International Journal of Research Publication and Reviews

Journal homepage: www.ijrpr.com ISSN 2582-7421

Image Resolution Enhancer using Deep Neural Networks

Dr. P. Srihari[1], K. Surya Kiran[2], K. Vishal[3], K. Rajasekhar[4], M. Sai Joshitha[5], G. Durga Nivas[6]

[1]Information Technology, GMR Institute Of Technology Rajam, Andhra Pradesh, India. srihari.p@gmrit.edu.in
[2]Information Technology, GMR Institute Of Technology, Rajam, Andhra Pradesh, India, 20341A1253@gmrit.edu.in
[3]Information Technology, GMR Institute Of Technology, Rajam, Andhra Pradesh, India 20341A1247@gmrit.edu.in
[4]Information Technology, GMR Institute Of Technology, Rajam, Andhra Pradesh, India 20341A1248@gmrit.edu.in
[5]Information Technology, GMR Institute Of Technology, Rajam, Andhra Pradesh, India 20341A1261@gmrit.edu.in
[6]Information Technology, GMR Institute Of Technology, Rajam, Andhra Pradesh, India 20341A1232@gmrit.edu.in

ABSTRACT –

Image super-resolution is a vital image processing technique in computer vision, as it works to enhance the resolution of images significantly. Super resolution has advanced significantly over the past 20 years, in large part because to the use of deep learning techniques. In a variety of industries, including computer graphics, medical imaging, security, space, and satellite, image super resolution is crucial. The main goal of this project is to boost and increase an image's resolution so that it can be useful in the aforementioned sectors. With an emphasis on contrasting the performance of the CNN models to that of a previously established method, sparse coding, we intend to leverage pioneers of convolutional neural networks to improve the resolution of the image in this. The model utilizes a neural network architecture that incorporates convolutional layers to learn and extract features from the low-resolution input image, and then uses these features to reconstruct a high-resolution output image. The sparse coding approach, on the other hand, utilizes a sparse representation of the image to reconstruct the high-resolution output. The findings of this study will provide light on the efficacy of CNN models for improving image resolution and their potential as a replacement for conventional sparse coding techniques.

Index Terms -: Convolutional Neural Networks (CNN), Sparse Coding, Image super-resolution, Computer Graphics, Medical Imaging, Security

I. Introduction

Image processing is a term that describes the methods used to manipulate, analyze, and interpret digital pictures. It includes using mathematical techniques to a picture in order to analyze it and improve its quality, increase feature extraction, or analyze its data. Techniques used often in image processing include filtering, detection of edges, segmentation, and extraction of features. Satellite imagery, computer vision systems, and medical imaging are just a few of the many uses for image processing. Image super-resolution, or the technique of recovering high resolution pictures from low resolution photos, is a prominent family of image processing methods in computer vision and image processing. It may be used for safety, monitoring in medical imaging and security surveillance among other useful applications. Along with improving picture perception quality, it aids in a number of computer vision tasks. Since there are frequently many HR photographs that correlate to a single LR image, the problem is essentially ill-posed and highly challenging to solve. Many super-resolution technologies, such as modern learning-based approaches and early classical approaches, are proposed as remedies for this problem. Traditional approaches include those that rely on regularization and interpolation. Recently, to solve the image SR problem, numerous convolutional neural network-based techniques have been proposed. The main metrics for image resolution include PSNR and SSIM. Peak-to-noise ratio is used to calculate noise in image with respective another image. In our case we will have clean image, low resolution image and high-resolution image. Structural Similarity index is a metric which is used to compare the structure similarity between two images.

II. Related Work

Z. X et al. investigated how sparse representation and color restrictions could improve the resolution of single-color images. In a number of applications, including high-definition digital television, mobile internet, remote sensing, medical imaging, cultural relic preservation, and display technology, they underlined the importance of enhancing image quality. The authors presented a unique method for reconstructing high-resolution color images that combines color channel limitations and L2/3 sparse regularization. In their method, a low-resolution color image is first converted from RGB to YCbCr before focusing on rebuilding the brightness channel using the L2/3 sparse regularization model. The process eliminates any unwanted color artifacts using boundary similarity limits among the RGB color channels, greatly enhancing the quality of the reconstructed color image. Overall, by combining sparse representation methods with color restrictions, the suggested method offers a potential direction for furthering study in the area of improving single-color image resolution.[1]

