Code Deep Dive

Ramtin Mojtahedi Saffari - 20307293

Assessed link: ConvNet

Self-reflection

- **Short Summary**

The provided link focuses on the idea that convolutional neural networks (CNNs) make use of spatial proximity by establishing a pattern of local connections between neurons in nearby layers of the network. Their design guarantees that the taught "filters" will give the greatest response to a spatially confined input pattern resulting from the learning process. The provided link uses this idea and compares the performance of CNNs and fully convolutional networks (FCNs) on permuted and non-permuted MNIST data. The results show that FCN performs better on permuted and shuffled MNSIT data compared to the CNN structure.

- **Hypothesis and Expectation**

I hypothesize that using a dilated convolutional network and a smaller batch size will improve the accuracy and performance of the proposed CNN network when using permuted data.

- **What I Achieved and Learned**

CNNs can learn local features and are not able to learn from global dependent features. Therefore, the accuracy and performance of the CNN network are being decreased on the permuted input images. However, the FCN network works better than the CNN on the permuted data as FCN can preserve spatial information without using prior knowledge. They are not heavily dependent on the spatial feature or randomization on the pixel-wise randomization [1].

One of the ways to improve the performance of the CNN on the permuted data is using dilated CNN (DCNN) networks. Dilation is a method for increasing the size of the kernel (input) by introducing gaps between the parts of the kernel that are consecutive. The process is similar to convolution in that it skips pixels in order to cover a broader region of the input. The size of the receptive field in a dilated CNN will be determined by the depth of the network [2]. It is preferable to use this approach rather than conventional convolution because:

- Increased receptive field size (i.e., no loss of coverage)

- Exceptionally computationally efficient (as it provides a larger coverage on the same computation cost)
- Memory utilization is reduced (as it skips the pooling step).
- implementation
- There is no reduction in the resolution of the produced picture (as we dilate instead of performing pooling).
- The structure of this convolution aids in the preservation of the data's chronological sequence.

I did two types of analysis, both using FCN, CNN, DCNN on permuted and non-permuted data. The results are as follows:

1- Permuted and non-permuted MNIST Data for CNN and FCN

As shown in the following figure, the Performance of the FCN in the permuted data is significantly less than the Performance of the FCN due to the explanations mentioned above. In the next trial, we will compare the performance of the CNN and DCNN concerning the permuted data, and it is anticipated that the Performance of the DCNN achieved better than the normal CNN.

