Siamese Neural Networks for One-shot Image Recognition

Gregory Koch Richard Zemel Ruslan Salakhutdinov

Department of Computer Science, University of Toronto. Toronto, Ontario, Canada.

Introduction

- Siamese neural networks which employ a unique structure to **naturally rank similarity between inputs.**
- Powerful discriminative features to generalize the predictive power of the network not just to new data, but to entirely new classes from unknown distributions.

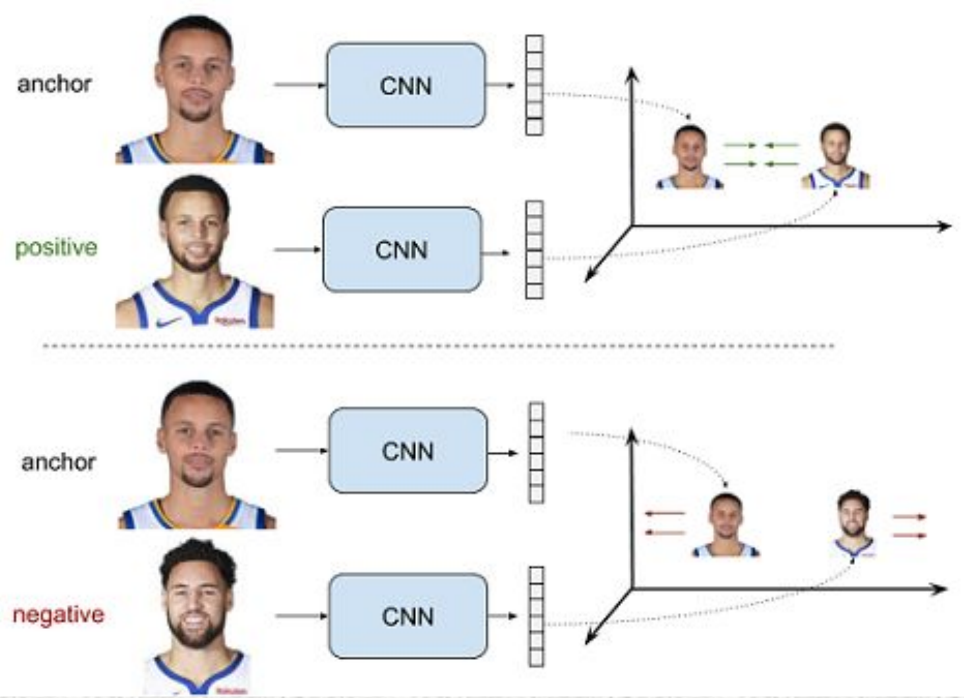
Application

Issue : Machine learning has been successfully used to achieve state-of the- art performance in a variety of applications such as web search, spam detection, caption generation, and speech and image recognition.

Limitation of Data

one-shot learning

- Proposed by (Fei-Fei et al., 2006; Lake et al., 2011).
- Only observe a single example of each possible class before making a prediction about a test instance
- If degree is difference
- d(img1,img2) <= τ
- d(img1,img2) >= τ



