A Brief Introduction to Few-shot Classification

Shaoxiong Ji

Problem Definition

Problem Definition

- N-way one-shot learning

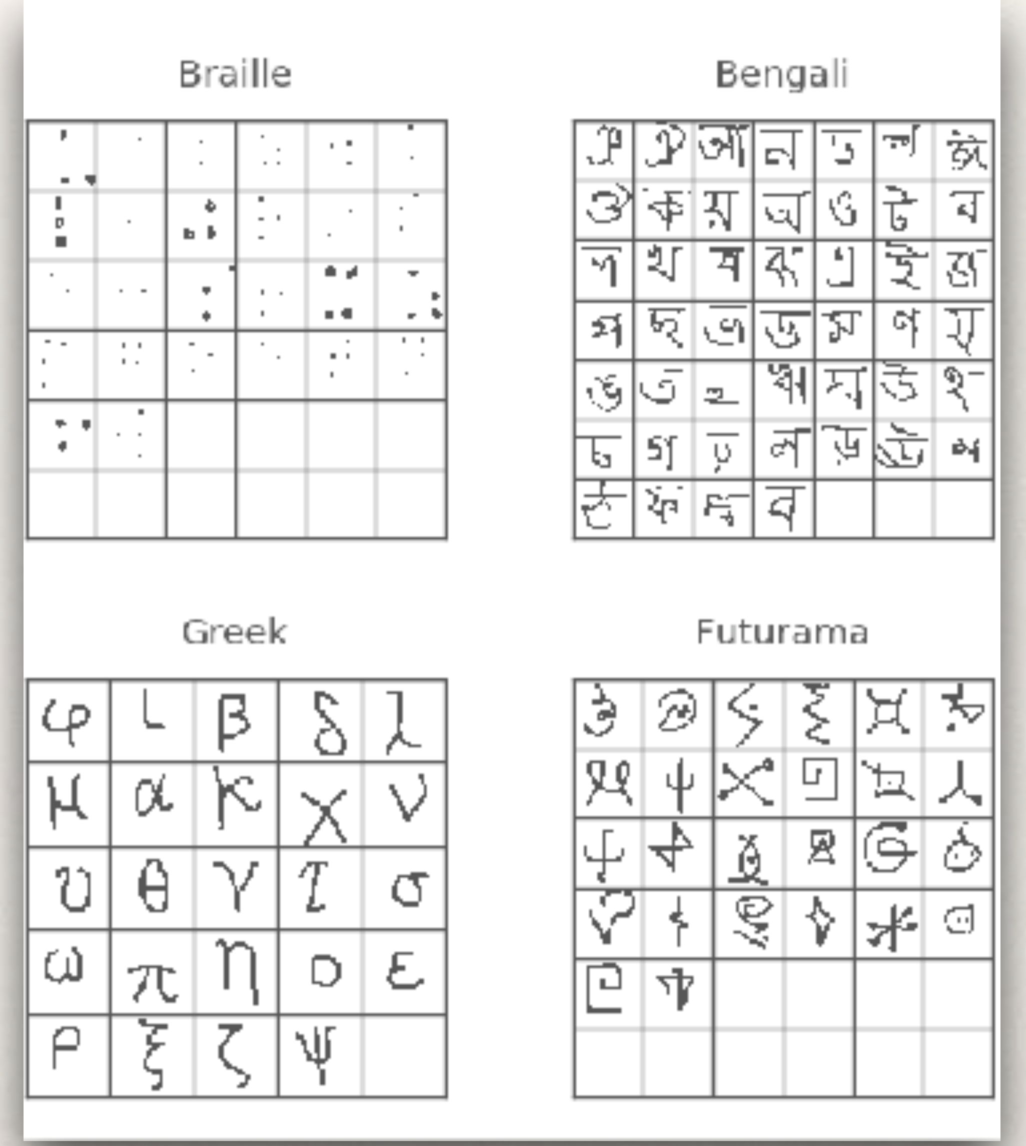
$$S = \{(x_1, y_1), \ldots, (x_N, y_N)\}$$

where N is the number of classes

- N-way few-shot learning

Benchmark: Omniglot

- a collection of 1623 hand drawn characters from 50 alphabets
- every character has 20 examples
- each example is drawn by different person at resolution 105x105



Simple baseline

- 1 Nearest Neighbor

$$C(\hat{x}) = \underset{c \in S}{\operatorname{argmin}} ||\hat{x} - x_c||$$

- calculate the Euclidean distance of the test example from each training example, then pick the closest one.

