



IJRASET

International Journal For Research in
Applied Science and Engineering Technology



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 10 **Issue:** VI **Month of publication:** June 2022

DOI: <https://doi.org/10.22214/ijraset.2022.44826>

www.ijraset.com

Call:  08813907089

E-mail ID: ijraset@gmail.com

Image De-Noising Using Convolutional Variational Autoencoders

M. Sreeteish¹, Afnan Mohammed², S. Shashidhar Reddy³, C N. Sujatha⁴

^{1, 2, 3, 4}Department of Electronics and Communication Engineering, Sreenidhi institute of science and technology Ghatkesar, Telangana, Hyderabad

Abstract: Typically, image noise is random colour information in picture pixels that serves as an unfavourable by-product of the image, obscuring the intended information. In most cases, noise is injected into photographs during the transfer or reception of the image, or during the capture of an image when an object is moving quickly.

To improve the noisy picture predictions, autoencoders that denoise the input images are employed.

Autoencoders are a sort of unsupervised machine learning that compresses the input and reconstructs an output that is very similar to the original input. The autoencoder tries to figure out non-linear correlations between data points. An encoder, a latent space, and a decoder all exist in autoencoders.

The encoder reduces the dimensionality of an original picture to its latent space representation, which is then used by the decoder to reconstruct the reduced dimensional image back to its original image. Basic Autoencoder, Variational Autoencoder, and Convolutional Autoencoder are the three approaches that were employed to denoise the picture. In the basic and convolutional autoencoders, there is only one loss parameter, however in the variational autoencoder, there are two losses: generative loss and latent loss. TensorFlow as the frontend and Keras as the backend are used to implement autoencoders in this project. The noisy pictures are trained on every convolutional variational autoencoder techniques to produce a decent prediction of noisy test data.

Keywords: autoencoder, latent space, convolutional autoencoder, variational autoencoders

I. INTRODUCTION

Deep learning-based autoencoders have progressed significantly in recent years as a result of remarkable advances in the field of data science. Autoencoders are a subset of unsupervised learning techniques, which are a sort of machine learning approach. Picture denoising, image production, dimensionality reduction, image colorization, and feature extraction are just a few of the applications where autoencoders are utilised. Autoencoders are used in a variety of ways in all of the applications. We employ neural networks to conduct autoencoders in such a manner that the neural network has a bottleneck that imposes a compressed knowledge representation of the original input.

To put it another way, see Autoencoders as a sandwich, with the lower bread acting as an encoder, the material in the middle acting as a bottleneck, and the top bread acting as a decoder. As a result, autoencoders must be symmetric.

The hidden layer between the encoder and decoder layers is known as the latent space. It's a compressed representation of the data being input. It may be used to learn data characteristics and to develop a simplified data representation for analysis. In the latent space representation of a variational autoencoder, the encoder network will convert the input samples into two parameters, which are designated mean and variance. There are no such parameters in latent space for basic autoencoders and convolutional autoencoders. There are no such parameters in latent space for basic autoencoders and convolutional autoencoders.

The main distinction between an autoencoder and a variational encoder is the latter's capacity to supply continuous data or a range of data in the latent space, which aids in the generation of new data or images.

Variational autoencoders are well-known for their ability to generate the supplied input. There are two types of losses: reconstruction loss and KL-divergence loss. We sum both losses in VAE and strive to reduce the overall loss.

II. LITERATURE SURVEY

According to reference [1], exaggerated in suggesting "Mixture of Experts similarity VAE" because of the growing popularity of deep clustering as a strategy for grouping high-dimensional data (MoE-sim-VAE). A Mixture of Experts architecture is used in this decoder. The model can execute clustering tasks, similarity-based representation, and create high-dimensional data rapidly and reliably. Adversarial training will be included to the MoE decoder in the future.

In paper [2], the author claimed that the normal VAE is unable to capture several desired qualities that are necessary for representation learning, thus he created the Auto-encoding variational autoencoder (AAVE), which produces resilient representation.

Rendering to reference [3], shown that employing warm-up for VAE with respect to KL loss produces far more convincing picture production outcomes. This publication also includes AE applications like picture morphing and reconstruction. The algorithm's efficiency is governed by kernel settings. The population-based image optimization technique with novel fitness functions is the emphasis of this portion of the image.

