Single Image Super Resolution using two-stage Neural Network architecture

Ashish Athimamula
ka42455@umbc.edu

Rohith Reddy Mada
rohithm4@umbc.edu

Hanuma Sashank Samudrala
hanumas1@umbc.edu

Abstract

This project endeavors to develop a robust neural network solution aimed at enhancing the resolution of low-resolution images. The primary objective is to establish a two-stage system that strategically leverages the strengths of both the SRCNN and SRGAN models, culminating in the generation of a final fused image output. The overarching methodology involves employing nine fusion techniques to merge the super-resolved output images generated by the aforementioned models. This fusion approach seeks to surpass the individual outputs of each model, with the ultimate goal of achieving a markedly improved resolution outcome. Notably, our results showcase that the Intensity Hue Saturation Image fusion technique consistently outperforms the PSNR, MSE, and SSIM metrics when compared with the output values obtained from both the SRGAN and SRCNN models. These findings underscore the efficacy of our fusion strategy in capitalizing on the unique strengths of each model, thereby pushing the boundaries of image resolution enhancement.The code used for this project can be found here: https://github.com/rohithmada00/SRGAN-SRCNN-FUSION

1. Introduction

Image Super-Resolution (SR) is an area of computer vision centered on restoring high-resolution (HR) images from low-resolution (LR) counterparts. This area is critical in a wide range of real-world applications, particularly medical imaging [2] and security [5].

While numerous classical methods have made significant progress in this domain, recent advances have moved the field toward the integration of deep learning techniques, which proved especially useful in addressing the complexity associated with SR. With the introduction of convolutional neural networks (CNNs), models such as SRCNN (Super-Resolution Convolutional Neural Network)[7] and SRGAN (Super-Resolution Generative Adversarial Network) [4] have emerged, demonstrating remarkable capabilities in generating realistic and high-quality HR images from LR inputs.

Despite these developments, the SR problem remains challenging, requiring further research into novel solutions. In this perspective, the combination of SRGAN and SRCNN outputs via nine separate fusion algorithms emerges as a promising area worth further exploration. This paper focuses further into the approach and gives experimental findings for these fusion methods. The primary objective is to contribute meaningfully to the rapidly growing field of single image super-resolution.

2. Related work

Many deep learning neural network models have been developed in recent years to address single image super-resolution.SRCNN model has gained significant traction in the field of image super-resolution. Yujing [7] Song and their colleagues have made significant contributions to the advancement of the SRCNN model for this purpose. They investigated SRCNN advancements by utilizing a technique known as Generative Adversarial Network (GAN) [4], which is well-known for its ability to incorporate detailed texture characteristics into high-resolution outputs.

Dong et al.[1] pioneered the first convolutional neural network (CNN) picture super-resolution technique, known as the Super-Resolution Convolutional Neural Network (SRCNN). As a result, a number of researchers have devoted a great deal of time and effort to the creation of CNN-based algorithms, gaining inspiration from the core notions provided by SRCNN.

Kaur, H, team [3] applied various image fusion techniques, including Simple Average, Laplacian Pyramid Fusion, Guided Filtering, Hue Intensity Saturation, Discrete Cosine Transform, and Maximum Technique. These techniques were utilized across various applications for image fusion.

Sahu and team [6] conducted research on an image fusion algorithm that combines Discrete Wavelet Transform (DWT) and Principal Component Analysis (PCA) with morphological processing. This approach enhances the quality of image fusion and holds promise as a prospective research trend in the field.

3. Methodology

Our approach involves the fusion of super-resolved output images generated by the SRCNN and SRGAN models, utilizing nine fusion techniques. The aim is to enhance the resolution of the resulting image beyond the individual outputs of each model. It is anticipated that the fused image will capitalize on the strengths of both models, thereby yielding an improved resolution outcome.

We use HR images given in the dataset to compute losses and train both the models. We followed the same architectures to design the model. Each image is a tensor with dimensions (width x height x num channels). Objective in training is to determine weights that can produce a (2*width x 2*height x num channels) image that's as close as to ground truth.