- ~28% accuracy in 20-way one-shot, Koch et al.

DL for one/few-shot

- deep neural networks will severely overfit on one-shot
- humans spend a lifetime to classify things
- knowledge transfer from other tasks

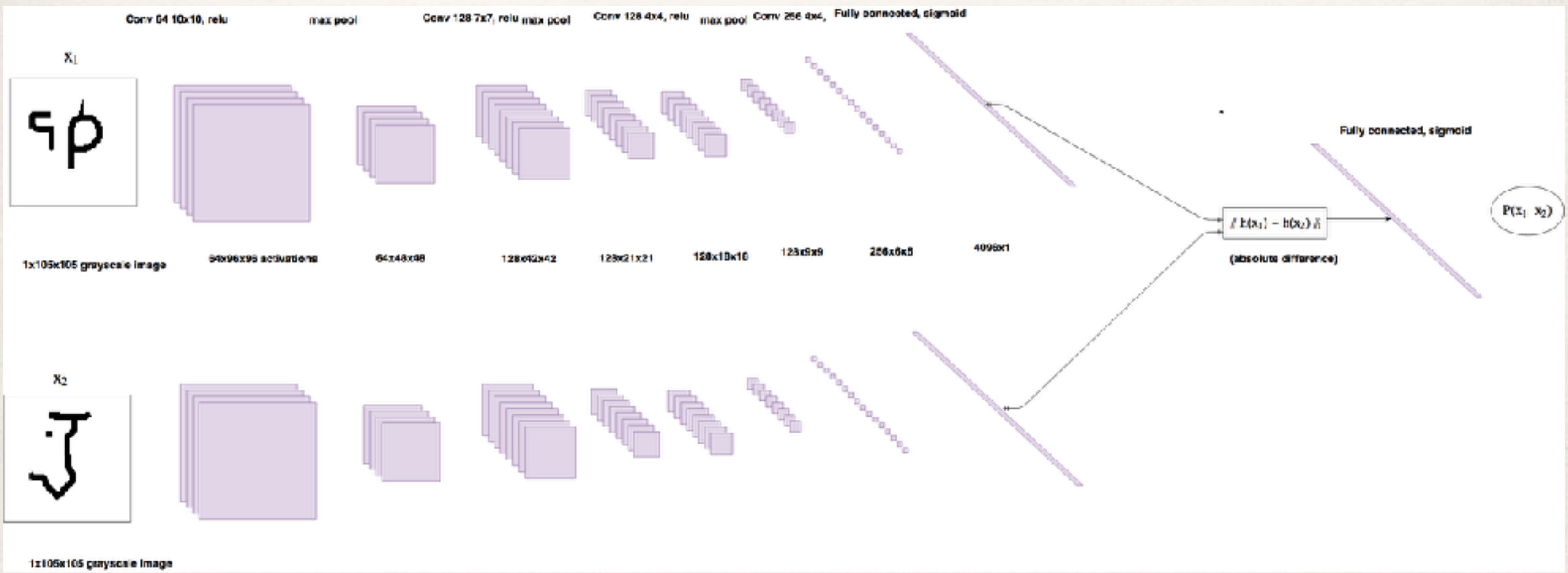
Siamese neural networks for one-shot image recognition

In ICML 2015 Deep Learning Workshop.

–Koch, G., Zemel, R., & Salakhutdinov, R.

Siamese Networks

- Given two images, guess whether they have the same category.



Siamese Networks

Table 2. Comparing best one-shot accuracy from each type of network against baselines.

