Restoration of Images using only noisy data

A Deep Learning Approach using U-NET

[1]Vivek Patil, [2]Pratik Katte, [3]Abhay Patil
[1]Research Student, [2]Research Student, [3]Project Guide
[1]Information Technology,
[1]MCT's Rajiv Gandhi Institute of Technology, Mumbai, India

Abstract: Digital Images have played a vital role in the research and development of technology. With increasing demand in quality for better visualization of the subject in the image, the main challenge in Image Processing is to remove the noise from the original image. There are many examples where images captured contains a lot of noise, for example, Satellite image, MRI, Mobile Captured images. The task of image restoration i.e. taking a corrupted image and estimating the cleaned original image is classical as well studied problem in Computer Vision and Image Processing. In this work, we propose to use Deep Learning, a unique algorithm can be developed that can fix the bad images by learning from noisy images only. Deep Learning models help in solving complex, classical problems which have not been solved for decades. We use U-Net[3], a fully convolutional network which is meant for segmentation of images. We aim to apply this neural network as an image restoration technique to remove Gaussian Noise. This approach can restore bad images to clean images without feeding clean high-quality images to the Neural Network. It will take milliseconds to render the image and produce the resulting clean image.

***Keywords* – Image Restoration, Image Processing, Deep Leaning, Neural Network.**

I. Introduction

In this growing age of Image Processing, it is important for the result to be crystal sharp and to be more informative. Nowadays, images are taken by anyone, anywhere. Right from our mobile phones to professional camera, all are used to capture images. They have various types of quality of hardware which affects the quality of the images. These all images are susceptible to some kind of noise. Image noise can be intrinsic (e.g. due to sensor) or extrinsic (e.g. due to environment) conditions which are often not possible to avoid in practical situations. Therefore, image de-noising plays and important role in a wide classification, where obtaining the original image content is crucial for strong performance.

One of the important challenges in the field of image processing and computer vision is image restoration, where the underlying goal is to estimate the original image by suppressing noise from a noise-contaminated version of the image. The problem of image noise removal remains an open challenge, especially in situations where the images are acquired under poor conditions where the noise level is very high. Image Restoration aiming to reconstruct a high-quality image from its low-quality observation has many important applications, such as low-level image processing, Medical imaging, remote sensing, surveillance, etc.

Creating a neural network model which will have only corrupted data as input and gives out a clean image which helps in improving medical imaging and astronomical imaging. This is a combination of two disciplines i.e. image processing and Deep Learning The image processing uses the concepts of Digital Signal processing for reconstruction of the image signal. These are straight sections of Information Technology.

It is essential to suppress noise from an image as far as possible. At the same time, its fine-details and edges are to be retained as much as practicable. The filtering algorithms to be developed must be of low computational complexity so that they can filter noise in a short time, and hence will find themselves suitable for online and real-time applications.

This research and implementation work focuses mainly on Gaussian noise. Usually, transform-domain filters consume much more time compared to the time taken by spatial-domain filters. Therefore, the following deep learning approach has been proposed.

II. Theoretical Study

Image restoration is different from image enhancement in that the latter is designed to emphasize features of the image that make the image more pleasing to the observer, but not from a scientific point of view. The objective of image restoration techniques is to reduce noise and recover resolution loss. Image processing techniques are performed either in the image domain or the frequency domain. The most straightforward and a conventional technique for image restoration is deconvolution, which is performed in the frequency domain and after computing the Fourier transform of both the image and deconvolution technique, because of its direct inversion of the Point Spread Function (PSF) and undo the resolution loss caused by the blurring factors. This deconvolution technique, because of its direct inversion of the PSF which typically has poor matrix condition number, amplifies noise and creates an imperfect de-blurred image. Also, conventionally the blurring process is assumed to be shift-invariant. Hence more sophisticated techniques, such as regularized de-blurring, have been developed to offer robust recovery under different types of noises and blurring function. It is of 3 types:

1. Geometric correction
2. Radiometric correction
3. Noise removal

Recent advances in the deep neural network models possible to restore images by only looking at corrupted images, models can learn to map corrupted observations to the unobserved clean images without explicitly training with the clean images and without likelihood models of the corruption images. In practices, we show that a single model can learn photographic noise removal, Gaussian noise – all corrupted by different processes – based on noisy data.

This can be done by training a regression-based convolution Neural Network Model with a large number of pairs (x, y) of corrupted images x and clean target y, minimizing the empirical risk

$$\underset{\theta}{\operatorname{argmin}} \sum_i L\left(f_\theta(\hat{x}_i),\, y_i\right),$$

The network should be able to minimize this loss L by solving the point estimation problem individually for every input given.