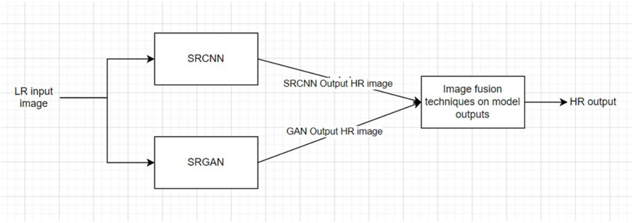


Figure 1. Overall design of the model.

3.1. Architecture of SRCNN

In the context of Super-Resolution Convolutional Neural Network (SRCNN) model algorithm, the procedure commences with the intake of a low-resolution input image, with variable dimensions. Subsequently, a bicubic up sampling operation is employed on the initial low-resolution image. Following this, a convolutional layer is applied, featuring a 9x9 kernel, serving as the initial processing step in the network. Successively, a second convolutional layer is introduced, utilizing a 3x3 kernel for further feature extraction and enhancement. Lastly, a third convolutional layer, incorporating a 5x5 kernel, is applied to generate the ultimate high-resolution output image. The resulting image represents the Super-Resolution Image achieved through the SRCNN model. This algorithmic architecture reflects the progressive application of convolutional operations to transform a given low-resolution input into a high-quality super-resolved image.

3.2. Architecture of SRGAN

Following the structure of the general Generative Adversarial Network (GAN) [14], SRGAN comprises a generator model (G) and a discriminator model (D). Drawing inspiration from the work of Christian Ledig et al. [10], SRGAN's generator is designed, using a combination of a residual network and a sub-pixel convolutional network.

Generator Structure: The core of SRGAN's generator is formed by a series of B residual blocks, each contributing significantly to the model's performance. Each residual block is composed of two convolutional layers with 3x3 kernels, generating 64 feature maps. Following each convolutional layer, a batch-normalization layer and the Parametric Rectified Linear Unit (ParametricReLU) activation function are applied. Notably, SRGAN introduces innovation by incorporating skip connections in each residual block.

Discriminator Structure: Discriminator SRGAN, has 8 convolutional layers with Leaky Rectified Linear Unit (LeakyReLU) activation . The number of channels doubles every second layer. Convolutions are used to lower image resolution during channel doubling. Two dense layers analyze 512 feature maps in the final level, with sigmoid activation for classification probability. In the adversarial training process, this discriminator design, which focuses on hierarchical feature extraction and resolution reduction, leads to effective discrimination between high-quality and produced images.

3.3. Fusion Techniques

We employed a total of 9 methods to experiment on:

3.3.1 Pixel Average Fusion

Pixel Average Fusion involves computing the average pixel values of corresponding pixels from multiple images. We calculated the mean of pixel values at each position to create a new fused image. At each pixel position (x, y), the pixel values from the same position in two input images are averaged to generate the corresponding pixel value in the fused image. This technique is most suited when the input images have relatively similar quality, content, and relevance but it may lack adaptability in handling complex image characteristics or emphasizing important details present in the input images.

3.3.2 Laplacian Pyramid Fusion

Laplacian Pyramid Fusion combines Gaussian and Laplacian pyramids for image fusion. The Gaussian pyramid represents images at different scales, while the Laplacian pyramid contains information about details and edges. Initially, Gaussian pyramids are created from the input images, followed by the generation of corresponding Laplacian pyramids. The levels of these pyramids are fused using blending functions like averaging, and the fused Laplacian pyramid is used to reconstruct the final fused image through inverse pyramid construction using pyrUp.

3.3.3 Principal Component Analysis Fusion(PCA)

Principal Component Analysis reduces data dimensions while retaining essential information. PCA captures significant pixel variations by identifying crucial components. It selects and combines these key components from input images to create a new set, impacting the quality based on the accuracy of the component selection and combination. However, it might not always fully represent all essential image information. We have observed some noise in fusion for our dataset.

3.3.4 Feature Level Fusion

Feature Level Fusion integrates specific features, like edges, from input images using the Canny algorithm. It selectively merges these extracted features to create a fused image, emphasizing crucial edge information. This process begins by identifying and extracting edges as key features from each input image. The final fused image is generated by combining the integrated edge information based on the chosen fusion strategy, enhancing edge details in the output.