Siamese convolutional neural networks

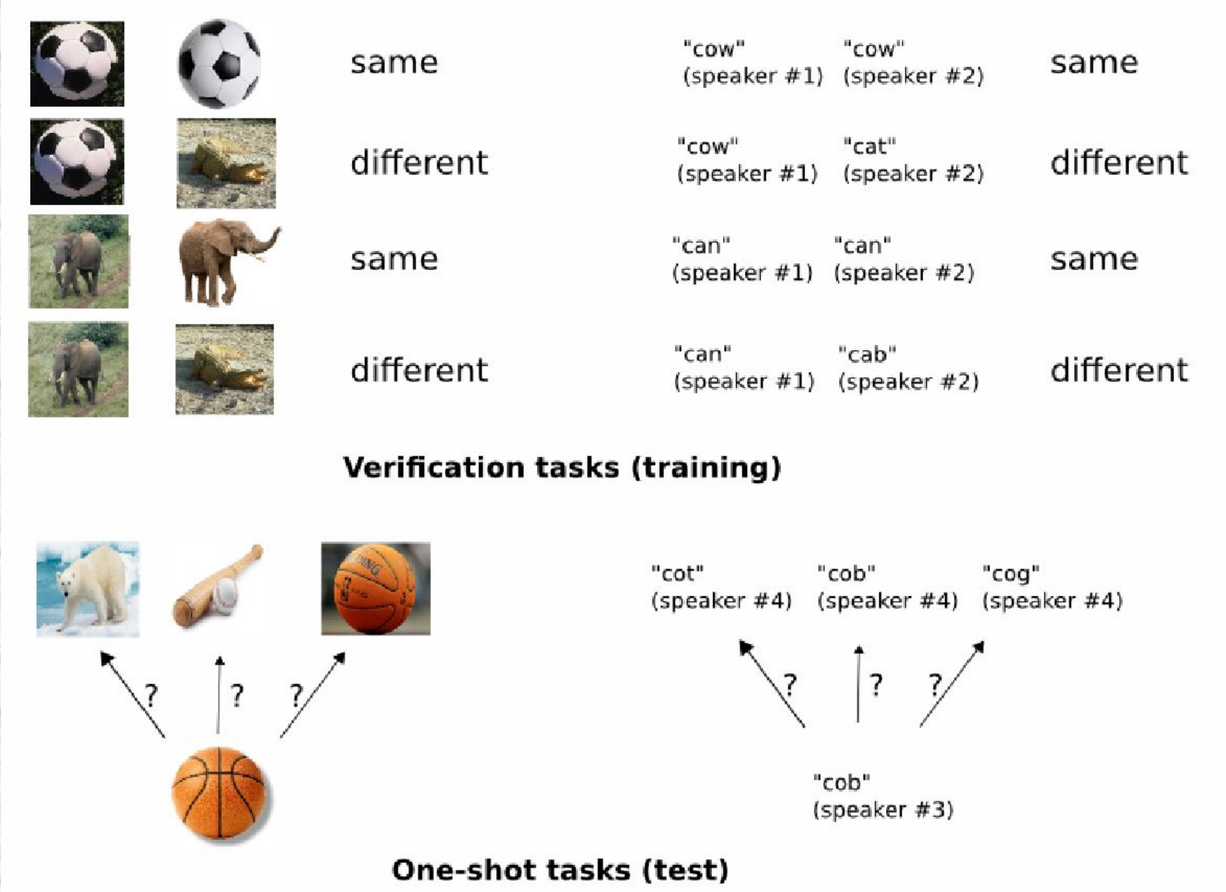
- Capable of learning generic image features useful for making predictions about unknown class distributions even when very few examples
- Easily trained using standard optimization techniques on pairs sampled from the source data
- Provides a competitive approach that does not rely upon domain-specific knowledge

Process

- Create a neural network that can discriminate between the class-identity of image pairs.

  The verification model learns to identify input pairs based on probability that they belong to the same class or different classes

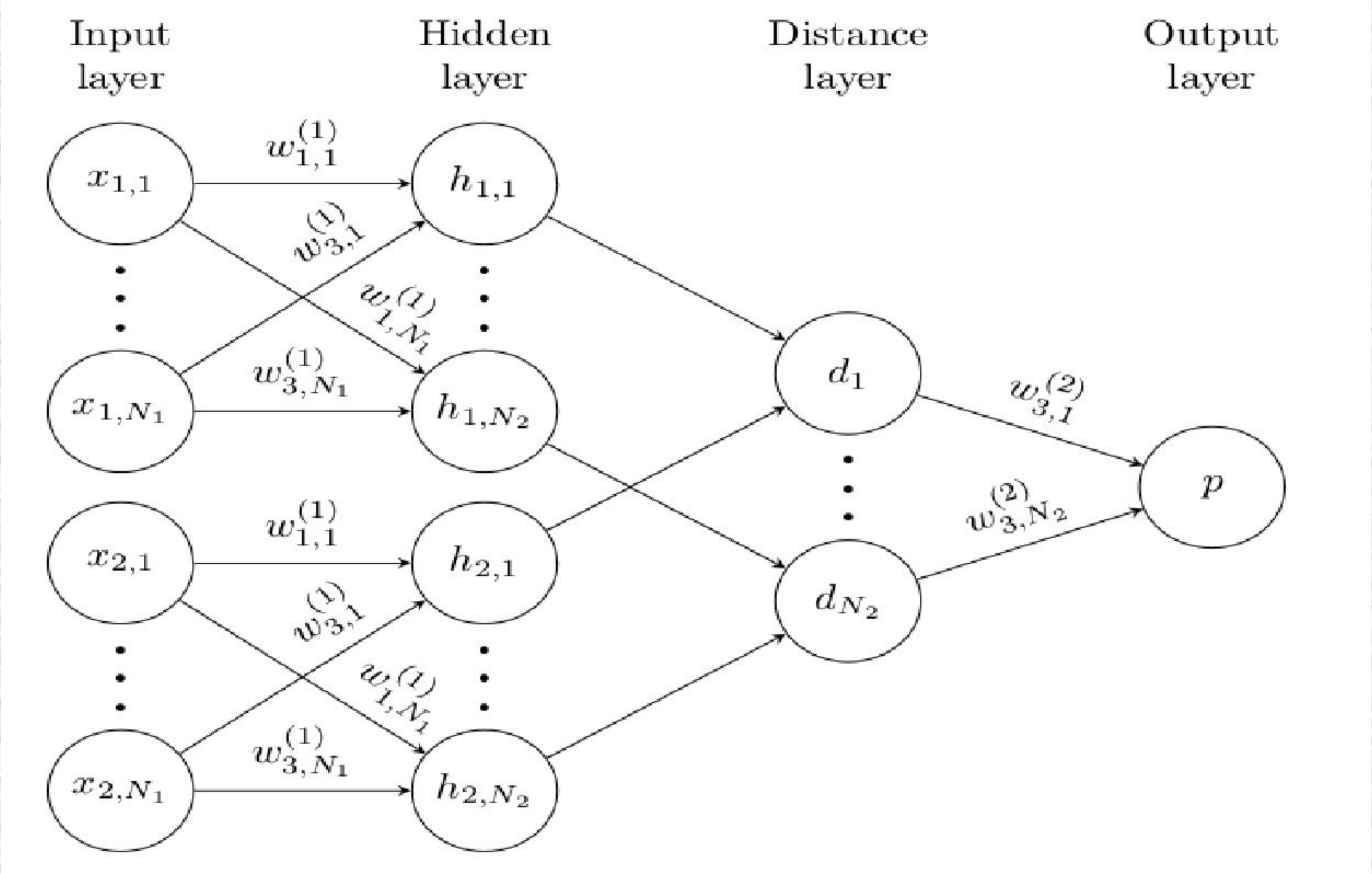
- Then model evaluates new images, exactly one per novel class, in a pairwise manner against the test image.