Method	Test
Humans	95.5
Hierarchical Bayesian Program Learning	95.2
Affine model	81.8
Hierarchical Deep	65.2
Deep Boltzmann Machine	62.0
Simple Stroke	35.2
1-Nearest Neighbor	21.7
Siamese Neural Net	58.3
Convolutional Siamese Net	92.0

Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks

In ICML2017.

–Finn, C., Abbeel, P., & Levine, S.

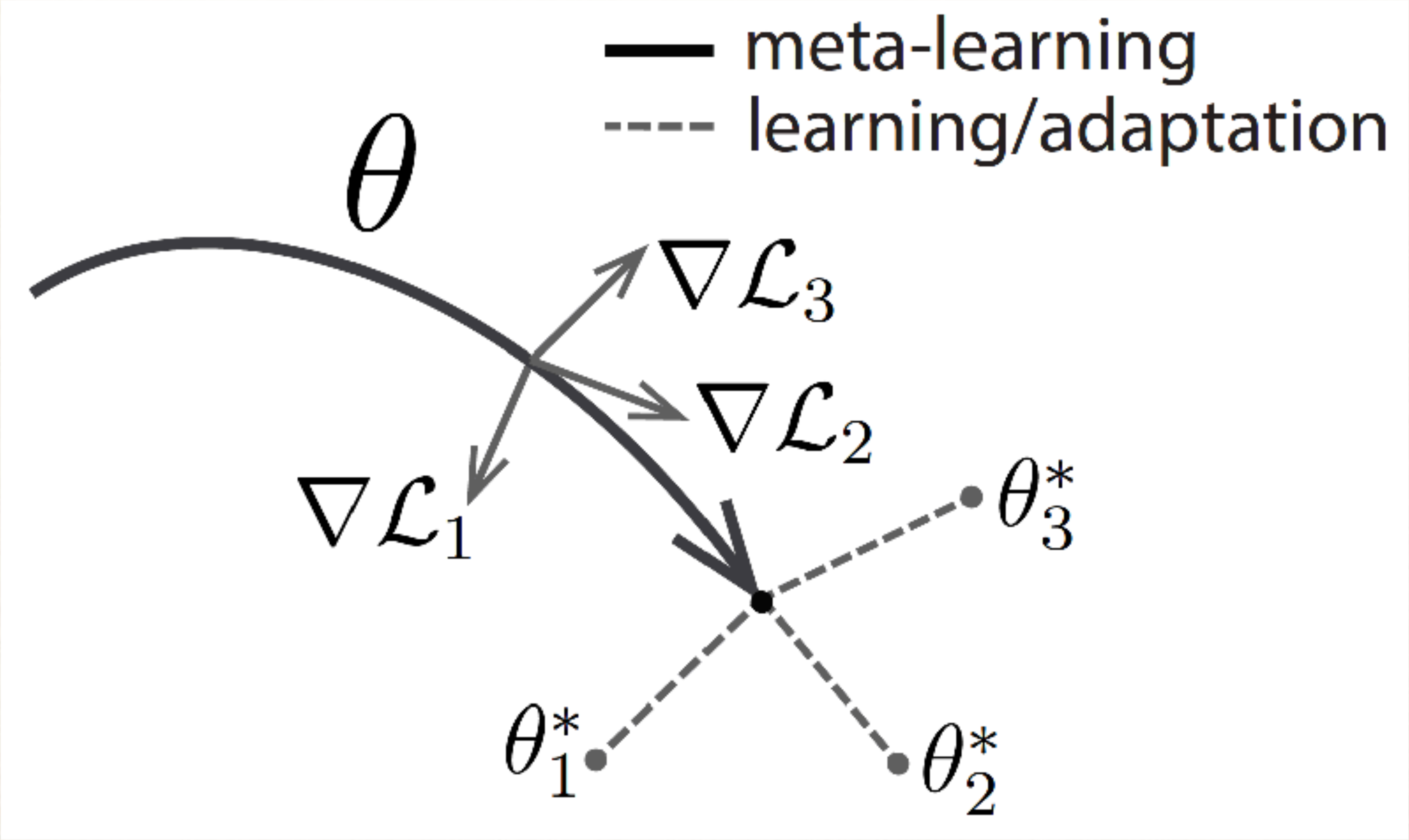
Model-Agnostic Meta-Learning

- learning initialization: optimize for an initial representation that can be effectively fine-tuned from a small number of examples
- useful for adapting to various problems
- be adapted quickly (in a small number of steps)
- and efficiently (using only a few examples)