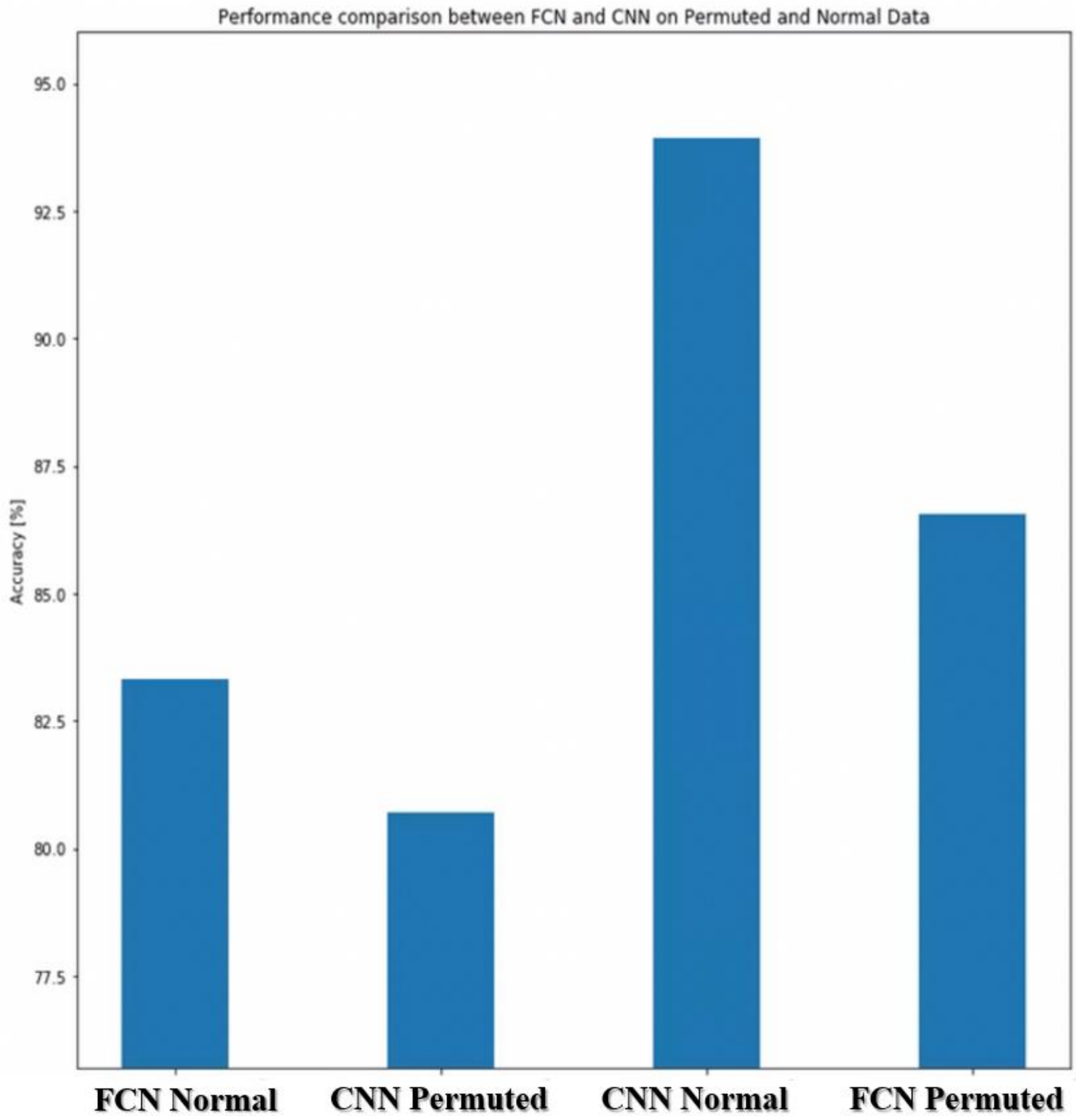

Fig. 1, The Performance of CNN and FCN on normal data (non-permuted)

2- Permuted and non-permuted MNIST data for CNN and DCNN

As shown in the following graph, the DCNN network could improve its performance three times compared to the CNN network. This means that DCNN can learn features that are not spatially dependent on each other. This process needs to be tuned by adjusting the convolution kernel and dilation sizes.