3.3.5 Region-Wise Fusion

Region-Wise Fusion merges specific regions in input images to create a cohesive fused image. Our method focuses on 4x4 regions, utilizing an averaging technique for fusion within these areas. This targeted approach selectively integrates information from identified regions, avoiding individual pixel fusion. The resulting fused image is constructed by combining information within these designated regions from the input images.

3.3.6 Guided Filter Fusion

Guided Filter Fusion employs guided filtering techniques to fuse images while preserving edge information and enhancing image quality. It uses a guided filter to combine information from multiple images while emphasizing structures and details present in the guidance image. We fused these two filters of individual images and obtained the fused image.

3.3.7 Intensity Hue Saturation(IHS) Fusion

IHS Fusion separates the input images into intensity (I), hue (H), and saturation (S) components. It combines the I component of one image with the H and S components of another to form the fused image, we have choosen the maximum intensity of two images. It typically involves transforming images from RGB color space to IHS color space, performing fusion on the respective components, and converting back to RGB.

3.3.8 Discrete Cosine Transform(DCT) Fusion

DCT Fusion utilizes the Discrete Cosine Transform to convert images into the frequency domain, where fusion occurs by manipulating DCT coefficients. We applied DCT to transform images, combined DCT coefficients using specific fusion rules, and then performed inverse DCT to reconstruct the fused image. It is suitable for frequency-based fusion, emphasizing certain frequency components and reducing artifacts.

3.3.9 Select Better Pixel Fusion

Select Better Pixel Fusion chooses the best pixel value from corresponding positions in multiple input images based on specific criteria like image quality or relevance. It compares pixel values at corresponding positions in input images and selects the pixel with higher quality, contrast, or relevance to form the fused image. It is simple and effective in retaining high-quality or significant information from input images.

4. Experiment

4.0.1 Dataset

We utilized Huggingface's DIV2K dataset (eugenesiow/Div2k) for training, comprising of 800 images in the train split, each with two image URLs for low resolution image and high resolution images each. Additionally, SET5 images, chosen due to their smaller size, were employed for testing and evaluating the project.

4.0.2 SRGAN

The experiment was performed with a scale factor of 2× between low and high resolution images. We trained the model on AMD Ryzen 7 7735HS Radeon Graphic Processor for 30 Epochs. Each epoch took about 20 minutes. These images are distinct from the testing images.We took a batch size of 1, a single image, so that both the models are able to capture fine details. The code was done in pytorch.Adam optimizer is used with learning rate 10^{-4}.

4.0.3 SRCNN

We've incorporated the pre-trained weights into our model which are obtained from a model which is trained as mentioned below The SRCNN model is trained on the above mentioned dataset with a 2x and 4x scale, employing an 80-20 cross-validation approach. The neural network architecture comprises a total of 85,889 parameters,encompassing both weights and biases. Here is an overview of the hyperparameters employed in the SRCNN model, which includes the use of the ADAM optimizer, a learning rate of $3*10^{-3}$, Rectified Linear Unit (Relu)activation function, Mean Squared Error (MSE) as the loss function, a training duration of 60 epochs, and a batch size of 10.

4.0.4 Evaluation Metrics

Our project underwent rigorous testing and evaluation, focusing on the Peak Signal-to-Noise Ratio (PSNR), Mean Squared Error (MSE), and Structural Similarity Index (SSIM) index values of the fusion outputs. We compared these values with the index values for our respective outputs of our individual models. If a given fusion output demonstrates superior values for any of the three metrics, we consider our objective fulfilled, indicating the effectiveness of the applied fusion technique.