He et al. developed an innovative technique for single image super-resolution (SISR) to convert low-resolution (LR) images into high-resolution (HR) outputs. Their approach uses an end-to-end trainable unfolding network made up of a convolutional sparse coding (CSC) model and a super-resolution (SR) model to merge deep learning and prior-based approaches. A reweighted method and a convolutional weighted iterative soft thresholding algorithm are two improvements included in the CSC model. The foundation of the SR model is learning weighted convolutional sparse coding with channel attention. The performance of the suggested system was compared to current SISR techniques through thorough experimentation. The inclusion of global residual learning is one of the authors' recommendations for additional enhancements. In conclusion, their method shows potential gains in the field[2].

Kim et al. presented two thin neural networks, the SR-ILLNN and SR-SLNN, for single image super-resolution (SISR). These networks were designed to find a balance between network complexity and SR accuracy (as determined by PSNR and SSIM). The first network From low-resolution and interpolated low-resolution pictures, SR-ILLNN learns feature maps. which enables it to comprehend image content more deeply and, as a result, provide better super-resolution outcome The second network, SR-SLNN, on the other hand, significantly reduces network complexity by utilizing only the low-resolution feature maps from SR-ILLNN. As a result, SR-SLNN is even more lightweight, albeit at the expense of slightly reduced super-resolution quality. The authors tested both networks on benchmark datasets (Set5, Set14, and BSD100), showing that their suggested models outperformed other cutting-edge lightweight SISR techniques by an equal or greater margin. These models stand out because they have faster inference times and much fewer parameters[3].

Maral, B. C. et al. provide an in-depth analysis of the single image super-resolution (SISR) discipline, exploring its various approaches, difficulties, and applications. They divide SISR methods into four primary categories: spatial domain, frequency domain, learning-based, and hybrid methods. This division offers an invaluable foundation for comprehending the various SISR methodologies. The evaluation of SISR techniques is also included in this study, along with metrics for assessing the caliber of reconstructed images. This discussion clarifies how the performance of SISR approaches is rated and contrasted. The authors also highlight the several applications of SISR, such as security, medical imaging, and image augmentation, emphasizing its potential benefits in a variety of fields. The paper concludes by discussing the future prospects of SISR research and identifying issues that must be tackled in order to improve the effectiveness of SISR techniques [4].

L. Jiang et al. introduced SADNN, a brand-new deep neural network created for single image super-resolution (SISR). Convolutional layers, pooling layers, and self-attention layers are all included in the SADNN design. In order for SADNN to reconstruct high-resolution images with improved texture details and hierarchical image structures, it is crucial to capture global relationships across various image positions. The network uses a loss function that combines the pixel-wise mean square error loss and perceptual loss, two separate components. In the former, the difference between the original high-resolution image and the reconstructed one is measured; in the latter, the distinction between the features of the original high-resolution image and the recreated one is quantified. According to the authors[5], SADNN could benefit from more improvements if a larger dataset and a more powerful optimizer were used.

William Symolon et al. focused on improving the resolution of low-resolution satellite photos taken by CubeSats, which often have constrained imaging capabilities. The researchers used Convolutional Neural Networks (CNN), a deep learning technique, to upscale these images and improve their sharpness and utility in order to solve this problem. In particular, they used the Efficient Sub-Pixel Convolutional Neural Network (ESPCN), which was created for effective low-resolution picture improvement, notably in super-resolution tasks. In their investigation, they used measures like PSNR and SSIM to assess the reconstructed images' quality. The results of this study show that CNN-based super-resolution is a technology that has tremendous promise for enhancing the resolution of CubeSat imaging, with possible applications in numerous fields[6].

Ahmed et al. developed a revolutionary technique called convolutional sparse coding (CSC) to improve multimodal images. They contend that traditional image enhancing techniques frequently fail to adequately capture the complex interactions between the various modalities included within an image. They suggest a novel approach that successfully makes use of these connections to overcome this constraint. Their method starts by employing a tensor-based approach to break down the input multimodal image into its individual modalities. The learning of convolutional filters that successfully capture the image's underlying structures is then made possible by applying CSC to each modality separately. The final improved image is produced by combining the upgraded modalities. A number of objective and arbitrary image quality metrics show that their solution outperforms existing ones in experimental evaluations on a variety of publicly accessible datasets. Their study essentially presents a persuasive and promising method for improving multimodal photos using CSC[7].

