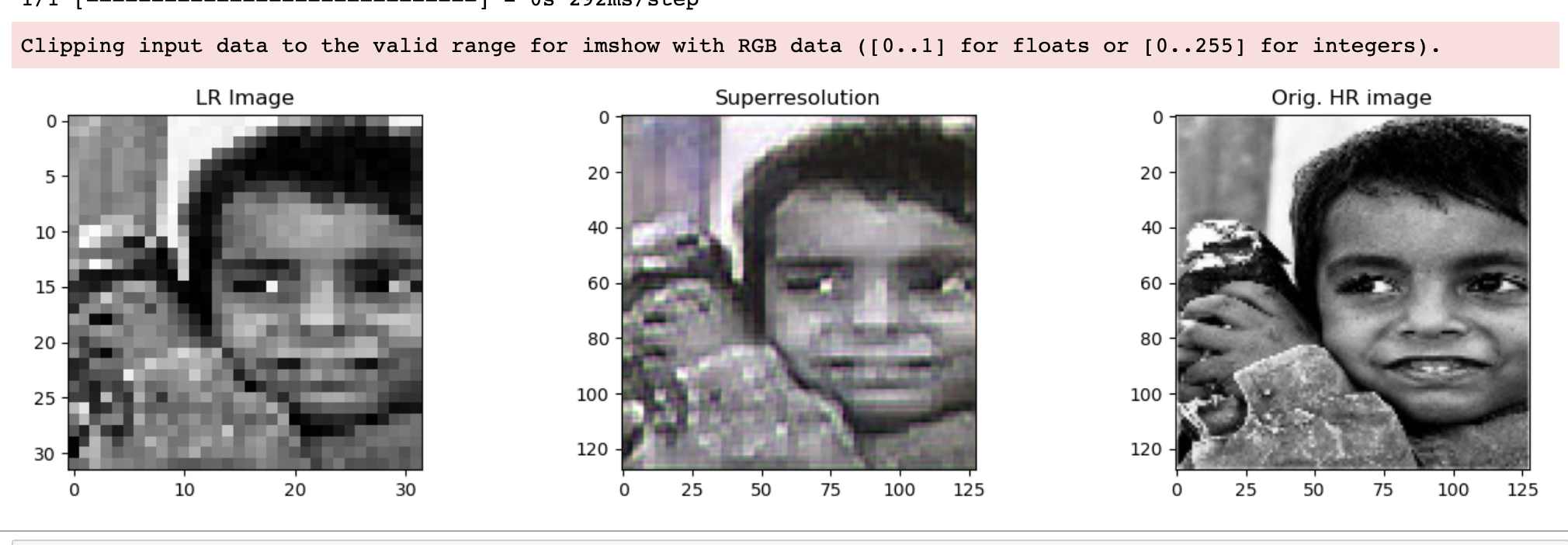
## 4.2.1 SRGAN Model Training

* Model was trained on training sample of 700 images and made prediction on test sample of 300 images.
* With total number of parameters for generator is 2044291 and for discriminator is 2325568
* Hyperparameter Tuning: we conducted experiment with ADAM optimizer, PReLU as activation function with learning rate of generator at 0.001, learning rate of discriminator at 0.001. Total epochs were 10
* The generator and discriminator are jointly trained using adversarial loss (binary\_crossentropy and content loss (MSE - Mean Squared Error) to achieve photo-realistic single-image super-resolution.
* Model was also trained on sample of 5000 sample images with 100 epochs, learning rate of both generator and discriminator at 0.003 with early stopping mechanism

