Siamese Neural Networks for One-shot Image Recognition

Alexander Smirnov

12.11.2019

Problem

- Learning good features computationally expensive
- Little data is available

Solution overview

- Siamese neural networks structure to rank similarity between inputs
- Once a network has been tuned it may be applied not just to predict on new data but to entirely new classes
- Correctly making predictions given only a single example of each new class

Example

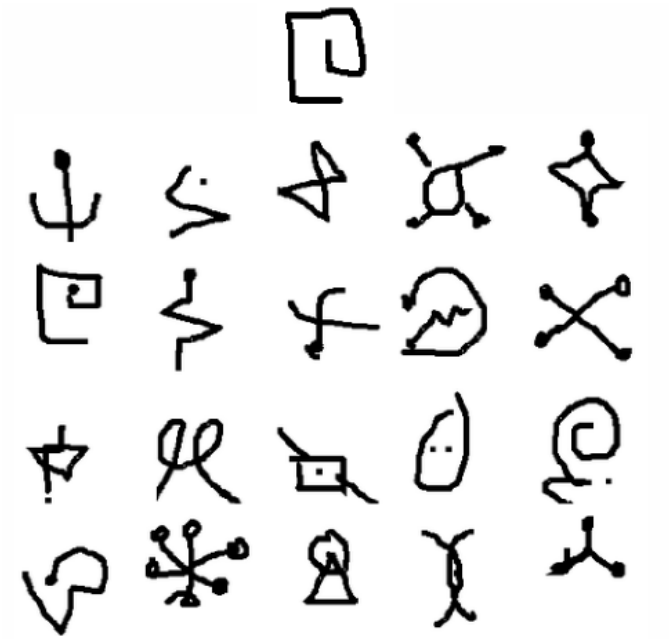


Figure: Example of a 20-way one-shot classification task using the Omniglot dataset. The lone test image is shown above the grid of 20 images representing the possible unseen classes that we can choose for the test image. These 20 images are our only known examples of each of those classes

Plan

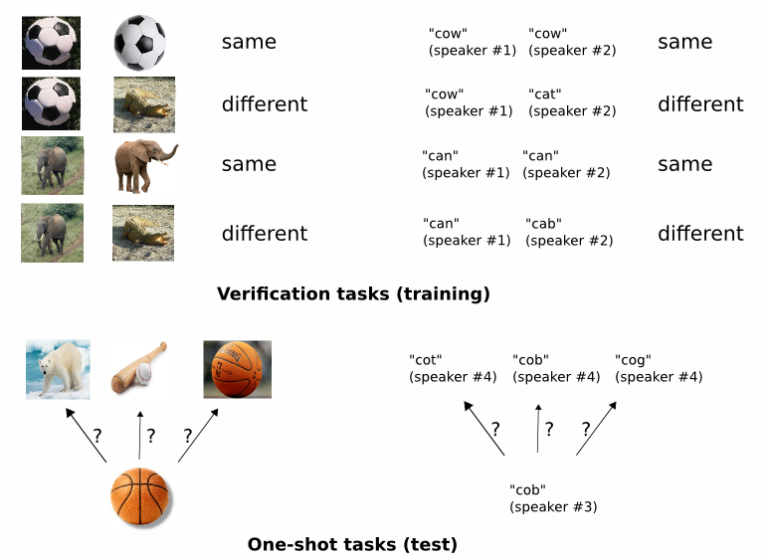


Figure: General strategy.

1. Train a model to discriminate between a collection of same/different pairs
2. Generalize to evaluate new categories based on learned feature mappings for verification

Approach

- Learn a neural network that can discriminate between the class-identity of image pairs (verification task)
- The verification model learns to identify input pairs according to the probability that they belong to the same class or different classes
- This model can then be used to evaluate new images, exactly one per novel class, in a pairwise manner against the test image
- The pairing with the highest score according to the verification network is then awarded the highest probability for the one-shot task

Model

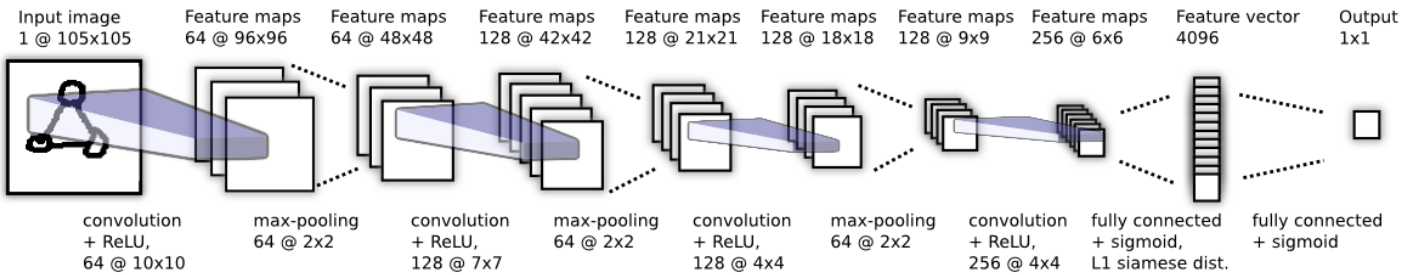


Figure: Best convolutional architecture selected for verification task. Siamese twin joins immediately after the 4096 unit layer where the L1 component-wise distance between vectors is computed

