Jenish Maharjan **Week 8**

Paper: *Siamese Neural Networks for One-shot Image Recognition*

Summary:

In this paper, the authors describe their approach of one-shot learning using Siamese neural networks to learn image representations via a supervised metric-based approach. In the task of classification, one-shot learning is when the model is allowed to only observe a single example of each possible class before making a prediction about a test instance. Models that implement one-shot learning tend to overfit to the training set and work great with similar instances but do not offer a robust solution. They implement Siamese convolutional neural networks which are capable of learning generic image features useful for making predictions about unknown class distributions, can be easily trained using standard optimization techniques on pairs sampled from the source data and provide a competitive approach that does not rely upon domain-specific knowledge.

In order to develop a model for one-shot image classification they first aimed to learn a neural network that can discriminate between the class-identity of image pairs and then to generalize the networks that will do well at verification stage will to one-shot classification. The model consists of a sequence of pairs of Convolutional Layers and Max pooling layers. They use the ReLU activation function for L-2 layers. The number of filters in a convolution layer is in multiple of 16. The convolution layers are followed by a final fully connected layer followed by a distance layer that computes the induced distance metric between each Siamese twin. The output of the layer is the probabilistic value of the 2 twins which tells whether they are similar or not. They use a mini-batch of size M with label vectors whose value is 1 when the twins are same images and 0 when they are different. Since the weights of the two twins are tied, the gradient in case of Siamese neural network is additive for the twin network and optimization is done by standard back-propagation. The authors note that the decaying of the model was uniform across the network even though different learning rates were applied for each layer. For hyper-parameter optimization, a Bayesian optimization framework was used.

For the experiments, they used the Omniglot dataset that contains 50 alphabets from well known languages to lesser known local dialects. The dataset is split into background set which is used for developing model by learning hyper-parameters and feature mapping and evaluation set which is used to measure one shot classification performance. The verification on the dataset was done by dividing the training set and forming additional copies using affine transformations. The approaches they implemented to monitor performances during training were: creating validation set for verification and generating 320 one-shot recognition trials. They also used the model trained on the Omniglot dataset on the MNIST dataset. The digits in the MNIST dataset were treated as an alphabet and was evaluated as a 10-way one-shot classification task. They also used a nearest neighbor approach which gave similar results to that found on the Omniglot dataset. Although the performance results of the Siamese NN model on MNIST dropped, it was still better than the nearest neighbor approach.

The approach presented in this paper is great for limited datasets and it can also be modeled on one set of images for a particular classification. They present comparative results that show that their metric learning approach is capable of human-level accuracy and that this approach should extend to one-shot learning tasks in other domains as well.

Strengths:

- The content of the paper is well organized and has a good flow to it.
- The paper does a good job of introducing Siamese networks and one-learning to readers who might be new to these concepts.

Weaknesses:

- They could have talked a little about some possible application areas of one-shot learning.

Confusions:

- What does "gradient is additive across the twin networks" mean?
- Is the distance vector their "supervised metric"?

Discussion Questions:

- Would one-shot learning still work if the image of an instance class in the training set is different in orientation than in the test set? Or is image augmentations of the dataset enough to overcome such an issue?
- They say that "the weights were drawn from a much wider normal distribution". What would be the intuition behind this?
- To measure the performance, they've performed 40 one-shot learning trials for each of the ten evaluation objects. How well does this evaluate the model's performance?