Ilesanmi et al. showed that Convolutional Neural Networks (CNNs) have been specifically utilized in recent advancements in picture denoising. They underline the growing significance of CNNs in the field of image denoising and credit their success to their ability to effectively learn complex image representations. The research presents a thorough investigation of different CNN-based methods for picture denoising, including techniques like fully convolutional networks, residual networks, and generative adversarial networks. The authors also thoroughly compare and contrast different approaches across various datasets, illuminating their individual advantages and disadvantages. The Berkeley Segmentation Dataset (BSD), Set12, and Kodak are well-known benchmark datasets that are used in this assessment. Performance is measured using a variety of measures, incorporating visual quality evaluation, structural similarity index, and peak signal-to-noise ratio .(VQA)[8].

Li, X., et al. reviewed contemporary deep learning techniques for real-time image super-resolution. They made a distinction between supervised Convolutional Neural Network (CNN) techniques and GAN-based strategies. By significantly relying on predetermined assumptions, supervised CNN algorithms seek to directly learn the mapping between low-resolution (LR) and high-resolution (HR) pictures. GAN-based networks, on the other hand, provide greater flexibility and display promising performance since they incorporate unsupervised training. The study claims that in terms of reconstruction outcomes, GAN-based super-resolution approaches have a significant advantage over their CNN-based equivalents. Even with scant labeled data, GANs, which were originally developed for image production, excel at producing photo-realistic HR images by capturing a variety of super-resolution patterns. A discriminator network that has been trained to distinguish between generated SR images and original natural scene images is used to do this. adversarial loss ensures that SRGAN (Super-Resolution GAN) can generate high-quality and photorealistic SR images[9].

C. Tian et al. created the Coarse-to-Fine Super-Resolution Convolutional Neural Network (CFSRCNN) to enhance low-resolution photos and ultimately restore high-resolution versions. The CFSRCNN architecture's feature extraction blocks, an enhancement block, a construction block, and a feature refinement block combine to create a powerful SR model. By using modular blocks, this model successfully mixes low-resolution and high-resolution features, reducing training instability and performance deterioration brought on by up-sampling procedures. The authors also examined several super-resolution techniques, including SRCNN, VDSR, Residual Dense Network (RDN), and Information Distillation Network (IDN). Each approach deals with a particular issue, such as enhancing training speed, lowering parameter count, and successfully recovering fine details in HR photos using hierarchical features. With their proposed 46-layer CFSRCNN model, which consists of stacked Feature Extraction Blocks (FEBs), an Enhancement Block (EB), a Construction Block (CB), and a Feature Refinement Block (FRB), they successfully transform low-resolution input (ILR) into high-resolution output (ISR)[10].

III. Methodology

The goal of this research is to create a super resolution model based on deep learning that can boost an image's resolution while maintaining its fine details. The desired result is a deep learning resolution image as based super resolution model that can handle images with significant fluctuations in scale and translation and can take a low input and produce a high-resolution image as output.

Autoencoder

For unsupervised learning, dimensionality reduction, and generative modelling, a neural network design known as an autoencoder is employed. In Fig. 1, an encoder and a decoder are depicted as the two primary parts of the system. A lower-dimensional representation known as a bottleneck or latent space is created by the encoder by compressing an input. Following that, the decoder uses the lower-dimensional representation to try and recreate the original input. In order to reduce the reconstruction error between an autoencoder's reconstructed output and original input, the autoencoder must be trained. Backpropagation, a method for accomplishing this, involves changing the network's weights. Consequently, the autoencoder learns to keep the crucial information in the input while removing the noise and the unnecessary information.