III. Related work

1) Learning Image Restoration without Clean Data. [1]

This paper uses RED30, a 30-layer hierarchical residual network with 128 feature maps which is designed for Image Restoration. The paper proposes to use the architecture to learn to reconstruct signals from only corrupted examples, without ever observing clean data.

2) Photo-realistic single image super-resolution using a generative adversarial network. [2]

This paper proposes to use General Adversarial Network (GAN) to recover finer texture details when super-resolving at large upscaling factors. Their algorithm is able to recover photo-realistic images from heavily down-sampled images.

IV. Proposed System Details

1) Implementation Methodology

We used SEMMA methodology which stands for "Sample, Explore, Modify, Model, and Assess"

- In the Sample and explore phase of the methodology we collect and understand the dataset.
- In the Modify phase, we prepare the dataset by adding the Gaussian noise to the image dataset. In the Model phase, we focus on applying various modeling techniques on the prepared datasets in order to create models that possibly provide the desired outcome.
- Assess Phase: we evaluate the modeling results that show the reliability and usefulness of the created models.

2) Theoretical framework

a) Gaussian Function

We used Gaussian functions to describe the Gaussian filters which are added on the image to create image dataset having Gaussian noise.
Gaussian functions are of the form: $f(x) = ae^{-\frac{(x-b)^2}{2c^2}}$

b) U-Net Model

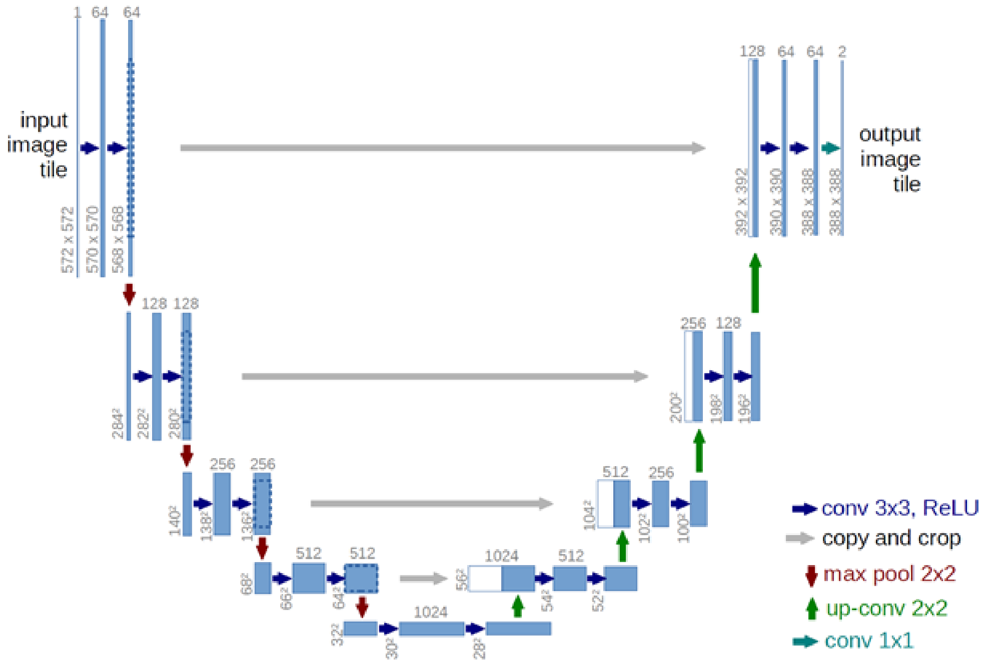


Fig. 1: U-Net Model

U-Net architecture is a Fully Convolutional Network. U-Net is symmetric in shape and hence owes its name **U**. The skip connections between the down-sampling path and the up-sampling path apply a concatenation operator instead of a sum which intend to provide local information to the global information with up-sampling. In image denoising, to get a finer and cleaner result it is very important to use low-level details while retaining high-level semantic information. The skip connections create an information path allowing signals to propagate between low and high levels in a much easier way and which also compensates low-level details to high-level semantic filter.

The U-net combines the low-level information from the down-sampling path with the contextual information in the up-sampling path to finally obtain general information which is necessary for image denoising task.

The U-Net consists of a contracting path and an expansive path. The contracting path follows the typical architecture of a convolution network. It consists of the repeated application of two 3X3 convolutions (unpadded convolutions), each followed by a rectified linear unit (ReLU) [4] and 2X2 max polling operation with stride 2 for down-sampling. At each down-sampling step, we double the number of feature channels.

Every step in the expansive path consists of an up-sampling of the feature map followed by a 2X2 convolution ("up convolution") that halves the number of feature channels, a concatenation with the correspondingly cropped feature map from the contracting path, and two 3X3 convolutions, each followed by a ReLU. The cropping is necessary due to the loss of border pixels in every convolution. At the final layer, a 1X1 convolution is used to map each 64-component feature vector to the desired number of classes. In total the network has 18 convolutional layers.