Related work

- **Bayesian framework** for one shot image classification using the premise that **previously learned classes can be leveraged to help forecast future ones** when very few examples are available from a given class.
- **Hierarchical Bayesian Program Learning:** by **drawing characters generatively to decompose the image into small pieces**. To determine a structural explanation for the observed pixels, inference under HBPL is difficult since the joint parameter space is very large, leading to **an intractable integration problem.**

Deep Siamese Networks for Image Verification



A simple 2 hidden layer Siamese network for binary classification with logistic prediction p.

The structure of the network is replicated across the top and bottom sections to form twin networks, with shared weight

Deep Siamese Networks for Image Verification

- Twin networks which accept distinct inputs but are joined by an energy function at the top.
- The function computes some metric between the highest-level feature representation on each side.
- Parameters between the twin networks are tied.

Why?

So that two extremely similar images could not possibly be mapped by their respective networks to very different locations in feature space because each network computes the same function.

Deep Siamese Networks for Image Verification

- Previously contrastive energy function was used which contained dual terms to decrease the energy of like pairs and increase the energy of unlike pairs.
- This paper used weighted L1 distance between the twin feature vectors h1 and h2 combined with a sigmoid activation, which maps onto the interval [0; 1].
- Best-performing models use multiple convolutional layers before the fully-connected layers and top-level energy function.