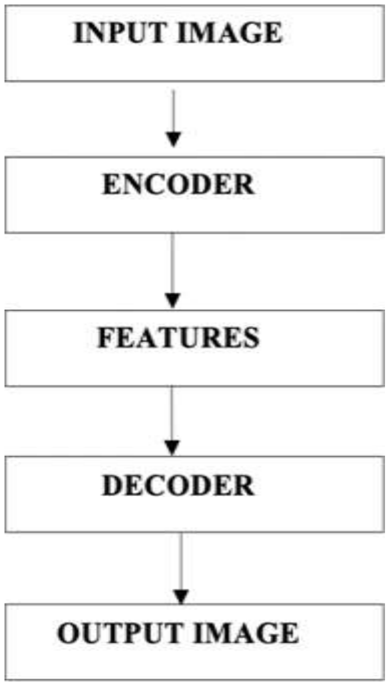

Fig. 1: Autoencoder Block Diagram

By mapping the input data into a lower-dimensional space, autoencoders can reduce the dimensionality of the data while maintaining its key characteristics. They can also be used for generative modelling, which involves encoding the bottleneck representation as random noise and decoding it into a high-dimensional image. Finally, autoencoders can be used to create new samples that closely resemble the original data.

There are four types of autoencoders, namely:

1. Vanilla autoencoder

2. Variational autoencoder
3. Denoising autoencoder
4. Convolutional autoencoder

Vanilla autoencoder is a simple sort of autoencoder that reconstructs the original input data as nearly as feasible. The vanilla autoencoder's purpose is to learn a compact representation of the input data that may be utilized for applications like dimensionality reduction, generative modelling, and denoising. An autoencoder intended to reduce noise from input data is a form of autoencoder. The denoising autoencoder is fed noisy versions of the input data during training, with the goal of reconstructing the original, clean input data. Variational autoencoders are a type of autoencoder that introduces a probabilistic interpretation to the bottleneck representation, allowing for sampling from the latent space and generating new data samples. Each form of autoencoder has advantages and disadvantages, and the ideal type to utilize depends on the task and the data. Convolutional autoencoders are the best for image super-resolution because they are created exclusively for image data and include layers that can capture spatial correlations in the input data, making them ideally suited for tasks like picture production and super-resolution.

Convolutional Autoencoder:

For the purpose of image compression and image super resolution a neural network model called convolutional autoencoder. Convolutional layers are ideally used to handle picture data and these layers are attached with the encoder and decoder architecture of a CAE. The encoder part of the model basically consists of a number of convolutional layers and pooling layers in order to decrease the spatial dimensions of the given input image. The decoder part has the ability to reduce the number of feature maps while increasing the spatial dimensions of the image which is encoded back to its actual size. By reducing the difference between the input and output of the decoder, the model can be trained to reconstruct an input image.

The architecture of the convolutional autoencoder is shown in the below fig. 2

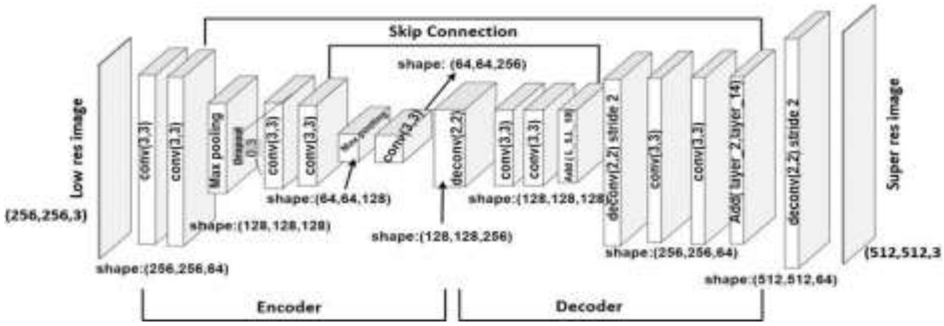

Fig. 2: Architecture of convolutional autoencoder

There are five layers in convolutional autoencoder:

1. Convolutional layers
2. Max-pooling layers
3. Drop out layers
4. Add layers
5. Up-sampling layers

Convolutional layers

A form of neural network layer used in computer vision applications to analyze visual images is called a convolutional 2D (Conv2D) layer. Because it uses a mathematical technique called convolution to combine the values of adjacent pixels in a picture to create a new set of values, the layer is known as a "convolutional" layer. An image or any 2D input that may be represented as a 3D tensor is convolutioned in the case of a 2D convolutional layer. Each filter that is applied to the input picture by the layer is also represented as a 3D tensor. The convolution procedure is carried out at each place as the filters move over the input picture. A feature map, a 3D tensor, is the end product. The depth of the feature map is dependent on how many filters are used during the convolution procedure. In addition to the Conv2D layer, there are other types of convolutional layers that can be used in a computer vision model, such as the convolutional 1D (Conv1D) layer, which is used for analyzing 1D signals like audio, and the transposed convolutional 2D (Conv2DTranspose) layer, which is used for up sampling feature maps. The Conv2D layer has a number of parameters that can be set when it is instantiated. In general, the Conv2D layer's primary job is to 23 extract characteristics from the input picture, such edges, textures, and forms. As a result, the model will be better able to recognize key elements in the image and anticipate its contents.