The author of this article [4] demonstrates how PCAs ease the difficulty of reducing high-dimensional data without sacrificing the original input data. In the case of severe noise and distortion in the brain picture, a segmentation approach integrating the two aforementioned algorithms was devised and assessed.

The author [5] detailed how AE are used to denoise an image in this article. He also depicted the loss it suffered. We should also see how the model performs when different noisy photos are shown to it. The residual network's effectiveness and efficiency may explain its widespread use. The residual network was utilised by the researchers to restrict the number of convolutions in their network.

The author detailed [6] the architecture as well as the use of VAE in this paper. In the end, he came to the conclusion that VAEs are the best for generation applications. A total variety and data terms are used to solve the computer complexity of the approximation function. A predetermined mask is used to designate a damaged pixel in order to keep the noise-free pixels intact. This specified mask ensures that there are no non-zero differences between the original and reconstructed images.

According to reference [7], the deep models with several layers of dependent stochastic variables are difficult to train, he invented ladder VAE, which corrects the model's generative distribution iteratively. He also shown that for training variational models with multiple stochastic layers, batch normalization and deterministic warm-up are essential.

Pham et al. suggested a new method for denoising a picture.[8] The importance of the approximation function is emphasized in the development of an effective numerical method. The approximation function determines how close the original and recreated images are. Damaged pixels are not taken into account when determining closeness. The edge is also protected since overall picture fluctuation is kept to a

minimum. A total variety and data terms are used to solve the computer complexity of the approximation function. A predetermined mask is used to designate a damaged pixel in order to keep the noise-free pixels intact. This specified mask ensures that there are no non-zero differences between the original and reconstructed images.

According to Mishra et al., a segmentation of the [9] picture using fuzzy clusters seeks to tackle the noise density problem as well as the initial cluster reliance. For designating the picture with knowledge of local spatial information, a novel kernel fuzzy clustering approach that employs the entropy of the kernel is utilized. The K means clustering approach may be used to represent the image with computational simplicity. The algorithm's efficiency is governed by kernel settings. The population-based image optimization technique with novel fitness functions is the emphasis of this portion of the image.

Ferlay et al. [10] divided the subject into five groups while working on it, based on their segmentation methodologies and concepts. Patil and Bhalchandra employed the Fuzzy C-Means method and the Support Vector Machine algorithm (SVM) for image segmentation. In the case of severe noise and distortion in the brain picture, a segmentation approach integrating the two aforementioned algorithms was devised and assessed.

A. The Main Objectives Of This Project Are Listed Below

To understand functionality of autoencoders. To understand the accuracy of the models to implement in the project. To train and test the given dataset.

B. Main Motive To Choose This Project

Previously we have done our project based on comparison the accuracy of PCA and ICA algorithms so we have obtained PCA having the most accuracy than ICA. To obtain 99% of more accurate image we use this image de-noising. This will also help us in re constructing the actual detailed image from noises and it can be helpful to us in many bases like in security purpose, re generating details of an old image and mostly cameras cannot capture high detailed images in nights, so that we get high noise in the image which can be easily reconstructed.

The fig (a) is a noisy image which is sent as the input to the PCA and fig (b) is the denoised or the output image which is the desired image.



fig (a):

fig (b):

Figure: 1 (a) noisy image and (b) de-noised image

C. Addition Of Noise To Data In Real World

An image is $N \times M$ dimensional array with $N \times M$ pixels that ranges from 0 (for white) to 255 (for black), the random variation of color information in pixels of an image will lead to an addition of noise in that image.

Even when we compress the memory of the image, sometimes we encounter blurry ness in that image so with the help of this image de-noising phenomenon we can recover the details of images

III. METHODOLOGY FOR IMAGE DE-NOISING

A. Auto Encoders

Autoencoders were initially presented as a neural network design that primarily focuses on encoding, in which the input data is compressed and then rebuilt using a decoder. Autoencoders are similar to Principal Component Analysis in that they have the same generalized or latent representation as PCAs and are similarly an unsupervised learning approach that does not require labels.

B. Variational Auto Encoders

VAEs are neural networks that receive input data and convert it to its latent representation, which has mean and variance as hidden layers. These two hidden layers are then transformed to sampled latent vectors, through which the original data is rebuilt. In VAE, we must concentrate on the two losses it causes. Here fig (3) shows the variational autoencoder where the encoder samples the input to its latent space representation and we have a decoder which samples the latent space and reconstructs the input image.

