Gradient-based learning applied to document recognition

November 21, 2023

1 Summary

0. Convolutional Neural Networks introduced in the paper are shown to outperform all other techniques in handwritten character recognition. A new learning paradigm called Graph Transformer Networks (that uses Convolutional Neural Networks) allows currently used multi-module systems to be replaced by globally trained Gradient-Based methods that perform better.

1. The paper motivates by citing that the traditional (at this point in 1998) machine learning method of first extracting feature vectors from raw inputs and then training classifiers is ineffective. Availability of large databases, computational power, and better learning methods are also cited as motivations for newer approaches.

2. Convolutional Neural Networks (also called LeNet-5) combine three architectural ideas to ensure some of degree of shift, scale, and distortion invariance: *local receptive fields*, *shared weights*, and spacial or temporal *sub-sampling*.

3. The LeNet-5 algorithm works as follows: A *receptive field* is parsed over a small area of a previous layer's plane to generate an output unit in the next layer's plane (called *feature map*). This trainable receptive field when parsed over the image captures the same 'feature' in different parts of it.

4. A convolutional layer is followed by a sub-sampling layer in which the planes are shrunk by a factor of 2. The paper justifies this by saying "reducing the resolution of the feature map and reducing the sensitivity of the output to shifts and distortions".

5. After alternating with convolutional and sub-sampling layers the algorithm ends with full connected layers. The paper goes on to show that this performs better than existing algorithms in every metric and it's power only increases with training data volume and better computational power.

6. In later part of the paper, they apply this algorithm in document recognition using multi-modular Graph Transformer Network. We ignore this part as it has not been influential in later years.

2 Analysis

The LeNet-5 is a seminal paper in computer vision. The design philosophy in coming up with the idea is an interesting aspect to learn from this paper. Specifically this paper has shown that ideas like local connections and to preserve properties like shift invariance can be successfully incorporated into design.