Siamese Neural Networks for One-Shot Image Recognition

by

Gregory Koch

Abstract

Siamese Neural Networks for One-Shot Image Recognition

Gregory Koch
Master of Science
Graduate Department of Computer Science
University of Toronto
2015

The process of [illegible])28(ery)-420(com)-1(c of each new class.

In this paper, we explore a method for r *siamese neur networks* which employ a unique

Contents

Chapter 1

Introduction

1.1 Overview

Figure 1.1: Learning discriminative features about tigers should help when classifying other felines that are unfamiliar to the model.

The problem of one-shot learning can be directly addressed by developing domain-speci c features or inference procedures which possess highly discriminative properties for the target task. As a result, systems which incorporate these methods tend to excel at similar instances but fail to o er robust solutions that may be generally applied to other types of problems. In this paper, we present a novel

More recently, Lake et al. approached the problem of one-shot learning from the point of view of cognitive science, addressing one-shot learning for character recognition with a method called Hierarchical Bayesian Program Learning (HBPL) [16]. In a series of several papers, the authors modeled the process of drawing characters generatively to decompose the image into small pieces [14, 15]. The goal of HBPL

Figure 1.2: Our general strategy. 1) Train a model to discriminate between a collection of same/di erent

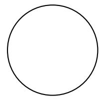
By additionally choosing a good metric at nal layer of the neural network and then imposing certain

Chapter 2

The Omniglot Dataset

2.1 Overview

The Omniglot data set was collected by Brenden Lake and his collaborators at MIT via Amazon's Mechanical Turk to produce a standard benchmark for learning from few examples in the handwritten character recognition domain [14].[1] Omniglot contains examples from 50 alphabets ranging from well-established international languages like Latin and Korean to lesser known local dialects. It also includes



3.3 Learning

Loss function. Let M represent the minibatch size, where i indexes the ith minibatch

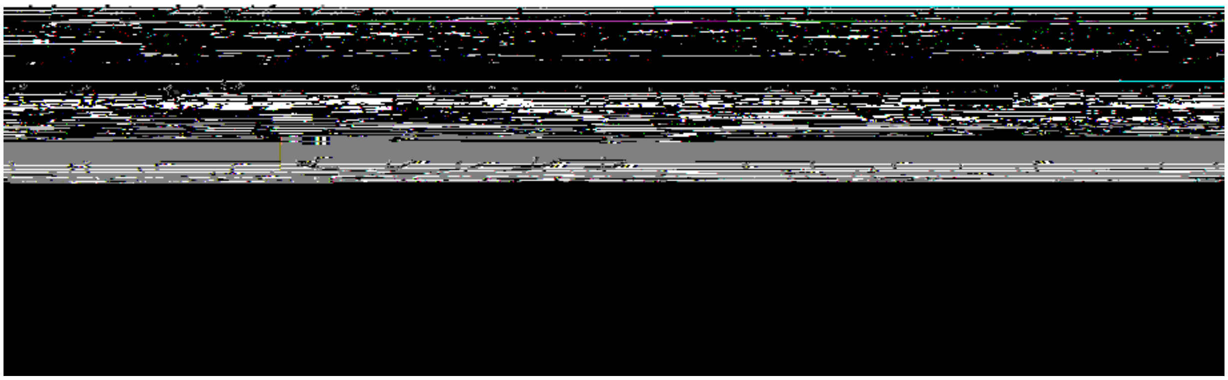


Figure 3.2: A sample of random a ne distortions generated for a single character in the Omniglot data set.

A ne distortions. In addition, we augmented the training set with small a ne distortions (Fig-

Chapter 4

Learning Deep Convolutional Feature Hierarchies

4.1 Overview

So far, we have restricted ourselves to a standard feedforward architecture using tied weights on each side of the siamese network, but with all neurons attached indiscriminatively without any local connectivity. Here, we describe a straightforward extension of the model presented in the previous section to include *convolutional* layers. Convolutional neural networks have achieved top-level results in many large-scale computer vision applications, particularly in image recognition tasks [2, 11, 21, 22]. Several factors make convolutional networks especially appealing:

1. Local connectivity can greatly reduce the number of parameters in the model, which inherently

Figure 4.1: Best convolutional architecture selected for veri cation task. Siamese twin is not depicted,