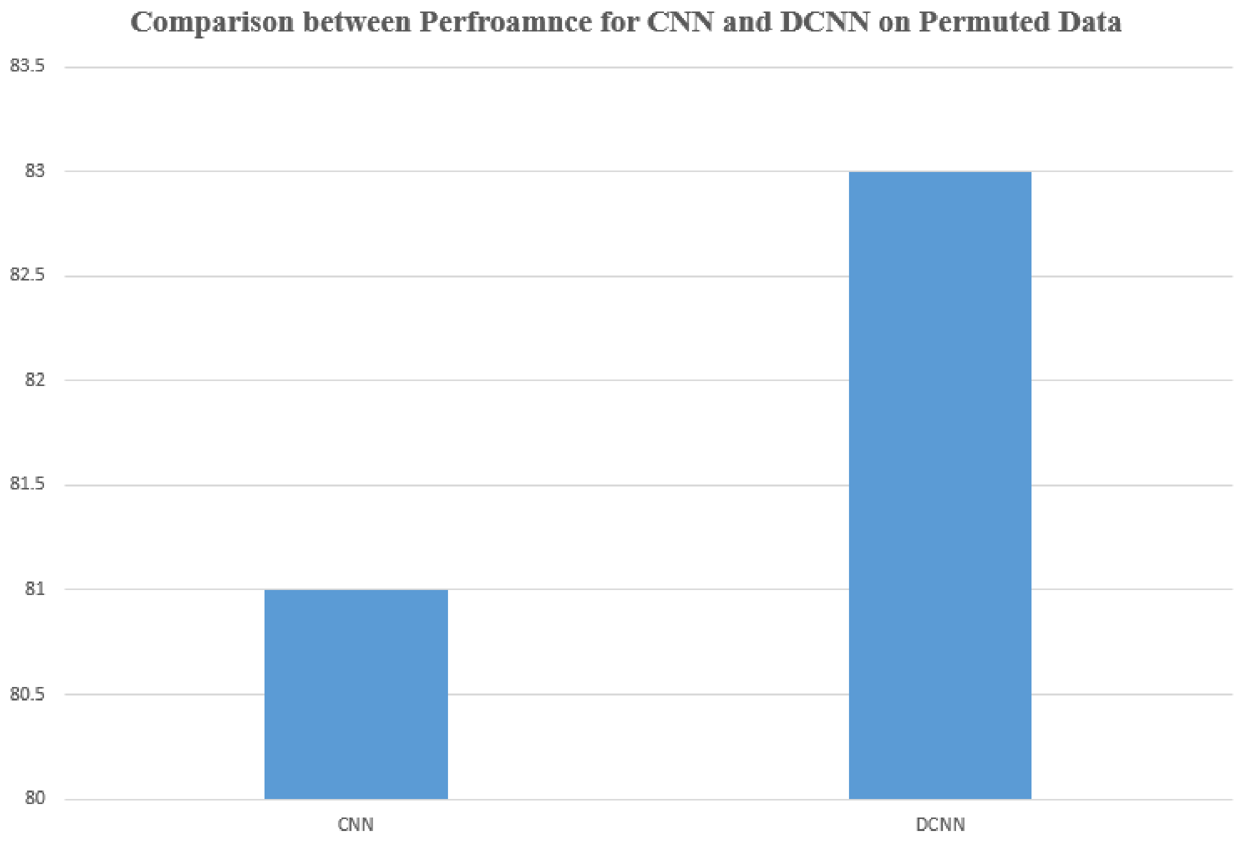

Fig. 2, Comparison between CNN and DCNN Performance on permuted data

Analysis of the batch size

One of the important factors that affect the performance of the network is the batch size. Here, I tested the input images on two different batch sizes, including 64 and 2. Smaller batch size can make a more generalizable model and improve the accuracy. However, it will significantly increase the training time. Smaller batch sizes also provide more up-to-date gradient calculations, which give more stable and reliable training [3]. Figure 3 and Figure 4 show the impact of using smaller batch(BS:2) and larger batch size (BS:64) on the performance and accuracy.

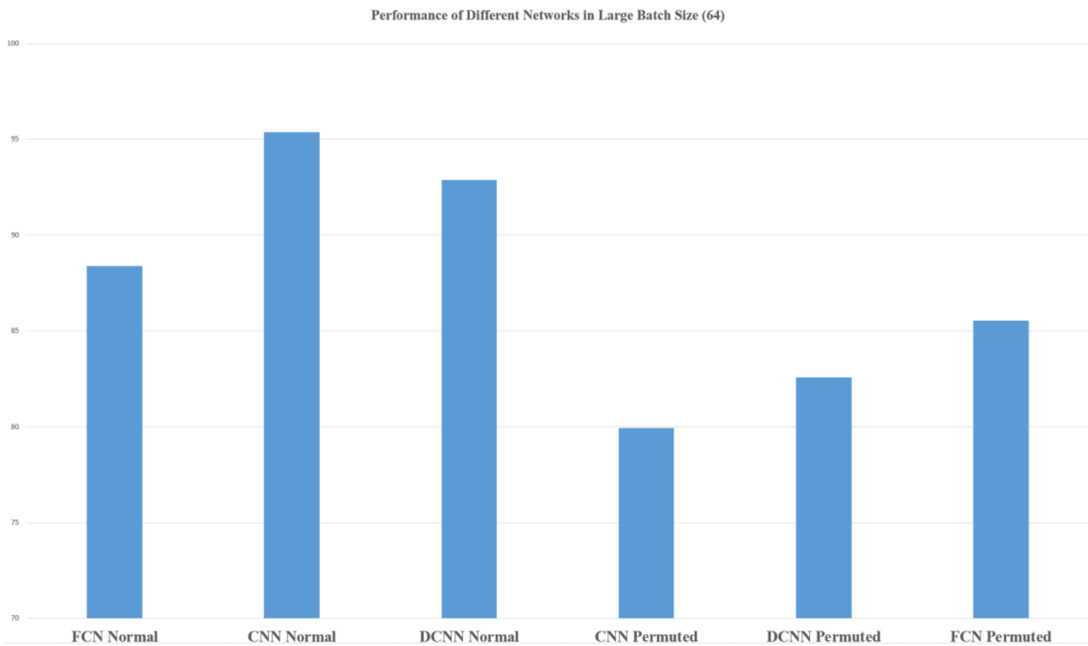

Fig. 3, The performance of different networks in large batch size (64)