NAME	N_{out}	FUNCTION
INPUT	n	
ENC_CONV0	48	Convolution 3×3
ENC_CONV1	48	Convolution 3×3
POOL1	48	Maxpool 2×2
ENC_CONV2	48	Convolution 3×3
POOL2	48	Maxpool 2×2
ENC_CONV3	48	Convolution 3×3
POOL3	48	Maxpool 2×2
ENC_CONV4	48	Convolution 3×3
POOL4	48	Maxpool 2×2
ENC_CONV5	48	Convolution 3×3
POOL5	48	Maxpool 2×2
ENC_CONV6	48	Convolution 3×3
UPSAMPLE5	48	Upsample 2×2
CONCAT5	96	Concatenate output of POOL4
DEC_CONV5A	96	Convolution 3×3
DEC_CONV5B	96	Convolution 3×3
UPSAMPLE4	96	Upsample 2×2
CONCAT4	144	Concatenate output of POOL3
DEC_CONV4A	96	Convolution 3×3
DEC_CONV4B	96	Convolution 3×3
UPSAMPLE3	96	Upsample 2×2
CONCAT3	144	Concatenate output of POOL2
DEC_CONV3A	96	Convolution 3×3
DEC_CONV3B	96	Convolution 3×3
UPSAMPLE2	96	Upsample 2×2
CONCAT2	144	Concatenate output of POOL1
DEC_CONV2A	96	Convolution 3×3
DEC_CONV2B	96	Convolution 3×3
UPSAMPLE1	96	Upsample 2×2
CONCAT1	$96+n$	Concatenate INPUT
DEC_CONV1A	64	Convolution 3×3
DEC_CONV1B	32	Convolution 3×3
DEV_CONV1C	m	Convolution 3×3, linear act.

Fig. 2: Network Architecture

c) LOSS FUNCTION

Means squared error loss is the average squared difference between the two variables. It is the most commonly used regression loss function.

$$MSE = \frac{\sum_{i=1}^{n} (y_i - y_i^p)^2}{n}$$

d) OPTIMIZATION FUNCTION

Adam optimizer[5] is an optimizer algorithm which is specifically used for training neural networks to update the network. Adam stands for Adaptive Moment Estimation which computes adaptive learning rates for each parameter. Adam works very well in practice compared to other optimization functions as it converges very fast and rectifies every problem that is faced in other optimization techniques such as vanishing learning rate, slow convergence.

e) ACTIVATION FUNCTION

The activation function is non-linear functions used to learn and understand the complex data and get the output from the node in the network. We used ReLU (Rectified Linear Unit) activation function which is defined as:

$$g(z) = \max\{0, z\}$$

The function gives a linear output to values greater than zero. ReLU is preferred due to its computational simplicity as well as due to its linearity in nature which helps in optimizing the neural network.

f) MAX POOLING

Max pooling is a downsampling technique which is used on the image to reduce computational cost as well as to reduce the dimensionality of the image and allowed to make decisions about the features in the sub-region of the image.
Max pooling is done by applying a max filter to a sub-region of the image.

3) Dataset

For the training purpose of the neural network model, we used ImageNet[6] Validation dataset as our primary dataset. ImageNet is basically a visual database of images designed for visual objection recognition. The database consists of more than 14 million images. The images are primarily of living beings and non-living beings.

We train the neural network using 50,000 images in the ImageNet validation set. The images are cropped down to 256X256. We used 30,000 images randomly from the 50,000 images. We used the 80/20 formula to divide the image dataset into training and validation dataset.

4) Training

During the training process, the dataset is passed through the U-Net models with a batch size of 12 simultaneously by 5 workers processes with learning rate 0.01. The loss value is obtained using the mean square error loss between the input image and output image obtained from the U-Net model.

For the experiments with images, the number of input and output channels were n=m=3. The network weight was initialized randomly and no batch normalization, dropouts or regularization techniques were used. We used ADAM optimizer with parameter values β_1= 0.9, β_2= 0.99, ε= 10^8 to update the weights of the network in order to improvise the task of noise removal from the given input image.

5) Validation

During the validation process which is performed after every epoch of training the network with training dataset, the validation dataset is used to test and validate the network how much it has learned and evaluated its improvement. In order to evaluate the network, we calculate the PSNR value of the output image obtained from the U-Net model which can hence show the improvement in the removal of the noise from the image.

6) Visualization

To visualize the development of the network we used Tensorboard python library. We plot the PSNR values obtained in the validation process and loss values obtained during the training phase.
We also try to visualize and understand the development of the network by observing the true input image and the output image obtained from the validation phase.