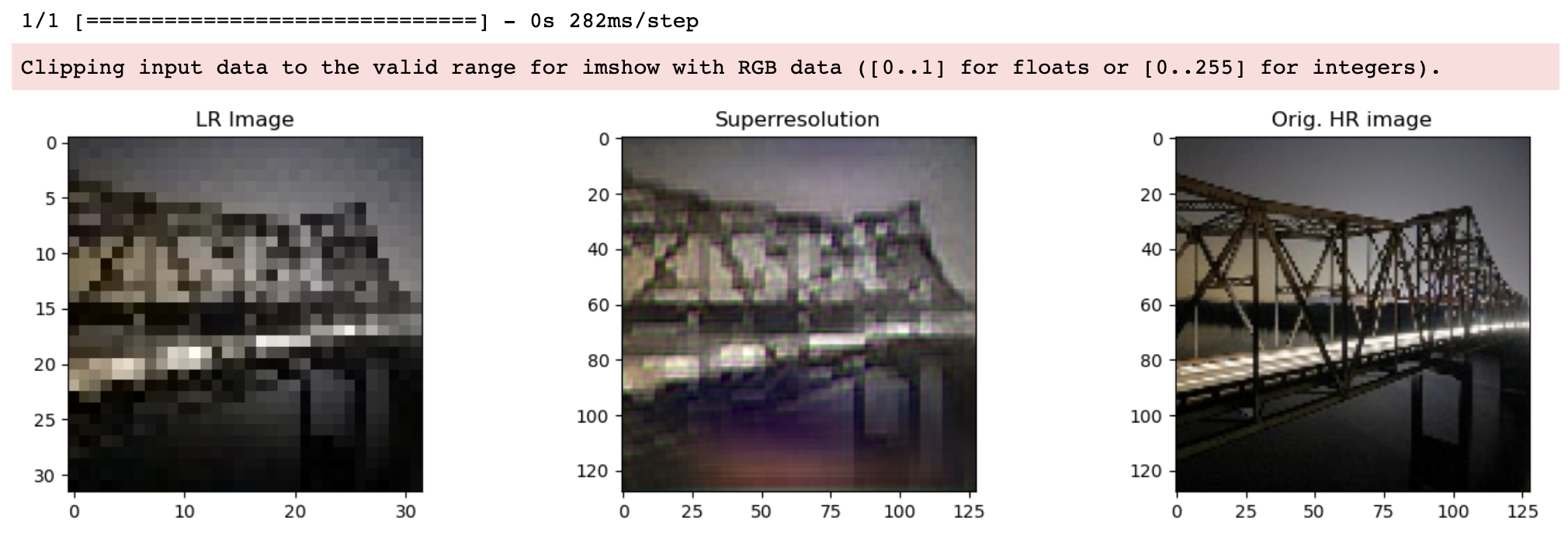
## 4.2.2 SRGAN Results and Analysis

Model after being trained on 700 train sample data with 10 epochs generated following results for randomly selected test sample.

The output is assessed on subjective method assessment based on pair wise similarity judgement.



[Figure 5: Original, degraded and SRCNN generated image from test dataset]



[Figure 6: Original, degraded and SRCNN generated image from test dataset]

# Conclusion

In conclusion, our project demonstrated working capability of SRCNN and SRGAN deep learning architecture and was able to successfully create a model.

The output generated from SRCNN was super resolution image with improved parameters of MSE, PSNR and SSIM.

The output generated from SRGAN is achieved by different combination of batch size, early stopping mechanism and learning rate and epochs. while the model ran for significant duration of time with limited sample size. Model could have achieved the better results if it were to be trained on large sample size with increased no of epochs.

The project is gradually producing improved resolution of images from Low resolution images with limited train and test dataset and epochs on train data.

# future work

* Improving results of the models by training the models on more time and larger datasets.
* Working on different deep learning algorithms such as FSRCNN, ESPCN, RDN, RDN and ESRGAN according to their architecture.
* Build a Web app with stable server that can handle the models and predict fast super resolution images.
* Modification of the architecture used in building the model so that we can achieve higher accuracy and less information loss
* Working on video Super Resolution.

##### References

Top of Form

1. Dong, C., Loy, C. C., He, K., & Tang, X. (2016). Image super-resolution using deep convolutional networks.
2. Ledig, C., Theis, L., Huszár, F., Caballero, J., Cunningham, A., Acosta, A., ... & Wang, Z. (2017). Photo-realistic single image super-resolution using a generative adversarial network.
3. Kim, J., Kwon Lee, J., & Mu Lee, K. (2016). Accurate image super-resolution using very deep convolutional networks.
4. Lim, B., Son, S., Kim, H., Nah, S., & Lee, K. M. (2017). Enhanced deep residual networks for single image super-resolution.
5. Zhang, K., Zuo, W., Chen, Y., Meng, D., & Zhang, L. (2018). Beyond a Gaussian denoiser: Residual learning of deep CNN for image denoising
6. Wang, X., Yu, K., Wu, S., Gu, J., Liu, Y., & Dong, C. (2018). ESRGAN: Enhanced super-resolution generative adversarial networks.
7. Shi, W., Caballero, J., Huszár, F., Totz, J., Aitken, A. P., Bishop, R., ... & Wang, Z. (2016). Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network.
8. Dong, C. C. "Image Super-Resolution Using Deep Convolutional Networks." Multimedia Lab, The Chinese University of HongKong. http://mmlab.ie.cuhk.edu.hk/projects/SRCNN.html
9. Geusebroek, J.M., van den Boomgaard, R., and Smeulders, A.W.M. "MIR Flickr Retrieval Evaluation Downloads." Leiden Institute of Advanced Computer Science (LIACS), Leiden University. http://press.liacs.nl/mirflickr/mirdownload.html
10. “Figure 1: System Architecture Diagram 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, no. 2, pp. 295-307, Feb. 2016.
11. " Figure 2: C. Ledig, L. Theis, F. Huszár, J. Caballero, A. Cunningham, A. Acosta, A. P. Aitken, A. Tejani, J. Totz, Z. Wang, and W. Shi, "Photo-Realistic Single Image Super-Resolution Using a Generative Adversarial Network," in Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 2017.
12. " Figure 3: SRCNN results – DIV2K dataset"
13. " Figure 4: SRCNN results – Custom dataset"
14. “Figure 5: Degraded, Super resolution and SRGAN image from test dataset”
15. “Figure 6: Degraded, Super resolution and SRGAN image from test dataset”