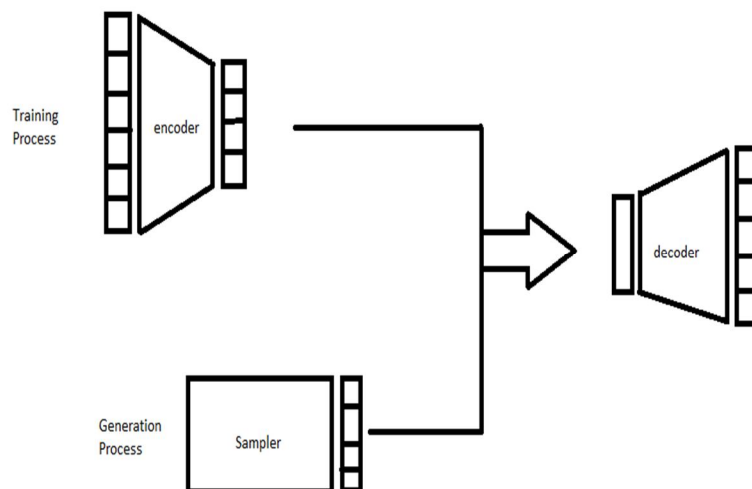


Fig (3): Architecture of variational autoencoder

C. Convolutional Variational Auto Encoders

Convolutional neural networks excel in a variety of tasks, therefore it's only reasonable to consider them for picture denoising. In convolutional autoencoders, max-pooling layers are used in conjunction with convolutional layers to compress the size of the original input and represent it in a latent space, and up-sampling is used in conjunction with convolutional layers to reconstruct the original input from the latent space at the decoder side. We can also use a transposed convolution layer instead of up-sampling, but the up-sampling method reduces the number of training parameters in the network and is suitable for a wide range of problems, especially when the training data is sufficient.

Using Python built-in methods, import essential libraries such as NumPy, TensorFlow, Keras, and Matplotlib for all three types of autoencoders. The mnist data was utilised in this experiment. You can also use any other information.

Fig (4) shows the architecture of our model where it consists of an encoder given with input image (256,256,3) which gives a latent space representation and the decoder decodes the latent space representation to give the original input image back.