Learning

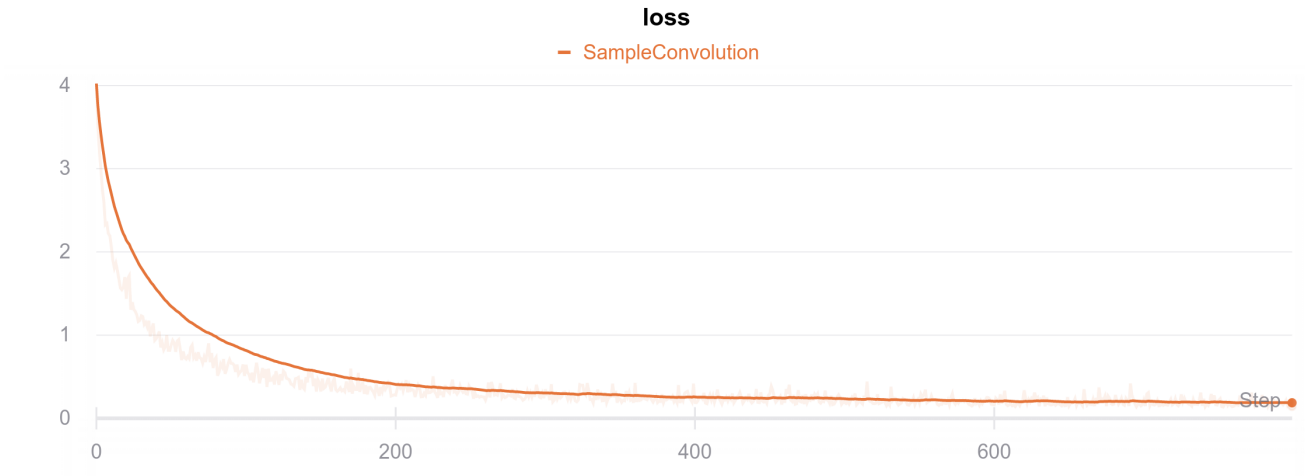
- Loss function
  - Regularized cross-entropy
- Weight initialization
  - Weights in the convolutional layers initialized from a normal distribution with zero-mean and a standard deviation of 10^{-2}
  - Biases were also initialized from a normal distribution, but with mean 0.5 and standard deviation 10^{-2}
- Schedule
  - 20000 iterations
  - 20 — support set quantity
  - 30 — number of validations

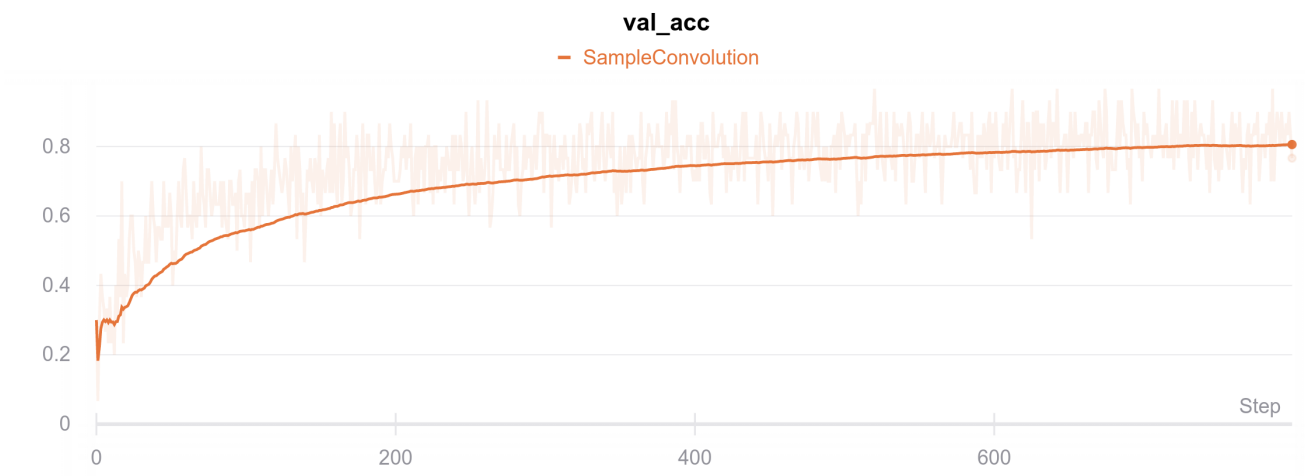
The Omniglot Dataset

Figure: The Omniglot dataset contains a variety of different images from alphabets across the world.

- The number of letters in each alphabet varies considerably from about 15 to upwards of 40 characters
- All characters across these alphabets are produced a single time by each of 20 drawers
- Splitting the data into a 30 alphabet training set and a 20 alphabet evaluation set

Training


loss
SampleConvolution
4
3
2
1
0
0
200
400
600
Step


val_acc
SampleConvolution
0.8
0.6
0.4
0.2
0
0
200
400
600
Step

Comparison

Figure: Comparing our approach to nearest neighbour approach

Conclusions

- We got about 81% average accuracy on evaluation set on 20-way one-shot classification

Links

- Original paper – https://www.cs.cmu.edu/~rsalakhu/papers/oneshot1.pdf
- Implementation – https://github.com/SmirnovAlexander/OneShotLearningSiameseNetworks