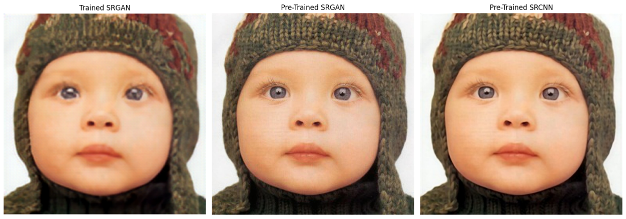
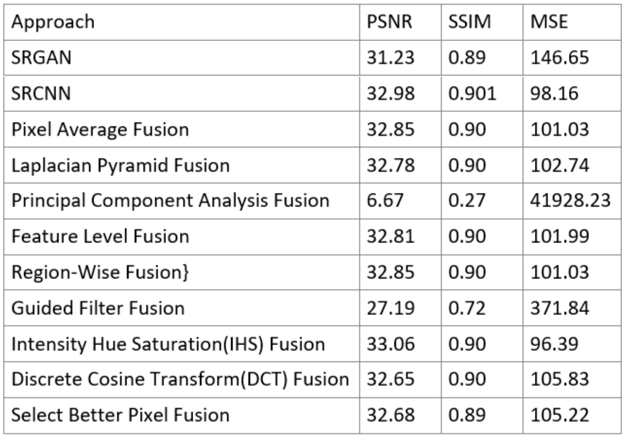


Figure 3. HR images with trained model SRGAN, pretraind SRGAN, pretrained SRCNN



Approach	PSNR	SSIM	MSE
SRGAN	31.23	0.89	146.65
SRCNN	32.98	0.901	98.16
Pixel Average Fusion	32.85	0.90	101.03
Laplacian Pyramid Fusion	32.78	0.90	102.74
Principal Component Analysis Fusion	6.67	0.27	41928.23
Feature Level Fusion	32.81	0.90	101.99
Region-Wise Fusion}	32.85	0.90	101.03
Guided Filter Fusion	27.19	0.72	371.84
Intensity Hue Saturation(IHS) Fusion	33.06	0.90	96.39
Discrete Cosine Transform(DCT) Fusion	32.65	0.90	105.83
Select Better Pixel Fusion	32.68	0.89	105.22

Figure 4. PSNR, SSIM, MSE for all our experiments

5. Conclusion

Our comprehensive analysis has revealed that the Image Fusion technique employing Intensity, Hue, and Saturation has demonstrated superior performance when compared to the output values obtained from both SRGAN and SRCNN models. Specifically, the assessment metrics, including PSNR, MSE, and SSIM, consistently exhibited higher values for the Image Fusion approach.

Upon closer examination of the two model outputs, it was observed that the SRCNN-produced image displayed commendable metrics with values of [32.98, 98.16, 0.901] for PSNR, MSE, and SSIM, respectively. In comparison, the fused output image demonstrated slightly improved results with values of [33.06, 96.3976, 0.9005]. Notably, the differences between these values were marginal, with variances of [0.08, 1.77, -0.0095] for PSNR, MSE, and SSIM.

This nuanced discrepancy in metrics signifies the delicate trade-offs and subtle distinctions in the performance of the models and fusion technique. The marginal improvements observed in the fused output underscore the potential efficacy of combining multiple approaches to achieve optimal results in the context of image enhancement.

6. Limitations

1. In our method, super-resolved output pictures generated by the SRCNN and SRGAN models are fused together utilizing nine different fusion strategies. Beyond the particular outputs of each model, the goal is to improve the resolution of the final image. It is believed that the combined image will take advantage of the advantages of both models, resulting in a better resolution result.
2. Although our models' basic architectures remained unchanged, we looked at a different approach by changing the generator to take advantage of an SRCNN model's capabilities. The objective of this modification was to increase the capacity of the model to generate better results.
3. It may require a while to individually train two models. We are looking into a method where the models work together to learn from each other's errors in order to address this. The goal of this collaborative learning paradigm is to optimize the model's efficiency by streamlining the training process.
4. Alternative fusion procedures have been actively investigated in our search for improved outcomes. These methods go beyond traditional strategies and combine the advantages of various models or components to optimize the end product. We think that by investigating fusion approaches, we can improve upon our current picture enhancing task's outcomes.

7. Individual Contributions

Team's individual member contributions are as below:

1. Rohith: SRGAN Model Implementation
2. Hanuma Sashank: SRCNN model implementation
3. Ashish: Fusing the two obtained output images to get a more accurate image.

