Image Restoration and Super resolution using Deep Convolutional Neural Networks

Mohit Saraf
Indiana University
sarafm@iu.edu

Paritosh Morparia
Indiana University
pmorpari@iu.edu

1. Introduction

The task of image restoration and super resolution is known to be an ill posed problem in the field of image processing and Computer Vision. Recently Deep learning has shown immense applications with substantial results in Computer Vision. Thus, the question arises that whether these networks generalize well over varying levels of corruption and thus we plan to implement a general model based on encoder decoder architecture with skip connections which has been adopted from the paper [1] by Xiao et al. We also observe the importance of skip connection in terms of recovered image quality when compared to networks without skip connections. Also, the architecture proposed in the paper [1] is mostly experimented on data with uniform blur which is far from what we get in real life. We studied that the problem of solving motion blur has been studied by Nah et al[12] using Generative Adversarial Networks (GAN) and they successfully produced State of Art results for the same on GOPRO dataset created by them. We also try to implement our architecture on the GOPRO datasets and analyze the results for the same.

2. Background and related work

Extensive work has been done related on image. de-noising and restoration. Many approaches in pre-deeplearning era, were based on traditional approaches such as Total Variation [6, 7] and BM3D [8]. More recent works include DNN based approach for image de-noising. Stacked denoising auto-encoder [9] and [13] area well-known approaches for image denoising.

Xu et al. proposed an encoder-decoder architecture for image de blurring where the network learned localized blur kernels. In another approach Sun et al. proposed sequential deblurring approach by generating pair of blurry and sharp patches with 73 blur kernels and then learn likelihood of blur kernel in the given patch of image. The image was constructed by methods described in [10]

It is interesting to know that most of the works use artificially generate noise and blur to estimate the output and learn the kernel that affects the image.

3. Dataset

We have used GOPRO [13] dataset for training and testing our model on deblurring task as it gives more natural and localized blur where Nah et al. have averaged over several images using high speed camera to create motion blur. The dataset is composed of 3214 pair of blurry and sharp images in 1280x720 resolution. We scaled it down to 256x256 to make it trainable on limited resources.

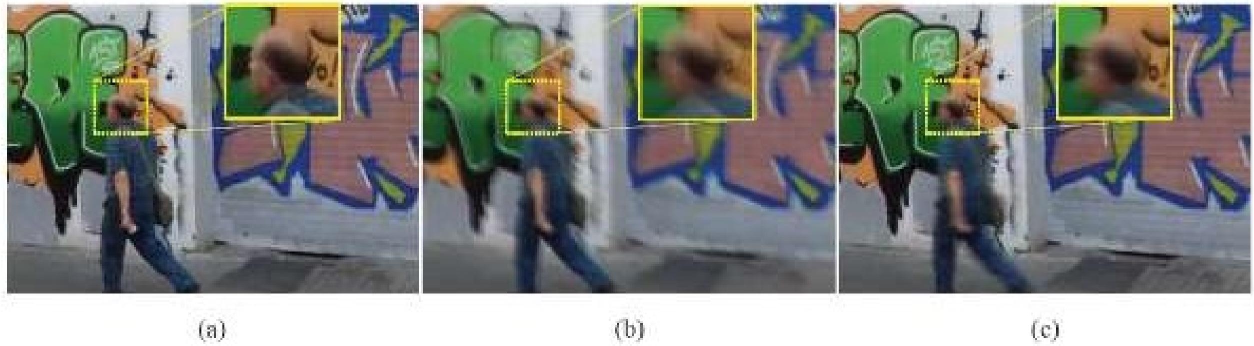

Figure 2. (a) Ground truth sharp image. (b) Blurry image generated by convolving a uniform blur kernel. (c) Blurry image by averaging sharp frames. In this case, blur is mostly caused by person motion, leaving the background as it is. The blur kernel is non-uniform, complex shaped. However, when the blurry image is synthesized by convolution with a uniform kernel, the background also gets blurred as if blur was caused by camera shake. To model dynamic scene blur, our kernel-free method is required.

Figure 1: Source: Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring Seungjun Nah Tae Hyun Kim Kyoung Mu Lee.[11]

For Image denoising task, we have used sharp images from GOPRO dataset to generate noisy images by adding Gaussian Noise of varying standard deviation ($\sigma = 10, 30,50,70$) to the sharp images of the complete dataset of 3214 images. Noise is added to each image to generate total of 4204 pair of images for training and 2222 pair of images for testing. The main aim of our denoising model is to denoise over different strengths of noise using a general model. So, the dataset for noise correction is a combination of images with varying noise strength and is as following:

Standard Deviation	Training Image pairs	Test Image pairs
$\sigma = 10$	1102	611
$\sigma = 30$	1000	500
$\sigma = 50$	1102	611
$\sigma = 70$	1000	500

4. Network Architecture and Methods Used:

In our study we have implemented three different variants of a full connected Deep Neural Networks architecture for denoising and deblurring the corrupted images as mentioned in the paper.

