Jenish Maharjan **Week 2**

Paper: *Handwritten Zip Code Recognition with Multilayer Networks*

Summary:

In this paper the authors describe a back-propagation network implemented for recognizing handwritten zip codes. Most widely used methods for pattern recognition implemented a complex feature extractor which need a lot of design efforts and a classifier. The back-propagation network described by the authors improves on this traditional approach and simplifies the feature extraction and achieves good generalization.

The authors used a dataset consisting of segmented numerals from handwritten zip codes supplemented by printed digits as the input for the network. The input images of these numerals were normalized using a linear transformation before feeding them to the network. Unlike the "naïve approach" of using fully connected network for the task at hand, they use locally connected network. Fully connected networks as explained in the 1998 paper discard the spatial organization of the input data which is not optimal in the case of images. The network instead uses "local connections" to detect local features which can then be combined to recognize higher level patterns. The network consists of an input layer, four hidden layers and an output layer. The output layer with ten units is able to identify the numeral in the input image. The first and the third hidden layers act as feature extractors which have shared weights while the second and the fourth layers act as averaging or subsampling layers. The network described in this paper is comparable to the convolutional network described in the 1998 paper.

Convolutional networks ensure to a certain degree an invariance in shift, scale and distortion by implementing the concepts of local receptive fields, shared weights and spatial or temporal subsampling. LeNet-5 is one of the convolutional networks that is used for recognizing characters. Similar to the earlier model, each unit in a layer of the model gets its inputs from a small neighborhood of units in the previous layer. The model has layers that extract features and subsampling layers that reduces the resolution of the feature map resulting in the shift, scale and distortion invariance. The function of subsampling layers can be intuitively understood as "blurring the input". The LeNet-5 network has feature extracting convolutional layers and subsampling layers stacked alternatively. As we move along the network, more abstract and higher-level features are obtained, finally activating one of the units among the ten in the output layer to identify the numeral in the input image. Inspired by Hubel and Wiesel's concept of "simple" and "complex" that was initially implemented by Fukushima's Neocognitron, the presented network is able to implement the convolution/subsampling combination to successfully categorize hand written numerals.

Strengths:

- The content is well organized – starting with the introduction which talks about the motivation for such a model, the following sections do a great job of taking you through the process data preprocessing and the subsequent layers for feature extraction and subsampling.
- The graphics help a lot in intuitive understanding of the network layers and it functions (in both the papers)
- The dataset used in this work is well defined.

Weaknesses:

- Although the graphics in the paper help in intuitive understanding of the network, a more detailed explanation of the internal representations depicted in the images would have been helpful.
- The abstract of the paper can definitely use some rewording to give a better idea of the paper's content.

Confusions:

- What does skeletonization mean?
- Can the same activation functions be used in convolutional layers and subsampling layers?

Discussion Questions:

- What is the role of data normalization in deep learning networks?
- Taking weight sharing into account, how is the back-propagation algorithm modified?
- How are the parameters like number of units in each layer, kernel size decided?