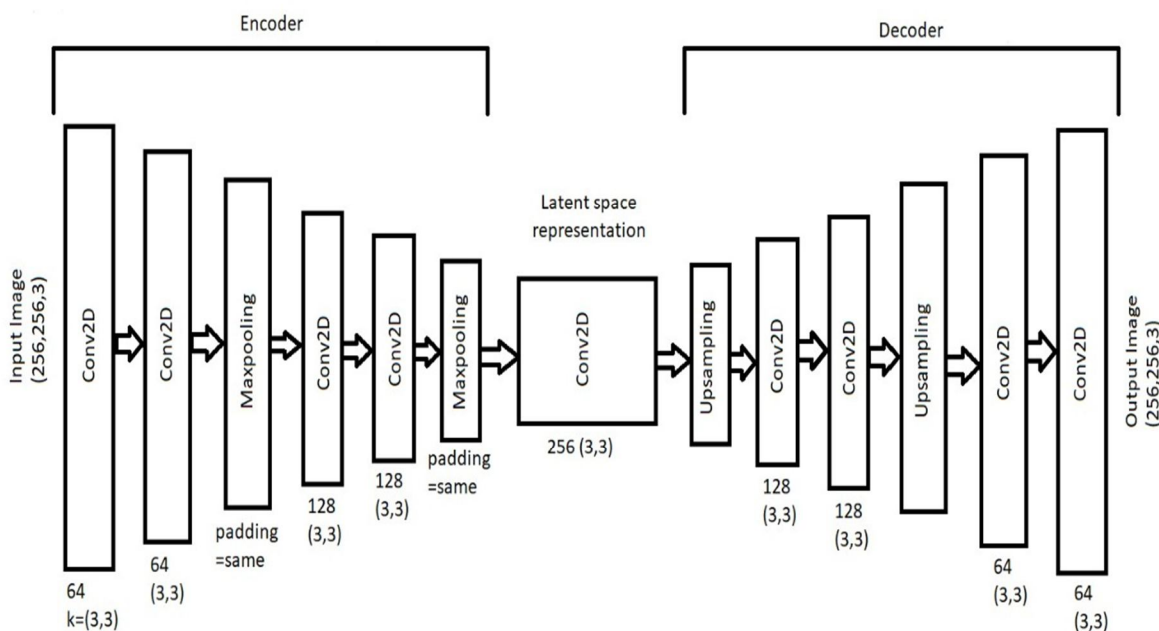


Fig (4): Our model architecture

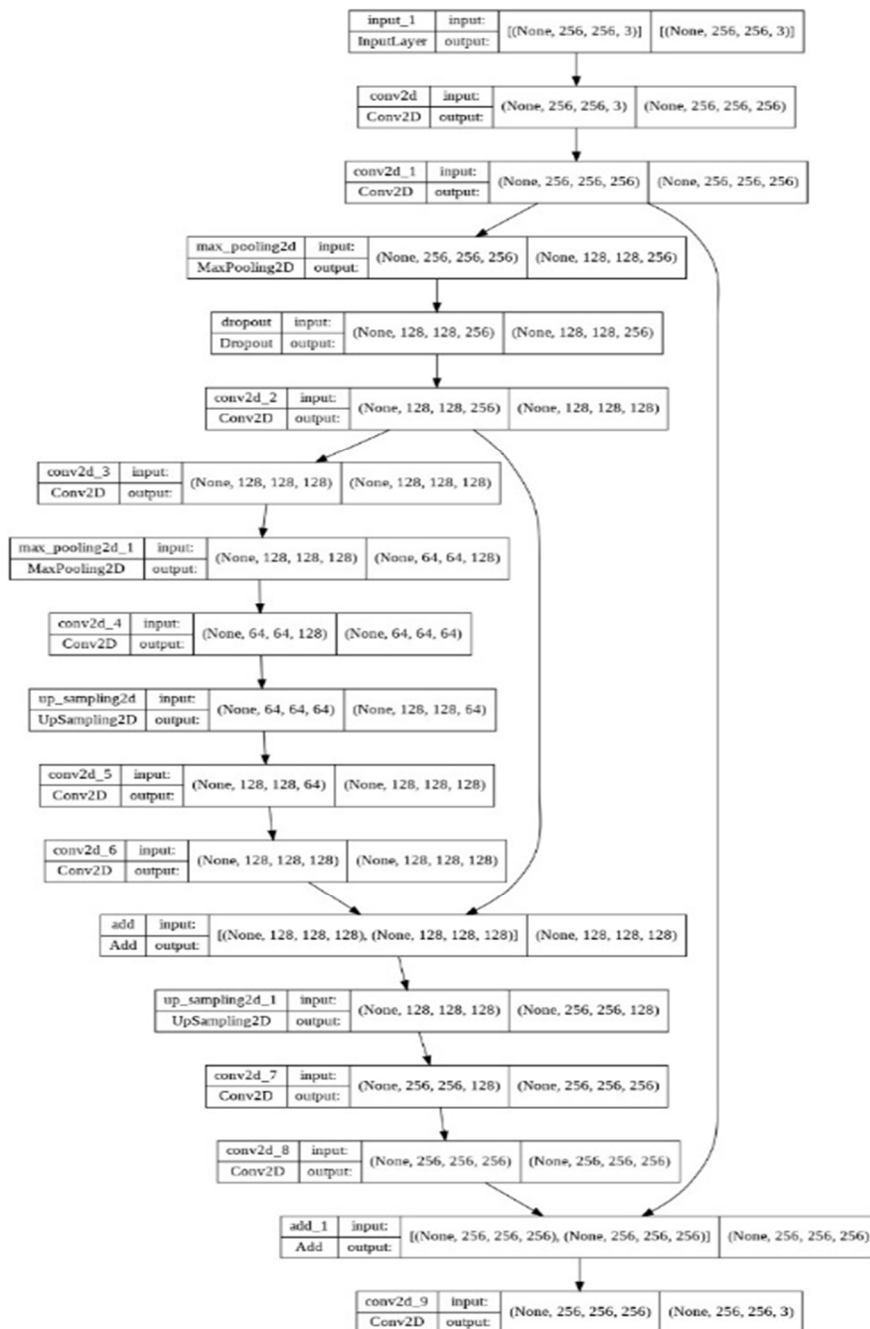
IV. PROPOSED METHOD:

To begin, import the required libraries, including NumPy, TensorFlow, keras, and matplotlib. Now save the files as (x_train,y_train), (x_test,y_test). Now we need to normalise the data so that the mean is near to zero. Normalization of data accelerates learning and convergence. For convenience, use them. Reshape () technique to turn the 3D data learned into 2D data. Now add some noise to the data and call it x_train_noisy and x_test_noisy respectively. Because the values of data can go beyond 1 when employing white gaussian noise, the clip function is used to clip the values between 0 and 1. Using the "plot function," see the data with and without noise.

There are four encoder layers in CAE, one latent space representation layer, and four decoder levels on the encoder side. The below fig shows the model architecture of our project.

The encoder is made up of filters with kernel sizes of 64,64,128,128 and kernel size of 3. Dropouts and maximum pooling are employed. To avoid data overfitting, regularisers are utilised, and 28 filters with k size = 3 are contained in the latent space representation. Now, in order to rebuild the supplied input from its latent space, we employed four layers of decoders, each having 128,128,64,64 filters and a k size of 3. To gradually scale up the layers on the decoder side, we employed up-sampling.

Fig (5) is the flow diagram of our CVAE model architecture which show the each and every step that takes place in the model.



Fig(5): CVAE model architecture flow diagram

V. RESULT

The training loss and validation loss abruptly fallen down from 0.12 to 0.023 after 16 epochs, training accuracy and validation accuracy abruptly increased from 0 to 0.807 after 16 epochs. Later we have seen that there is jump in the validation accuracy from 0.807 to 0.816 after 20 epochs. Invalidation loss and validation accuracy there is no change after 20 epochs. The obtained above results are written based on the below shown graph figures which are obtained in our project. The below shown graph figures are plotted against training loss vs validation loss [fig(f)], training accuracy vs validation accuracy [fig(g)], training loss vs training accuracy[fig(h)], validation loss vs validation accuracy[fig(I)] Here we can see how images are converted from noisy image to de-noisy image by reducing the noise in the images. Reduction of noise can make an image detail more visible and Clear, one of the advantages of image denoising is we can obtain the true image out of noisy image.

Table 1: Basic Auto Encoder

Epoch	Training		Testing with noise	
	Loss	Accuracy	Loss	Accuracy
50	0.012	0.014	0.013	0.014
100	0.008	0.015	0.011	0.013

For the above basic autoencoder table (1) for 50 epochs and 100 epochs we got the training loss as 0.012 and 0.008 whereas with noise we got 0.014 and 0.013 as training accuracy respectively, which concludes that there is not much change and hence it cannot be used for image denoising application.

Table 2: Variational auto encoder

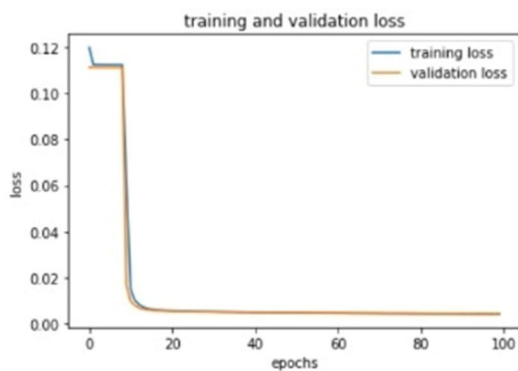
Epoch	Training		Testing with noise	
	Loss	Accuracy	Loss	Accuracy
50	62.154	0.013	62.012	0.133
100	60.769	0.016	60.805	0.143

For the above variational autoencoder table for 50 epochs and 100 epochs we got the training loss as 62.15 and 60.78 whereas with noise we got 1.133 and 0.143 as training accuracy respectively, which concludes that there is not much change and hence even this cannot be used for image denoising application.