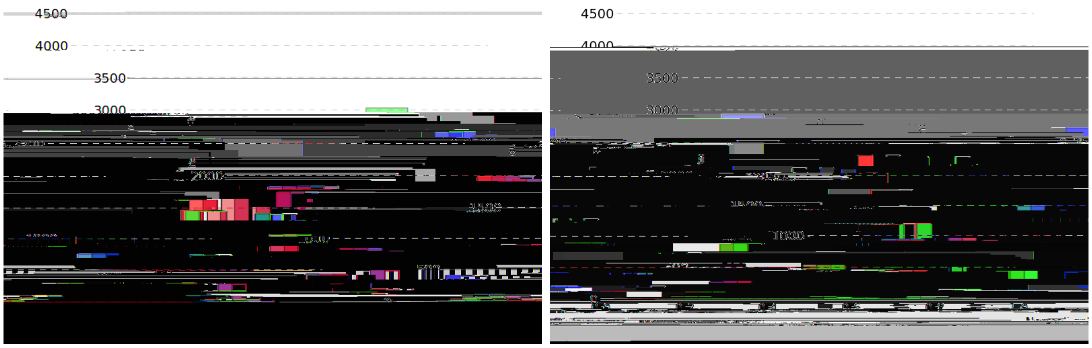


Figure 4.3: **Left**: Before training. Initially the distance is roughly in the middle depending on the weight initialization of the convolutional network, but there is no clear pattern to the distances between the computed siamese twin feature vectors. **Right**: After training. (blue: y=1, same; red: y=0, di erent). Note that since this distance is weighted and activated by a sigmoid, the network is only indirectly motivated to fully separate these distributions.

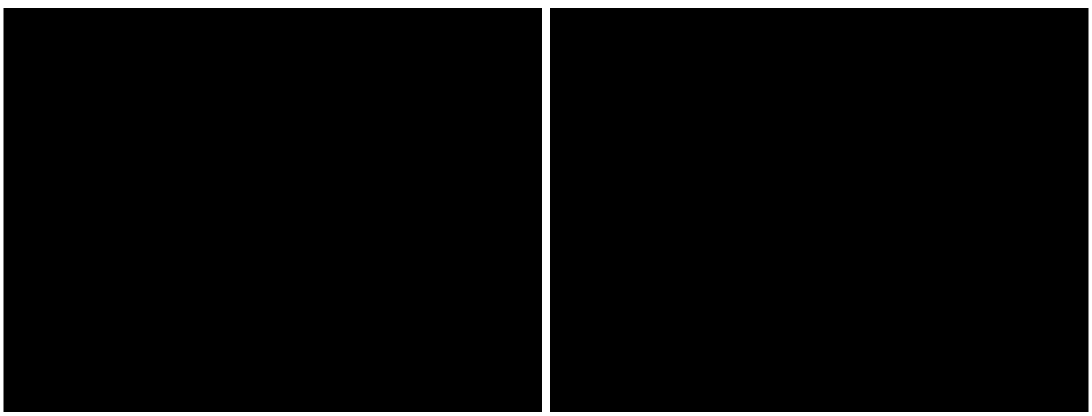


Figure 4.4: **Left**: Before training. **Right**: After training. Once optimized, the veri cation network not only achieves high accuracy but exhibits a considerable degree of con dence in its predictions on the test set (blue: y=1, same; red: y=0, di erent).

	Test (6-layer)
30k training	
no distortions	90.61
a ne distortions x8	91.90
90k training	
no distortions	91.54
a ne distortions x8	93.15
150k training	
no distortions	91.63
a ne distortions x8	**93.42**

Table 4.1: Accuracy on Omniglot veri cation task (siamese convolutional neural net)

Chapter 5

Using Veri cation Networks for One-Shot Image Recognition

5.1 Overview

Once we have optimized a siamese network to master the veri cation task, we are ready to demonstrate the discriminative potential of our learned features at one-shot learning. Recall that given a pair of images, our siamese network outputs a prediction p

Figure 5.2: One-shot validation and evaluation accuracy measured over the course of training a deep convolutional siamese network on 30,000 training examples with 8x a ne distortions on the inputs. We also include the training and validation curves for veri cation.

Our non-convolutional method fares no better than the majority of the other methods, but at 92 percent our convolutional method is stronger than any model except HBPL itself. which is only slightly

Chapter 6

Conclusions

6.1 Summary

We have presented a strategy for performing one-shot classification by first learning deep convolutional siamese neural networks for verification. We outlined new results comparing the performance of our networks to an existing state-of-the-art classifier developed for the Omniglot data set. Our networks outperform all available baselines by a significant margin and come close to the best numbers achieved

Bibliography

Appendix A

Data Generation

```
function generateOneshotTrials()
  data J/F58 9.9626 Tf 24.837 0 Td [(data)]TJ/F8 9.9626 Tf 24.243 2a
```