First, we tried a simple Auto-encoder neural network architecture which consisted of ten convolution layers followed by ten deconvolution layers. We used a filter size of 3x3 and number of filters used were 64. We have used batch normalization after each layer in our architecture to minimize covariant shift and speed up training process along with continuously normalizing the values in each layer independently. It also adds a little bit regularization during the training process. We have used ReLU as activation function in our convolution and deconvolution layers.

After this we tried using symmetric skip connections in our previous network where there was a skip connection between a convolution layer and the corresponding deconvolution layer after every step of 2. We also added ReLU activations after each point where a skip

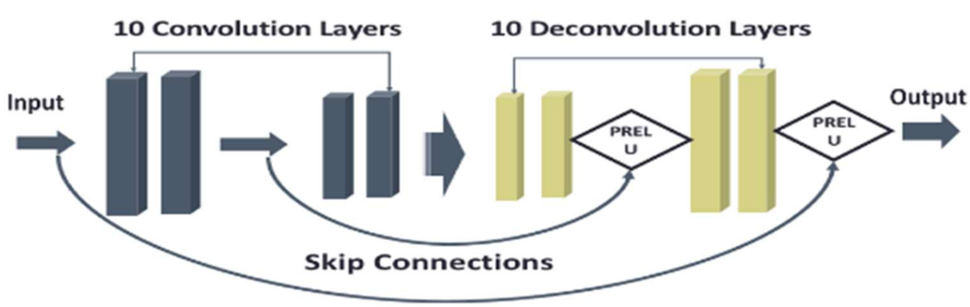

Figure 2 **. Fully Connected Deep Neural Network with skip connections**

connection between a convolution and deconvolution layer meets. Basically, the intuition behind skip connections comes from highway networks and deep residual networks. Skip connections help us overcome the problem of gradient vanishing as we go deeper in terms of depth in a neural network. Another advantage of using skip connections is that it helps us minimize the loss of image details as we go deeper into the networks while recovering image using the deconvolution step.

We also tried an architecture using a fully convolutional neural network with skip connections but Max pooling after convolutional layer instead of convolutional strides to down sample the image and upsampling with convolutional layer instead of deconvolutional layer. We believe this will speed up convergence and thus can be useful in implementation on mobile devices with lower processing power.

For each architecture we have used sigmoid as activation function in our output layer as it helps us bound the range of values between 0 to 1 which is required for images. We used Adam optimizer for our study and tested for learning rates from 10-1 to 10-5. We used mean square error as the loss function for our model. We also tried out PReLU for our architectures as an activation function.

5. Results:

We used two variations of GOPRO dataset in to experiment with our model architecture.

First, we trained our convolution deconvolution model with skip connections on noisy dataset that we created and tested to check if the model works on de-noising the image. We trained our model on 4204 sharp-noisy pairs of images and remaining 2222 pairs were used for test. The results obtained showed clarity in the image with improvement in image quality. We could have performed much better on a deeper network and with more processing power and computing resources.

We then tried running our model on same dataset using up sampling instead of deconvolution and Max pooling instead of using convolutional strides and observed increase in training speed with comparable results.

We also tried using convolution and deconvolution model without skip connections and obtained worse results.

We then tried our original convolution deconvolution model with skip connections for deblurring task and found insignificant result minimal change in the blur image.

Architecture Type	**Task**	**MSE**	**PSNR**
Without Skip Connections	De- Noising	0.00677	-4.74
Without Skip Connection	De-Blurring	0.0045	-1.19
With Up sampling and Max Pooling	De- Noising	0.00093	14
With Up sampling and Max Pooling	De- Blurring	0.004	-0.17
With Skip Connections	De-noising	0.00090	15
With Skip Connections	De-Blurring	0.00356	0.84
Nah et al.	De-Blurring	--	29.08

Table 1

Figure 4Prediction for De-noising the image on skip connection architecture. Source: GOPRO dataset [12]