V. RESULTS

5.1 Result Analysis

We studied and experimentally performed the noisy target training with added Gaussian Noise. Since the algorithm was not fully trained, yet significantly good results were obtained. The key observation from the result was how the algorithm was trying to restore the image.

We observe that by applying basic statistical reasoning to perform image restoration task using deep neural network have proven to give clean images by only looking at the corrupted images. This also comes with the fact that the algorithm does not have prior knowledge about the corrupted images or any likelihood models.

Fig. 3: (Left) Original image. (Right) Recovered Image

Fig. 4: (Left) Original Image, (Right) Partially Recovered Image

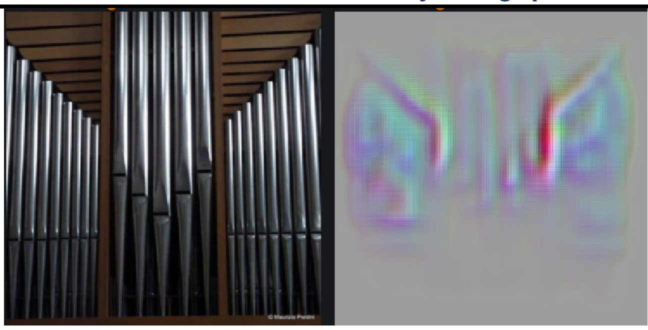
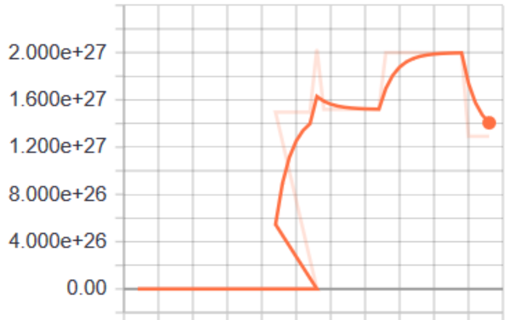


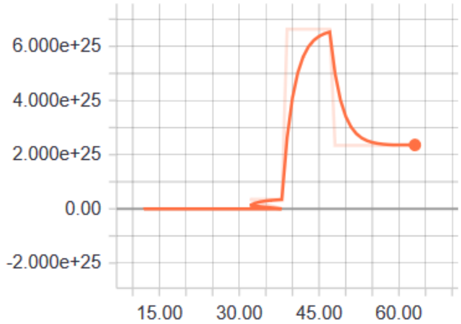
Fig. 5: (Left) Original Image, (Right) Recovering Image

5.2 Tensorboard Analysis

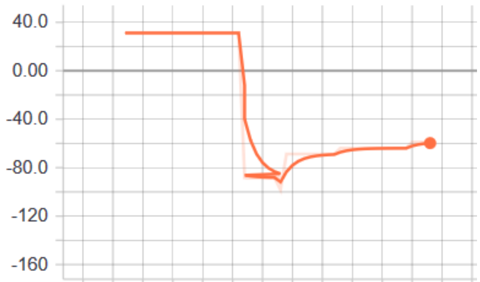
a) Training Loss



b) Validation Loss



c) PSNR Loss



V. Acknowledgment

Thanks to Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miikka Atittala, Timo Aila and the entire NVIDIA research team for this unique approach to consider for image restoration. Because of their original paper of Noise2Noise[1], we got inspired to research in this field and practically perform the experiment.

REFERENCES

[1] Jaakko Lehtinen, Jacob Munkberg, Jon Hasselgren, Samuli Laine, Tero Karras, Miika Aittala, Timo Aila. Noise2Noise: Learning Image Restoration without Clean Data. In arXiv: 1803.04189v1 [cs.CV] 12 Mar 2018.

[2] Ledig, Christian, Theis, Lucas, Huszar, Ferenc, Caballero, Jose, Aitken, Andrew P., Tejani, Alykhan, Totz, Johannes, Investment Performance of Common Stocks in Relation to their Price to Earnings Ratio: A Test of the Efficient Markets Hypothesis. Journal of Finance, 33(3): 663-682.

[3] Ronneberger, Olaf, Fischer, Philipp, and Brox, Thomas. U-net: Convolutional networks for biomedical image segmentation. MICCAI, 9351:234–241, 2015.

[4] Xavier Glorot, Antoine Bordes and Yoshua Bengio (2011). Deep sparse rectifier neural networks (PDF). AISTATS. "Rectifier and softplus activation functions. The second one is a smooth version of the first."

[5] Diederik P. Kingma, Jimmy Ba. Adam: A method for Stochastic Optimization. In aXiv:1412.6980v9 [cs.LG] 30 Jan 2017.

[6] J. Deng, W. Dong, R. Socher, L. Li, Kai Li and Li Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conference on Computer Vision and Pattern Recognition*, Miami, FL, 2009, pp. 248-255. doi: 10.1109/CVPR.2009.5206848.