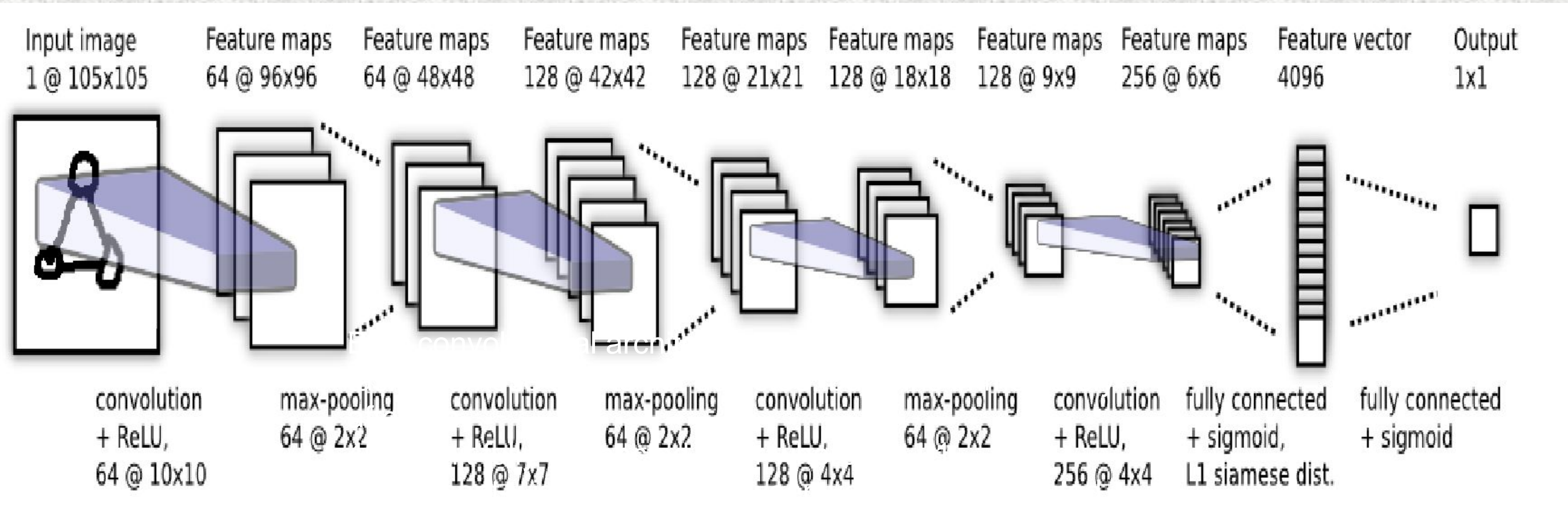
Model

- L layers each with Nl units
- Where h1,l represents the hidden vector in layer l for the first twin, and h2,l denotes the same for the second twin.
- Rectified linear (ReLU) units is used in the first $L-2$ layers and sigmoidal units in the remaining layers.
- Model consists of a sequence of convolutional layers, each of which uses a single channel with filters of varying size and a fixed stride of 1

Model

- Number of convolutional filters is specified as a multiple of 16 to optimize performance.
- It applies a ReLU activation function to the output feature maps, optionally followed by maxpooling with a filter size and stride of 2.

$$a_{1,m}^{(k)} = \text{max-pool}(\max(0, \mathbf{W}_{l-1,l}^{(k)} \star \mathbf{h}_{1,(l-1)} + \mathbf{b}_l), 2)$$
$$a_{2,m}^{(k)} = \text{max-pool}(\max(0, \mathbf{W}_{l-1,l}^{(k)} \star \mathbf{h}_{2,(l-1)} + \mathbf{b}_l), 2)$$



Best convolutional architecture selected for verification task.

Siamese twin is not depicted, but joins immediately after the 4096 unit fully-connected layer where the L1 component-wise distance between vectors is computed

Learning

- Loss function: M= minibatch size,
- i indexes the ith minibatch

- $\mathbf{y}(x_1^{(i)}, x_2^{(i)})$ = a length-M vector which contains the labels for the minibatch
- The value is assumed to be 1 when x1 and x2 are from the same character class and zero otherwise.
- a regularized cross-entropy objective on a binary classifier of the following form is imposed

$$\mathcal{L}(x_1^{(i)}, x_2^{(i)}) = \mathbf{y}(x_1^{(i)}, x_2^{(i)}) \log \mathbf{p}(x_1^{(i)}, x_2^{(i)}) +$$
$$(1 - \mathbf{y}(x_1^{(i)}, x_2^{(i)})) \log (1 - \mathbf{p}(x_1^{(i)}, x_2^{(i)})) + \boldsymbol{\lambda}^T |\mathbf{w}|^2$$

Model

- Optimization :

  -Objective is combined with standard backpropagation algorithm

  -The gradient is additive across the twin networks due to the tied weights

$$\mathbf{w}_{kj}^{(T)}(x_1^{(i)}, x_2^{(i)}) = \mathbf{w}_{kj}^{(T)} + \Delta\mathbf{w}_{kj}^{(T)}(x_1^{(i)}, x_2^{(i)}) + 2\lambda_j |\mathbf{w}_{kj}|$$

$$\Delta\mathbf{w}_{kj}^{(T)}(x_1^{(i)}, x_2^{(i)}) = -\eta_j \nabla w_{kj}^{(T)} + \mu_j \Delta\mathbf{w}_{kj}^{(T-1)}$$

Model

- **Weight initialization:** initialized all network weights in the convolutional layers from a normal distribution with zero-mean and a standard deviation of 10^2.
- Biases were also initialized from a normal distribution, but with mean $0{:}5$ and standard deviation 10^2.
- In the fully-connected layers, the biases were initialized in the same way as the convolutional layers
- The weights were drawn from a much wider normal distribution with zero-mean and standard déviation $2 \cdot 10$^-1.

Model

- Learning schedule
- Hyperparameter optimization
- Affine distortions

Experiment

Method	Test
30k training	
no distortions	90.61
affine distortions x8	91.90
90k training	
no distortions	91.54
affine distortions x8	93.15
150k training	
no distortions	91.63
affine distortions x8	**93.42**

Table 1. Accuracy on Omniglot verification task (Siamese convolutional neural net)

Conclusion

The paper presented a strategy for performing one-shot classification by first learning deep convolutional Siamese neural networks for verification.

It outperforms all available baselines by a significant margin and come close to the best numbers achieved by the previous author

Thank you! ☺