Single Image Super Resolution using two-stage Neural Network architecture.

Rohith Reddy Mada
Hanuma Sashank Samudrala
Ashish Athimamula

OBJECTIVE

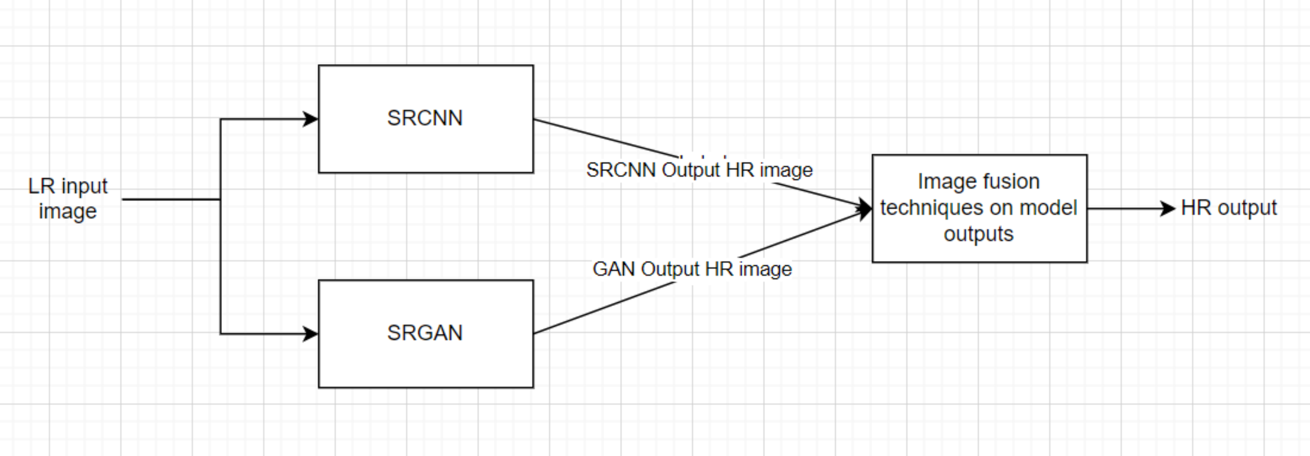
The objective of this project is to develop an effective neural network solution for enhancing the resolution of low-resolution images. The project aims to create a two-stage system that leverages the advantages of both model outputs to generate a final fused image output.

DATASET DESCRIPTION and PREPROCESSING

●The Huggingface's DIV2K dataset (eugenesiow/Div2k) that we are using has 800 images in the train split. Each of them has two image urls fields for both low resolution and high resolution images.

● We used SET5 images owing to its smaller size to test and evaluate the project

PROPOSED APPROACH



After going through some of the existing Single Image Super Resolution methods, we came to the conclusion that a two-stage system(generating SR images using two SISR models and fuse these outputs to produce final SR image) can improve the image resolution to furthermore, that leverages the advantages of both model outputs to generate a final fused image output.

Models Used: Super Resolution Generative Adversarial Network (SRGAN) and

Super Resolution Convolutional Neural Network (SRCNN)

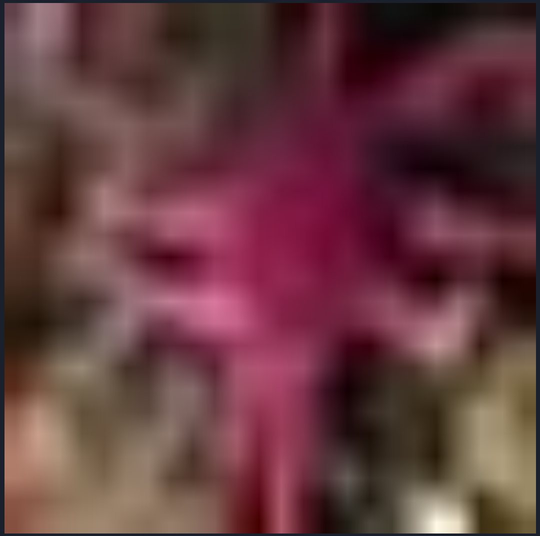
Methodology

For each of the two models:

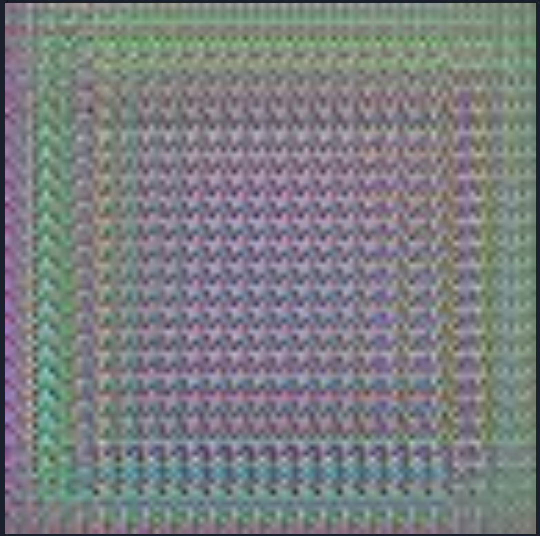
1. Made a custom preprocessor to resize our input images to 24x24 to train. This was done because of limited RAM for training.
2. Defined model architecture and data loader. For each batch we had to check the maximum dimensions and add padding if necessary.
3. Set our hyperparameters, define our optimizers and loss functions.
4. Train and evaluate the models
5. Use the fusion techniques decided to fuse both outputs and calculate metrics.

EXPERIMENTS

Our original plan was to train our models from scratch and the result was :



Expected



Result

SRGAN MODEL INPUT OUTPUT AND RESULTS

Input LR Image

SRGAN Output HR Image

Original HR Image

Peak Signal-to-Noise Ratio(PSNR): 31.23 (32.05)
Mean Squared Error(MSE): 146.65
Structural Similarity Index(SSIM): 0.89 (0.9019)

Note: The presented outcomes were achieved by integrating pretrained weights into the model. The performance of the model we attempted to train independently (on total 1554,883 params) proved to be suboptimal, primarily due to resource limitations, as illustrated in the preceding slides.

SRCNN MODEL INPUT OUTPUT AND RESULTS

Input LR Image

SRCNN Output HR Image

Original HR Image

Peak Signal-to-Noise Ratio(PSNR): 32.98
Mean Squared Error(MSE): 98.16
Structural Similarity Index(SSIM): 0.901

Note: The presented outcomes were achieved by integrating pretrained weights into the model. The performance of the model we attempted to train independently(on total 85,889 params) proved to be suboptimal, primarily due to resource limitations, as illustrated in the preceding slides.

Image Fusion

Image fusion: Merging multiple images into a single composite image to generate a more informative, comprehensive, or enhanced representation.

SRCNN output + GAN output are inputs

Fused output of 2 images

Experiment 1

Pixel Average Fusion

PSNR, MSE, SSIM

[32.8574, 101.0335, 0.9061]

Experiment 2

Laplacian Pyramid

Fusion

Gaussian + Laplacian pyramids

PSNR, MSE, SSIM

[32.7844, 102.7461, 0.9069]

Experiment 3

Principal Component

Analysis

Merge principal components

PSNR, MSE, SSIM

[6.6770, 41928.2314, 0.2706]

Experiment 4

Feature Level Fusion

Edges as features

PSNR, MSE, SSIM

[32.8163, 101.9921, 0.9007]

Experiment 5

Region Wise Fusion

4 rows 4 columns as a region

PSNR, MSE, SSIM

[32.8574, 101.0335, 0.9061]

Experiment 6

Guided Filter Fusion

Getting filters for both images

PSNR, MSE, SSIM

[27.1984, 371.8465, 0.7293]

Experiment 7

Intensity Hue Saturation

Fusion

Maximum intensity

PSNR, MSE, SSIM

[33.0614, 96.3976, 0.9005]

Experiment 8

Discrete Cosine Transform

Fusion

Dct channels in frequency domain

PSNR, MSE, SSIM

[32.6559, 105.8318, 0.9067]

Experiment 9

Select Better Pixel

Fusion

Getting better pixel of both images

PSNR, MSE, SSIM

[32.6807, 105.2278, 0.8953]

RESULTS

We have observed that Intensity Hue Saturation Image fusion technique resulted with better PSNR, MSE and SSIM values compared with SRGAN and SRCNN output Values

Among the two model outputs SRCNN output image is having better values as [32.98, 98.16, 0.901] for PSNR,MSE and SSIM respectively whereas the fusion output image is having values as [33.06, 96.3976, 0.9005]

Note: There is differences of [0.08, 1.77,-0.0095]

LIMITATIONS AND FUTURE SCOPE

1. Our objective was to build the model from scratch but due to limited resources, high hyperparam sensitivity of models, we are unable to train our models well.
2. We never changed the architecture of our models but we could update it in a way like generator could be a srcnn model and that could produce a better output.
3. Training two models is time consuming and we will need to find an approach to make them learn together and infer from each others losses
4. Find other fusion techniques to improve the final result.

THANK YOU