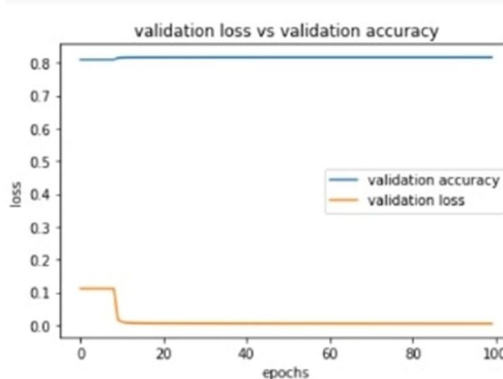
Table 3: Convolutional auto encoders

Epoch	Training		Testing with noise	
	Loss	Accuracy	Loss	Accuracy
50	0.0049	0.815	0.0047	0.814
100	0.0043	0.815	0.0043	0.815

For the above convolutional variational autoencoder table for 50 epochs and 100 epochs we got the training loss as 0.0049 and 0.0043 whereas with noise we got 0.814 and 0.815 as training accuracy respectively. As we have an accuracy of 81.4%, we expect convolutional variational autoencoders will perform well in denoising the input image, to our surprise as expected we got good results from C-VAE which are shown in the below figures.

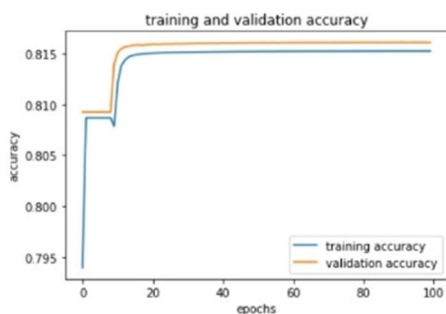


Fig(6)

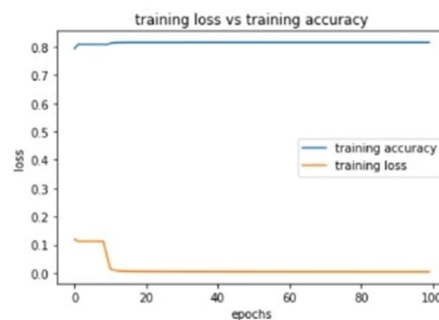


fig(7)

Graph 1 or the fig (6) indicates the validation loss and training loss in the model we made where the training loss and the validation loss are same where it drops from the peak and in the graph 2 or the fig (7) shows validation loss versus validation accuracy where validation accuracy is almost constant with validation loss at 0.



fig(8)



fig(9)

Graph 3 or the fig (8) show the training and validation accuracy where training accuracy increases with validation accuracy and fig (9) show the training loss versus training accuracy training accuracy is almost consent with training loss at 0