Max-pooling layers

A convolutional neural network (CNN) has a pooling operation type called max pooling. By pooling, the number of parameters and computation in the network are reduced as well as the spatial dimensions of the convolved feature maps. In particular, max pooling employs the highest value from each patch of the input feature map as the output of the pooling process. After one or more convolutional layers, a max pooling layer is commonly implemented, and the pooling is done separately for each channel of the input feature map. A tiny kernel, such 2x2 or 3x3, with a stride the same size as the kernel are commonly used for pooling. As a result, the spatial dimension of the feature map is decreased by a 24-kernel size factor. The resulting feature map will have dimensions (H/2)x(W/2)xC, for instance, if the input feature map has dimensions HxWxC and the max pooling is carried out using a 2x2 kernel and a stride of 2. The network becomes more effective as a result of the reduction in the number of spatial dimensions and the associated parameters and calculation. The additional benefit of offering some sort of translation invariance is another advantage of max pooling.

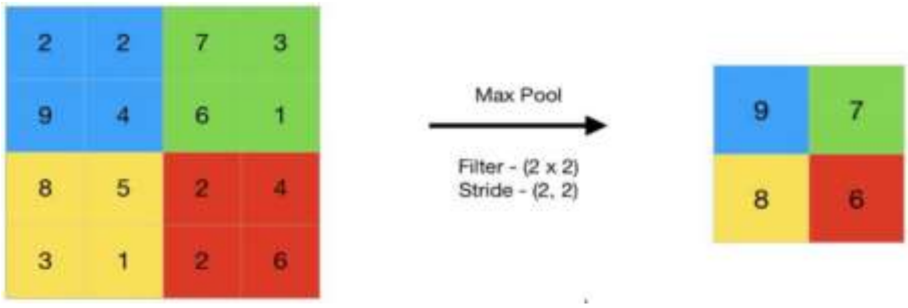

Fig. 3 Converting 4*4 pixel matrix to 2*2 pixel matrix

Drop out layers

By randomly removing (i.e., setting to zero) a specific proportion of neurons during training, the regularization approach known as "dropout" is utilized in deep learning to minimize overfitting. The dropout layer of a neural network is often 25 positioned in between the fully connected layers. When a dropout layer is added to a neural network, it randomly changes a particular proportion of the neurons' output to zero on each forward pass during training. The dropout rate is a percentage that is normally between 0.2 and 0.5. As a result, the model's complexity is decreased by lowering the number of neurons that contribute to the final output. When many sub-models are trained concurrently and their outputs are merged during inference, dropout may be thought of as a type of model averaging. The network is efficiently trained using numerous thinned versions with distinct neurons being removed at each forward pass using the dropout layer. The whole network is used during testing, but each neuron's output is scaled by 1/(1-p), where p is the dropout rate.

The fig. 4 shows the neuron structure before and after applying the dropout layer.

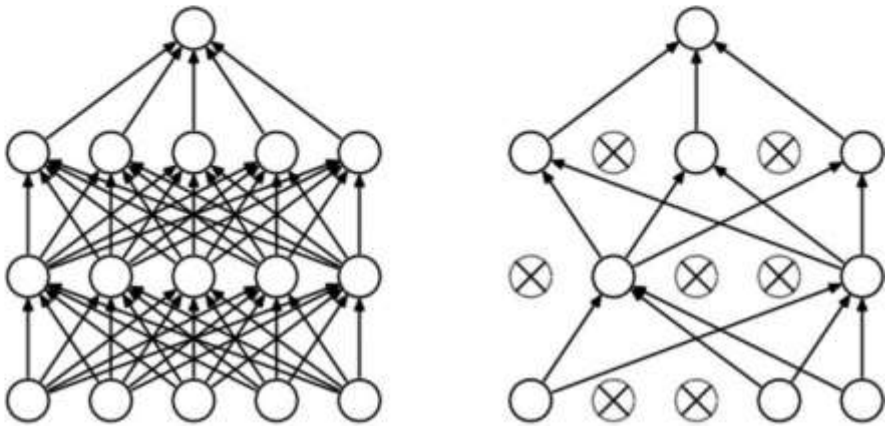

Fig.4: Before and after applying dropout layer

The increasing number of neurons during testing is balanced in part by this scaling. Dropout is very helpful in deep neural networks since it makes it more difficult for any one neuron to become overly significant for the output, hence lowering the risk of overfitting

