EE608: Final Report

Anika Shrivastava (S22006)
School of Computing and Electrical Engineering (SCEE)
IIT Mandi
s22006@students.iitmandi.ac.in

Shilpa Chandra (S22004)
School of Computing and Electrical Engineering (SCEE)
IIT Mandi
s22004@students.iitmandi.ac.in

***Problem Statement*—The goal of this paper is to develop a blind-spot neural network architecture using an autoencoder model with varying dilation rates which is passed through several 1×1 convolutions, specifically tailored for noisy images.**

***Index Terms*—Blind-Spot, denoising, convolution, dilation, encoder, decoder**

I. Introduction

Image denoising is an important technique in computational photography, but traditional methods rely on clean and noisy image pairs for training deep neural networks. However, obtaining clean samples can be challenging in practical scenarios. Blind-spot neural network architectures enable training directly on noisy images.

We take into consideration Honzátko, et al. [1] as our base paper, who propose a new neural network architecture for image denoising, which is the process of removing noise or unwanted artifacts from an image.

Our architecture **Blind Spot Dilation Architecture** uses an autoencoder model with varying dilation rates which is passed through several 1×1 convolutions. In image processing, **blind spots** refer to regions of an image that are occluded or corrupted and cannot be recovered using traditional image denoising methods. Our architecture trains the network to predict these missing regions of the input signal from the available information in the non-missing regions.

II. Dataset

BSDS300 dataset was used for results and comparison with the baseline paper. We have added Additive white Gaussian noise with **mean 0** and **variance 0.15** to the dataset.

III. Proposed Method

We have solved the problems using two ways:

A. *PRIMARY ARCHITECTURE: Blind Spot Dilation Architecture*

1) In the encoder, the input image is passed through several convolutional layers with decreasing dilation rates such as dilation_rate = 9,7,6,6 for consecutive Conv2D layers as shown in Fig 1. The feature maps from the first two convolutional layers are combined using element-wise addition. The subsequent convolutional layers further capture and refine the image features. The encoded representation is obtained using a convolutional layer with 256 filters with dilation_rate = 4.
2) The decoder part of the model reverses the encoding process. Transpose convolutional layers with increasing dilation rates such as dilation_rate = 6,6,7,9 for consecutive Conv2DTranspose layers are applied to gradually upsample the encoded representation. Similar to the encoder, element-wise addition is used to combine feature maps at certain stages. The final reconstructed image is obtained using a transpose convolutional layer with 64 filters and a sigmoid activation function.
3) Additional convolutional layers with decreasing filter sizes are included after the decoder for further refinement. The model is compiled with the Adam optimizer and the mean squared error (MSE) loss function.
4) The objective of this architecture is to train the denoising autoencoder to effectively remove noise from images and produce high-quality denoised outputs.

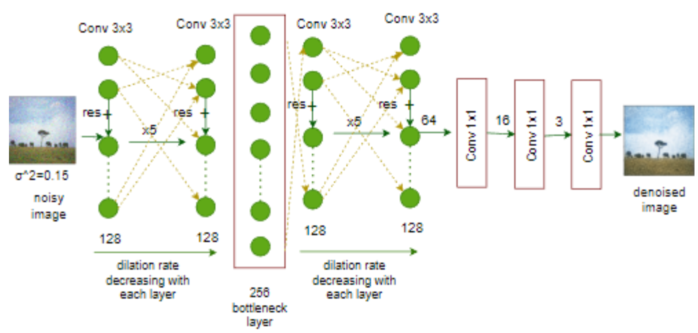

Fig. 1. Blind Spot Dilation Architecture

B. *SECONDARY ARCHITECTURE: Denoising Autoencoder*

The denoising autoencoder is a type of neural network architecture that is designed to remove noise or corruption from input data. It consists of two main components: an encoder and a decoder. The encoder learns to compress the input data into a lower-dimensional representation, while the decoder aims to reconstruct the original input from the encoded representation.

1) The encoder takes the input image and applies convolutional layers to extract relevant features. It consists of three convolutional layers with increasing channel sizes. The first layer has 128 filters, followed by a layer

with 256 filters, and the final layer with 512 filters. Each convolutional layer uses a 3x3 kernel and ReLU activation function. The second convolutional layer has a dilation rate of 2, and the third layer has a dilation rate of 4.
2) The decoder part of the model uses transpose convolutional layers to reconstruct the clean image from the encoded representation. It mirrors the architecture of the encoder in reverse. The encoded representation is passed through two transpose convolutional layers, each with decreasing channel sizes of 256 and 128. The final layer uses a 3x3 kernel, sigmoid activation function, and outputs the reconstructed image.
3) The model is trained using the mean squared error (MSE) loss and optimized with the Adam optimizer.

IV. Results and Comparisons

- In our study on denoising, we explored two different architectures to address the problem. We compared both our primary and secondary architectures with the base paper that served as a reference.
- For the purpose of our analysis, we focused on evaluating the performance of the models using a specific noise variance of 25. Additionally, we limited our experimentation to a single dataset, namely BSDS300. This decision was made due to constraints in our computational resources, which restricted the extent of our evaluation.
- By comparing the results of our proposed architectures against the base paper and conducting experiments specifically on the BSDS300 dataset with a variance of 25, we aimed to gain insights into the effectiveness and applicability of our approaches in the context of denoising.
- **Subsequently, the output of the autoencoder was passed through multiple 1x1 convolutions, serving to enhance and fine-tune the reconstructed output. These additional convolutions further refined the representation obtained from the autoencoder, improving the overall quality of the reconstructed images.**

Method	σ^2	PSNR of BSD300
Blind Spot Architecture [1]	25	31.02
Ours (Denoising Autoencoder)	25	30.15
Ours (Blind Spot Dilation Architecture)	25	31.09

TABLE I
Comparisons using PSNR scores

V. Conclusion

Introducing a blind spot dilation network architecture, we present a distinctive approach that eliminates the need for sequential inference and independent processing of directions. In the domain of image denoising, our method achieves results comparable to state-of-the-art techniques across various noise distributions. Notably, our approach surpasses these existing methods by a narrow margin, particularly when the input noise variance aligns with more realistic values.

References

[1] D. Honzátko, S. A. Bigdeli, E. Türetken and L. A. Dunbar, "Efficient Blind-Spot Neural Network Architecture for Image Denoising," *2020 7th Swiss Conference on Data Science (SDS)*, Luzern, Switzerland, 2020, pp. 59-60.