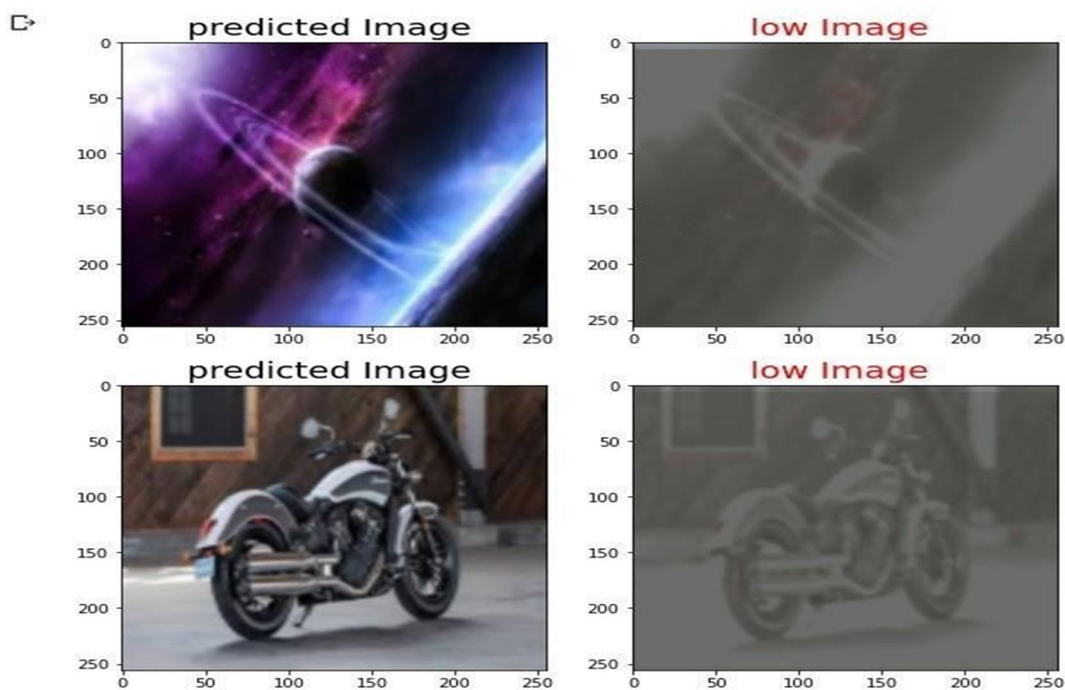


Fig (10): low image and predicted image

Here in the Fig (10) shows that the low image which is given as input in the model we made and the it also shows the predicted image which is accurate.

VI. CONCLUSION AND FUTURE SCOPE:

The GAN was found to be the most popular approach for CNN picture denoising. For extraction, several approaches employed the generator and discriminator. production of a clean image Surprisingly, some researchers believe that the GAN approach was integrated with the DCNN algorithms. The CNN and U-Net were also used as feedforward sources. The leftovers researchers utilised the network on various occasions. a justification

The residual network's effectiveness and efficiency may explain its widespread use. The residual network was utilised by the researchers to restrict the number of convolutions in their network.

So, with the help of this de-noising algorithm, there are many applications such as tracking, video processing, image analysis. We can also use this image de-noising in night vision images because in low light vision the images captured by the CCTV cameras are very noisy so we can easily remove noise from those images as well.



REFERENCES

- [1] Wikipedia contributors. "Structural similarity." Wikipedia, The Free Encyclopedia. Wikipedia, The Free Encyclopedia, 19 Nov. 2020. Web. 29 Jan. 2021.
- [2] Ledig, C., Theis, L., Huszar, F., Caballero, J., Cunningham, A., Acosta, A., . . . Shi, W. (2017, May 25). Photo-Realistic Single Image de-noising Using a Generative Adversarial Network. Retrieved vol. 6, no. 9, pp. 68–72, October 29, 2020, from <https://arxiv.org/abs/1609.04802v5>
- [3] Romano, Y., Isidoro, J. and Milanfar, P., 12 NOV 2020. RAISR: Rapid and Accurate Image Super Resolution. [online] arXiv.org.
- [4] Yang, C.Y., Ma, C., Yang, M.H.: Single-image super-resolution: A benchmark. In: European Conference on Computer Vision, pp. 372–386, 24 SEP 2019
- [5] Pham, T. X., Siarry, P., Oulhadj, H. (2018). Integrating fuzzy entropy clustering with an improved PSO for MRI brain image segmentation. Applied Soft Computing, 65: 230-242. <https://doi.org/10.1016/j.asoc.2018.01.003>
- [6] Shi, W., "Real-Time Single Image and Video Super-Resolution Using an Efficient Sub-Pixel Convolutional Neural Network", arXivpreprints, 17 JULY 2017.
- [7] E. Agustsson and R. Timofte, "NTIRE 29 JULY 2017 Challenge on Single Image Super-Resolution: Dataset and Study," 2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW), Honolulu, HI, 2017, pp. 1122-1131, doi: 10.1109/CVPRW.2017.150.
- [8] C. Dong, C. C. Loy, K. He and X. Tang, "Image de-noising Using Deep Convolutional Networks," in IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 38, no. 2, pp. 295- 307, 27th May. 2016, doi: 10.1109/TPAMI.2015.2439281.
- [9] Mishra, S., Mishra, D. (2015). SVM-BT-RFE: An improved gene selection framework using Bayesian T-test embedded in support vector machine (recursive feature elimination) algorithm. Karbala International Journal of Modern Science, 1(2): 86-96. <https://doi.org/10.1016/j.kijoms.2015.10.002>
- [10] Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M. (2015). Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. International Journal of Cancer, 136(5): E359-E386. <https://doi.org/10.1002/ijc.29210>
- [11] J. Huang, A. Singh, and N. Ahuja. Single image de-noising from transformed self- exemplars.



10.22214/IJRASET



45.98



IMPACT FACTOR:
7.129



IMPACT FACTOR:
7.429



INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Call : 08813907089  (24*7 Support on Whatsapp)