Prototypical one-shot and k-shot learning on the Omniglot dataset

(with a general application to non-parametric learning via high-dimensional Euclidean embeddings)



Stanislav Fort, sfort1@stanford.edu, CS 231N, Spring 2017

One-shot learning

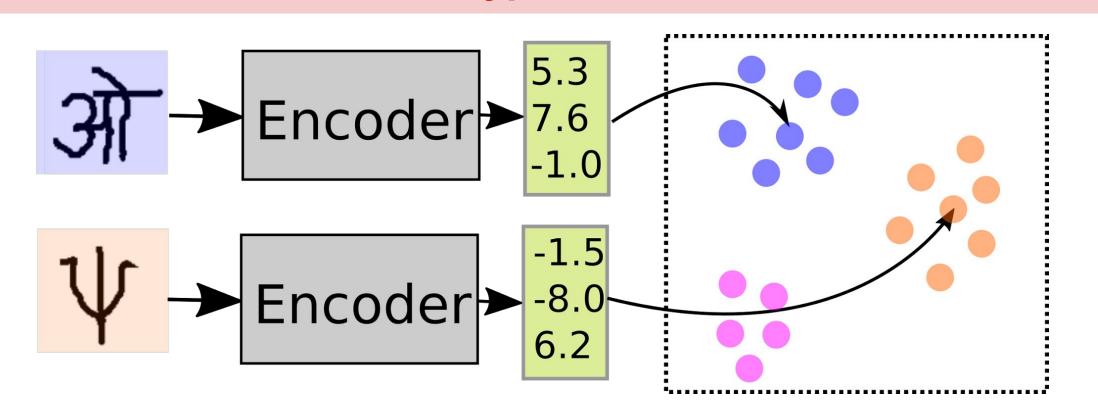
- classification on new classes not seen during training
- hallmark of human learning
- extends deep learning to data-poor problems
- non-parametric → avoids overfitting on small datasets
- protypical learning inspired by cognitive science and

Omniglot and data

- 1623 characters, 50 alphabets, 20 examples of each
- images 105 x 105 x 1 reduced to 28 x 28 x 1
- data augmentation by rotations of 90° → 6492 characters
- 4800 characters for training, 1692 for test

I TOPTEKEKA CAM INTPHICK 三川川氏のかやはちゃるのかららわかるのとはそ名 タストロニ1ココイトのロロアングトロトコマトロト) 434966189me1369126 > 6 >> 9m61369 圆山水小平田山山田本山东西安北西安江田湖日江江 विभ्राभ्रुग्डन्डन्स्टरद्वस्ट्उम्चम्त्रभ्दप्रविभ्र पित्रहर्भिण्टिमिष्डिमिष्ठिमेष्टल्य कुण्णाश्च ०००० क्त है भे ते के कि खे धे धे छे क्र कि ते ते ते ते वे धे धे चे छे क्र के ते ते ते वे धे धे चे छे के क्र के ते हैं ひ み る ひ み な な 了 口 口 口 甘 申 申 申 旦 足 肛 円 臼 両 月 目 助 掛 കഗപടെ ഥധഠഢഛഅസിഖ എമനണതെ ബിവ എയറ ഫസ ഷ ശ 200日至四丁丁七年十万万年7月世五日五年八四日日600 としているといっているととにってとしているといっと БВГАЕЖ5 3 7 Н І К Л М N О П Р Г Т ОУ 8 Ф Х U 百万百月双瓦双环电对对不不万式状()) 1000 Lampy 4 p ray of a common of a proposition of a common of a company of a company of a common of a company of 口行为己中中的阿马斯马斯马斯马斯马里的人名西西马

Protypical model



Encoder: Image → CNN encoder → Euclidean embedding Classification and training: (as described in arXiv 1703.05175)

- # of Support images embedded → their mean = prototype
- # of **query** images embedded
 - their embeddings compared to prototypes → **Euclidean distances**
 - the closest prototype determines the predicted class

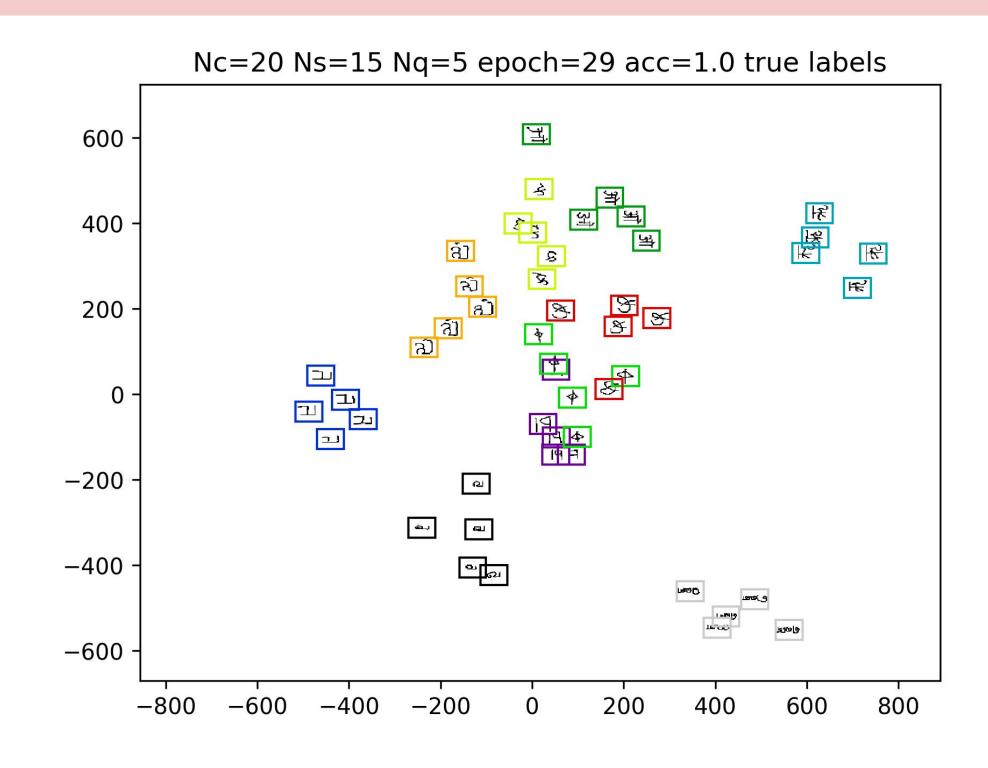
Encoders:

 $[3x3x64 \rightarrow Batch Norm \rightarrow ReLU \rightarrow 2x2 max pool] x4 \rightarrow 64-dim embed Loss = cross entropy on distances to prototypes Optimizer = Adam with learning rate 1e-3$

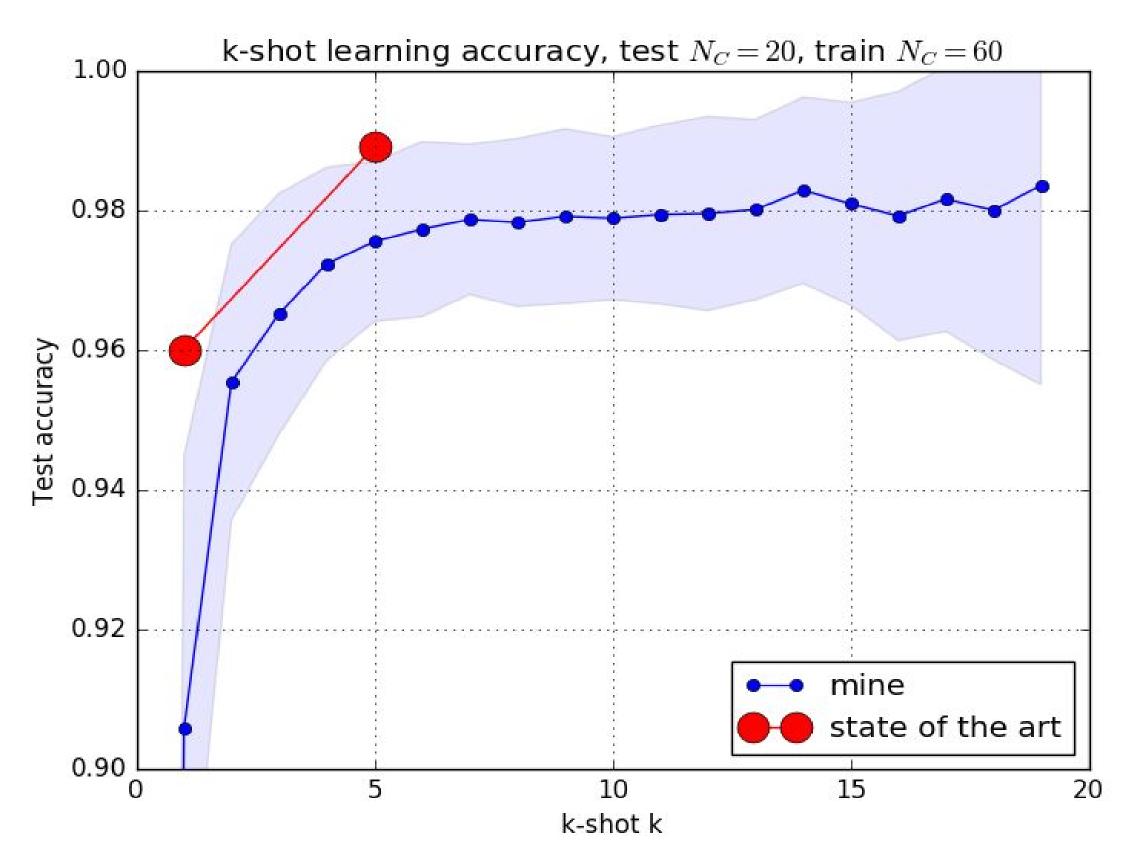
Also experimented with:

- fully-connected → embed (lack spatial awareness)
- CNNs → fully-connected → embed (similar performance)
- embedding sizes, norms (cos, Ln) Euclidean works the best

Results



k-shot learning performance for a model trained on 15 support points and 60 classes per batch, and tested on 20 classes per batch - within error of state of the art



	1-shot	5-shot	19-shot
mine	90.57 ± 3.90 %	97.55 ± 1.10 %	98.35 ± 2.86 %
1703.05175	96.0 %	98.9 %	not done

Conclusion

- prototypical learning works very well on Omniglot
- I was able to reach near state-of-the-art performance on the dataset
- encoder learns useful representation for clustering in the Euclidean embedding
- Euclidean distances work the best and are easily adaptable to new classes
- Will use the method on CIFAR-100 and ImageNet

Examples of characters from the test part of Omniglot.