Add layers

The add layers are also known as element wise addition layers. It is a straightforward layer in a neural network that adds two input tensors element-by-26 element. It is important that output tensor should have same shape as input tensor. It is add can layers outputs to other various layers output or other branches. It can be used to aggregate the results of different neural network branches, such as those found in a Siamese network . It can also be used as instance to integrate the results of many convolutional layers or the results of a convolutional layer with an attention layer.

De-convolutional or Up-sampling layers

Up-sampling is a method used in computer vision and image processing to enhance the spatial resolution of an image or feature map. The up-sampling layer in deep learning is used to boost a convolutional neural network's (CNN) feature map's spatial resolution. It is frequently employed in the network's decoder section, where the aim is to improve the feature maps' spatial resolution so that they may be utilized to produce high-resolution output images. Typical applications include object identification, super-resolution, and picture segmentation.

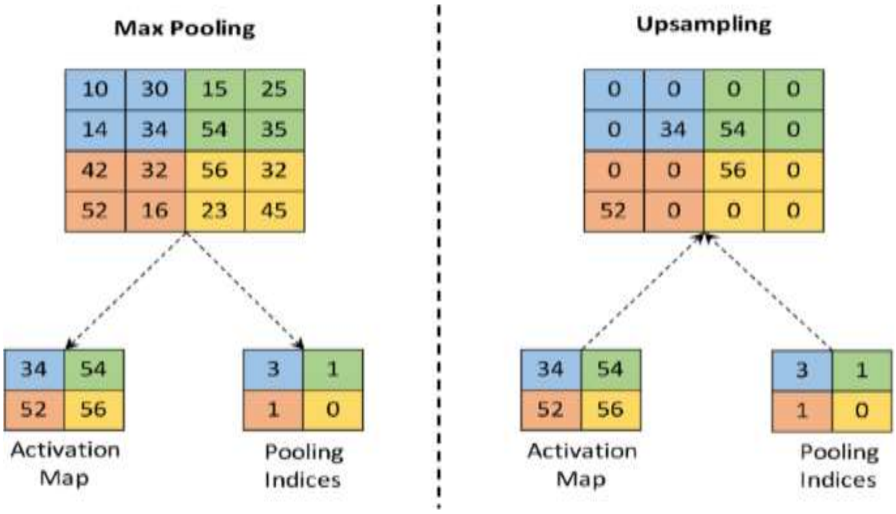

Fig. 5 Working of max pooling and up-sampling layers

Metrics used for evaluation:

Mainly two parameters are used to compare our predicted image with our original image. We will be employing peak-to-noise ratio and structural similarity index measure to compare the outcomes .

Peak-to-noise ratio

Peak-to-noise ratio is used to calculate noise in image with respective another image. In our case we will have clean image, low resolution image and high-resolution image. First, we need to apply PSNR for clean image and low-resolution image then we need to apply PSNR for clean image and high-resolution image. We can say that more the PSNR value less the noise. So, if PSNR of our output image is greater then PSNR of low-resolution image we can say the high-resolution output is achieved.

Structural Similarity Index

Structural Similarity index is a metric which is used to compare the structure similarity between two images. The values of SSIM always lies between -1 to 1. If the value is near to -1 then we can say that there is less similarity in structure between two images. If the value is near to 1 then we can say that there is more similarity in structure between two images. In our case we will try to achieve SSIM value of output image greater than SSIM value of low-resolution image.

IV. RESULTS AND DISCUSSIONS

There 3 models we trained based on convolutional auto-encoder technique. The three variations of auto-encoder are: First Model (Model_1): - One Convolutional layer with 64 Filters, Second Model (Model_2): - Two Convolutional layer with 32 Filters, Third Model (Model_3): - Two Convolutional layer with 64 Filters. The image fig. 6 is given as input to the models and the output of the models is shown in fig.7. The

Fig.6 : Input Image to Model

Fig.7 : Output image of Model

We can observe all the results of all variations of auto encoder models in table 1. Here we are using 2 images to understand which models performance is best. For the best comparison metrics we use PSNR and SSIM.