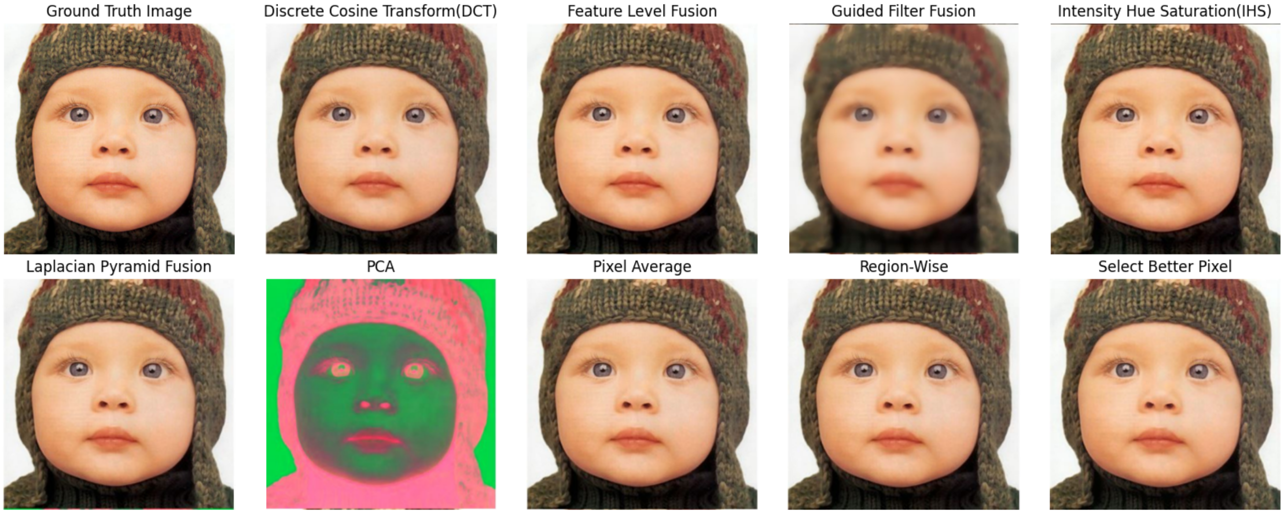


Figure 2. Super resoluted image for all experiments

References

[1] Chao Dong, Chen Change Loy, Kaiming He, and Xiaoou Tang. Learning a deep convolutional network for image super-resolution. In *Computer Vision–ECCV 2014: 13th European Conference, Zurich, Switzerland, September 6-12, 2014, Proceedings, Part IV 13*, pages 184–199. Springer, 2014. 1

[2] Yawen Huang, Ling Shao, and Alejandro F Frangi. Simultaneous super-resolution and cross-modality synthesis of 3d medical images using weakly-supervised joint convolutional sparse coding. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 6070–6079, 2017. 1

[3] Harpreet Kaur, Deepika Koundal, and Virender Kadyan. Image fusion techniques: a survey. *Archives of computational methods in Engineering*, 28:4425–4447, 2021. 1

[4] Christian Ledig, Lucas Theis, Ferenc Huszár, Jose Caballero, Andrew Cunningham, Alejandro Acosta, Andrew Aitken, Alykhan Tejani, Johannes Totz, Zehan Wang, et al. Photo-realistic single image super-resolution using a generative adversarial network. In *Proceedings of the IEEE conference on computer vision and pattern recognition*, pages 4681–4690, 2017. 1

[5] Pejman Rasti, Tonis Uiboupin, Sergio Escalera, and Gholamreza Anbarjafari. Convolutional neural network super resolution for face recognition in surveillance monitoring. In *Articulated Motion and Deformable Objects: 9th International Conference, AMDO 2016, Palma de Mallorca, Spain, July 13-15, 2016, Proceedings 9*, pages 175–184. Springer, 2016. 1

[6] Deepak Kumar Sahu and MP Parsai. Different image fusion techniques–a critical review. *International Journal of Modern Engineering Research (IJMER)*, 2(5):4298–4301, 2012. 1

[7] Yujing Song. Single image super-resolution. *Scholarly Horizons: University of Minnesota, Morris Undergraduate Journal*, 6(1):9, 2019. 1