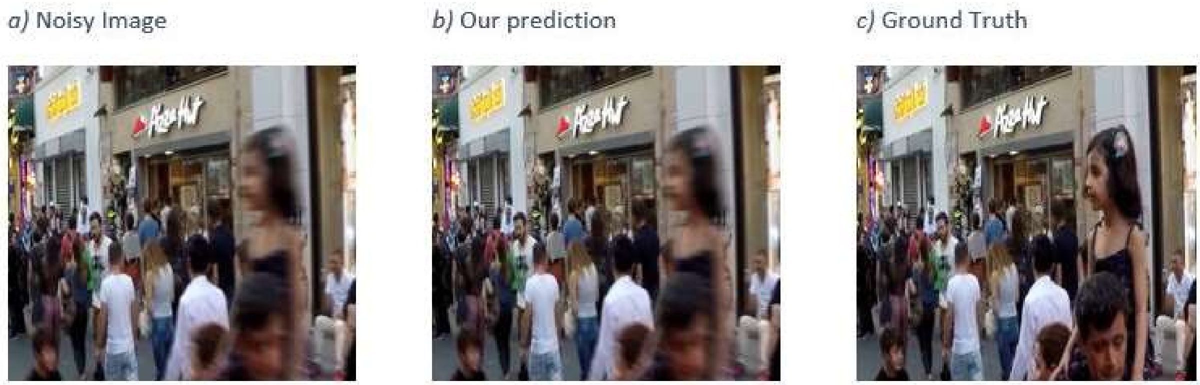

Figure 3: Prediction for De-Blurring the image on skip connection architecture. Source: GOPRO dataset [12]

6. Discussions:

In this paper we proposed using simplified architecture and test learning capability and generalization ability achieved by the models. After conducting the experiments on our 3 different model architectures, we found out that the models with skip connections performed well with the task of image denoising as compared to the models without skip connections.

6.1 Image Denoising:

Image denoising worked well using skip connections. From this we could infer that skip connections do a good job in passing information which is normally lost during convolutional layers. Here skip connections defer from residual networks as convolutional layers are connected to its corresponding deconvolutional layer in the proposed architecture by [1]. Also, in deeper architectures, there is a problem of vanishing gradient which is minimized by skip connections as the initial layers are connected to the end layers in the architecture we have implemented.

The difference between the architecture proposed in [1] and the previous architectures is that the existing methods first estimated the level of corruption and was trained for handling single noise level. While architecture proposed by Mao et al [1] successfully generalizes the noise in the image.

6.2 Image Deblurring:

In the paper [1] Mao et al. proposed a generalized model to handle different levels of corruption. This model

The results of this architecture deblurring architecture were unsatisfactory. There are a few possible reasons the model was not successful for this application:

- The proposed architecture could not generalize motion blur.
- Limited resources:
    - The convergence did not occur with the number of epoch (100) and given learning rate.
    - The architecture was not deep enough for the model to generalize motion blur. A deeper architecture

and better resources would have helped with de-blurring task

7. Conclusion and Future work:

From this experiment that we performed to determine a good model for image sharpening and restoration using different architectures, we can say that skip connections helped recovering clean images from the noise, while the architecture proposed in [11] is most proficient for handling the blur in the image. There is a possibility to train the model on more simple, generalized and deeper architecture given more resources.

This model has many real life applications such as performance enhancement in photography, better performance in security cameras and image quality improvement in handheld devices.

9. References:

[1] Image Restoration Using Very Deep Convolutional Encoder-Decoder Networks with Symmetric Skip Connections, Xiao-Jiao Mao, Chunhua Shen, Yu-Bin Yang

[2] A. Gupta, N. Joshi, C. L. Zitnick, M. Cohen, and B. Curless. Single image deblurring using motion density functions. In ECCV, pages 171–184. Springer, 2010. 1

[3] S. Harmeling, H. Michael, and B. Sch¨olkopf. Spacevariant single-image blind deconvolution for removing camera shake. In Advances in Neural Information Processing Systems, pages 829–837, 2010. 1

[4] K. He, X. Zhang, S. Ren, and J. Sun. Deep residual learning for image recognition. In CVPR, pages 770–778, 2016. 4

[5] M. Hirsch, C. J. Schuler, S. Harmeling, and B. Sch¨olkopf. Fast removal of non-uniform camera shake. In ICCV, 2011. 1, 3

[6] S. Osher, M. Burger, D. Goldfarb, J. Xu, and W. Yin. An iterative regularization method for total variation-based image restoration. Multiscale Modeling & Simulation, 4(2):460–489, 2005.

[7] L. I. Rudin, S. Osher, and E. Fatemi. Nonlinear total variation based noise removal algorithms. Phys. D, 60(1-4):259–268, November 1992.

[8] K. Dabov, A. Foi, V. Katkovnik, and K. O. Egiazarian. Image denoising by sparse 3-d transform-domain collaborative filtering. IEEE Trans. Image Process., 16(8):2080–2095, 2007.

[9] L. Xu, J. S. Ren, C. Liu, and J. Jia. Deep convolutional neural network for image deconvolution. In Advances in Neural Information Processing

[10] D. Zoran and Y. Weiss. From learning models of natural image patches to whole image restoration. In ICCV, pages 479–486. IEEE.

[11] Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring Seungjun Nah Tae Hyun Kim Kyoung Mu Lee.

[12] GOPRO dataset by Seungjun Nah, Tae Hyun Kim, Kyoung Mu Lee Deep Multi-scale Convolutional Neural Network for Dynamic Scene Deblurring

[13] J. Xie, L. Xu, and E. Chen. Image denoising and inpainting with deep neural networks. In Proc. Advances in Neural Inf. Process. Syst., pages 350–358, 2012.