	Input image 1		Input image 2	
	PSNR	SSIM	PSNR	SSIM
Low resolution image	27.258	0.882	30.293	0.852
Model_1 output	29.360	0.911	32.053	0.881
Model_2 Output	29.74	0.909	32.085	0.889
Model_3 Output	29.783	0.919	32.546	0.897

Table 1: Comparison of Models

Firstly, it's important to understand what PSNR and SSIM represent. PSNR is a 37 metric that measures the difference between two images by computing the ratio of the maximum possible pixel intensity to the mean squared error (MSE) between the two images. A higher PSNR value indicates that the images are more similar to each other. On the other hand, SSIM is a metric that quantifies the structural similarity between two images based on luminance, contrast, and structure. A higher SSIM value indicates that the images are more visually similar to each other. In the table, we have four rows corresponding to different images. The first row corresponds to a low-resolution image, which is used as the input for the three models.

Out of all the three models, the third model outperformed with its best output indicating that it produces more accurate output than the other models in terms of image super resolution. In summary, the table shows that Model_3 is the most effective of the three models in improving the resolution and quality of the input images. This information can be used to inform decisions about which model to use in a given application where high-quality image reconstruction is required.

V. Conclusion

Convolutional Autoencoder architecture has been proposed to restore a low-resolution image to a better quality image with higher resolution. Our study provides insight into the effectiveness of Convolutional Autoencoder for image super resolution and its potential as an alternative to traditional sparse coding methods. The Convolutional autoencoder model requires 2 convolution layers after every max-pooling or up-sampling layer. It provides a more cleaner image with higher resolution. To achieve best reconstructed image with high PSNR ratio and high SSIM value even on low performance centric device. By adding more convolutional layers and changing the filters we can get more accurate results i.e, high resolution in output of the images.

References

[1] Z. Xu, Q. Ma and F. Yuan, "Single color image super-resolution using sparse representation and color constraint" , in Journal of Systems Engineering and Electronics, vol. 31, no. 2, pp. 266-271, April 2020.

[2] He, J., Yu, L., Liu, Z. et al. Image super-resolution by learning weighted convolutional sparse coding. SIViP 15, 967–975 (2021).

[3] Kim, S.; Jun, D.; Kim, B.-G.; Lee, H.; Rhee, E. Single Image Super-Resolution Method Using CNN-Based Lightweight Neural Networks. Appl. Sci. 2021, 11, 1092.

[4] Maral, B. C. (2022). Single Image Super-Resolution Methods: A survey. arXiv (Cornell University).

[5] L. Jiang, M. Zhong and F. Qiu, "Single-Image Super-Resolution based on a Self-Attention Deep Neural Network," 2020 13th International Congress on Image and Signal Processing, BioMedical Engineering and Informatics (CISP-BMEI), Chengdu, China, 2020, pp. 387-391, doi: 10.1109/CISP-BMEI51763.2020.9263584

[6] William Symolon, Cihan Dagli, Single-Image Super Resolution Using Convolutional Neural Network, Procedia Computer Science, Volume 185, 2021, Pages 213-222, ISSN 1877-0509

[7] Ahmed, A., Kun, S., Ahmed, J. et al. Multimodal image enhancement using convolutional sparse coding. Multimedia Systems 29, 2099–2110 (2023).

[8] Ilesanmi, A.E., Ilesanmi, T.O. Methods for image denoising using convolutional neural network: a review. Complex Intell. Syst. 7, 2179–2198 (2021).

[9] Li, X., Wu, Y., Zhang, W. et al. Deep learning methods in real-time image super-resolution: a survey. J Real-Time Image Proc 17, 1885–1909 (2020).

[10] C. Tian, Y. Xu, W. Zuo, B. Zhang, L. Fei and C. -W. Lin, "Coarse-to-Fine CNN for Image Super-Resolution," in IEEE Transactions on Multimedia, vol. 23, pp. 1489-1502, 2021, doi: 10.1109/TMM.2020.2999182.