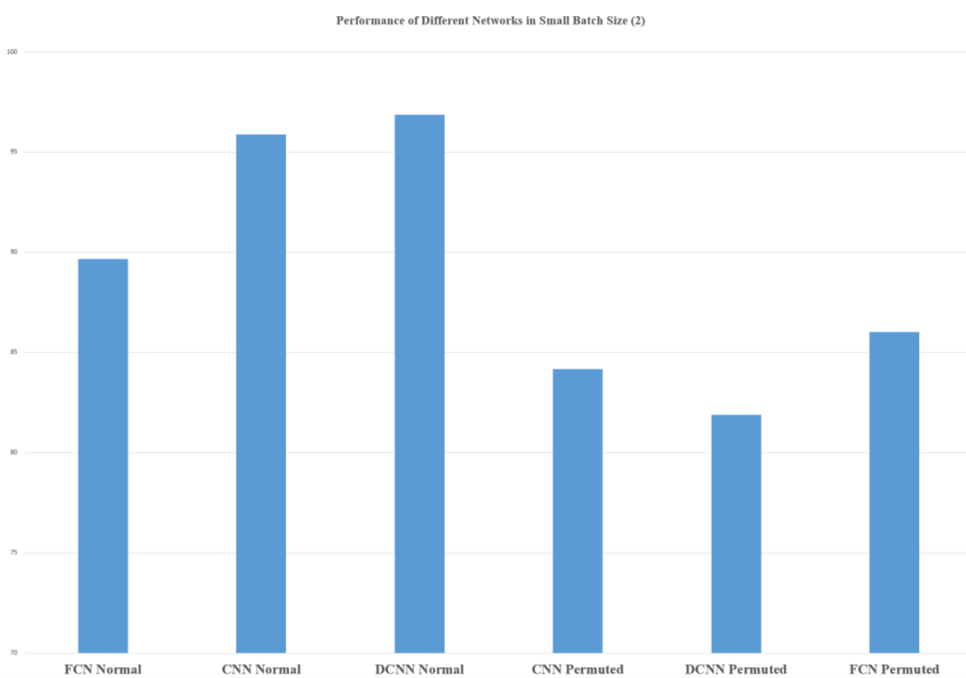

Fig. 4, The performance of different networks in small batch size (2)

Discussion and Future Recommendation

In this report, two studies have been analyzed, including a comparison between the performance of different networks on normal and permuted data and the impact of batch size on performance and accuracy. Considering the results of the two studies, it is found that dilated CNN shows better performance both on normal and permuted input data, which can learn both spatial and long-range features. Also, a smaller batch size could improve the performance and accuracy in all of the studied networks. In terms of having a trade-off between computational cost and Performance, DCNNs are recommended as FCNs are heavy computational and slower than DCNNs.

It is recommended for future analysis to work on using a fusion of transformers and CNNs to improve the network's performance on long-range features and prevent performance reduction when it comes to giving random pixel-wise and permuted input data to the networks.

Suggestion and Filling the Gaps

I found the notebook link very interesting and helpful in learning about the CNN and FCN on normal and permuted data and how they can learn local and long-range features. It is recommended that authors provide some extra information on using transformers for image classification and how they can learn long-range features and, as an example, use [4] as their reference.

Self-evaluation:

In this report, I have gone through an in-depth analysis of one of the provided notebooks. In my assessment, I have completely considered the required expectation for deep exploration, including proposing a hypothesis and what I expected, reporting on what I achieved and explored, in-depth discussion of what I have learned, and providing gaps and recommendations to fill them. Considering the quality and assessment level, I deserve myself to get the full mark (4 points) for this report.

In advance, thank you very much for your time and consideration of this report.

Best regards,

Ramtin

References

[1] Spinelli, I., Scardapane, S., & Uncini, A. (2020). Convolutional Neural Networks on Randomized Data. *Neural Networks*, *129*, 249–260. https://doi.org/10.1016/j.neunet.2020.06.005

[2] Lei, X., Pan, H., & Huang, X. (2019). A Dilated CNN Model for Image Classification. *IEEE Access, 7*, 124087–124095. https://doi.org/10.1109/access.2019.2927169

[3] Shen, K. (2018, June 20). *Effect of batch size on training dynamics - Mini Distill*. Medium. https://medium.com/mini-distill/effect-of-batch-size-on-training-dynamics-21c14f7a716e

[4] *Transformers for Image Recognition at Scale*. (2020, December 3). Google AI Blog. https://ai.googleblog.com/2020/12/transformers-for-image-recognition-at.html