Another benchmark: Mini-ImageNet

- 60,000 colorful images of size 84×84 with 100 classes, each having 600 examples

Model-Agnostic Meta-Learning



Algorithm 2 MAML for Few-Shot Supervised Learning

Require: $p(\mathcal{T})$: distribution over tasks

Require: α, β: step size hyperparameters

randomly initialize θ

while not done **do**

  Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$

  for all $\mathcal{T}_i$ **do**

    Sample K datapoints $\mathcal{D} = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$

    Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ using $\mathcal{D}$ and $\mathcal{L}_{\mathcal{T}_i}$ in Equation (2) or (3)

    Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$

    Sample datapoints $\mathcal{D}_i' = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$ for the meta-update

  end for

  Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$ using each $\mathcal{D}_i'$ and $\mathcal{L}_{\mathcal{T}_i}$ in Equation 2 or 3

end while

MAML for classification

- experiments on MiniImagenet

MiniImagenet (Ravi & Larochelle, 2017)	5-way Accuracy 1-shot	5-shot
fine-tuning baseline	$28.86 \pm 0.54\%$	$49.79 \pm 0.79\%$
nearest neighbor baseline	$41.08 \pm 0.70\%$	$51.04 \pm 0.65\%$
matching nets (Vinyals et al., 2016)	$43.56 \pm 0.84\%$	$55.31 \pm 0.73\%$
meta-learner LSTM (Ravi & Larochelle, 2017)	$43.44 \pm 0.77\%$	$60.60 \pm 0.71\%$
MAML, first order approx. (ours)	$\mathbf{48.07 \pm 1.75\%}$	$\mathbf{63.15 \pm 0.91\%}$
MAML (ours)	$\mathbf{48.70 \pm 1.84\%}$	$\mathbf{63.11 \pm 0.92\%}$

Reptile: a Scalable Meta-learning Algorithm

In arXiv preprint 2018

–Nichol, A., & Schulman, J.

Reptile

- Similar to MAML
- does not calculate any second derivatives
- takes less computation and memory than MAML

Algorithm 2 Reptile, batched version

Initialize ϕ
for iteration $= 1, 2, \ldots$ **do**
  Sample tasks $\tau_1, \tau_2, \ldots, \tau_n$
  for $i = 1, 2, \ldots, n$ **do**
    Compute $W_i = \mathrm{SGD}(L_{\tau_i}, \phi, k)$
  end for
  Update $\phi \leftarrow \phi + \epsilon \frac{1}{k} \sum_{i=1}^{n} (W_i - \phi)$
end for

Reptile: experiments

- Reptile and MAML yield similar performance
- slightly better in Mini-ImageNet, slightly worse in Omniglot

Algorithm	1-shot 5-way	5-shot 5-way
MAML + Transduction	$48.70 \pm 1.84\%$	$63.11 \pm 0.92\%$
1^{st}-order MAML + Transduction	$48.07 \pm 1.75\%$	$63.15 \pm 0.91\%$
Reptile	$47.07 \pm 0.26\%$	$62.74 \pm 0.37\%$
Reptile + Transduction	$49.97 \pm 0.32\%$	$65.99 \pm 0.58\%$

Results in Mini-ImageNet

References

1. Koch, G., Zemel, R., & Salakhutdinov, R. (2015). Siamese neural networks for one-shot image recognition. In *ICML Deep Learning Workshop* (Vol. 2).

2. Finn, C., Abbeel, P., & Levine, S. (2017, July). Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks. In *International Conference on Machine Learning* (pp. 1126-1135).

3. Nichol, A., & Schulman, J. (2018). Reptile: a Scalable Metalearning Algorithm. *arXiv preprint arXiv:1803.02999*.

4. One Shot Learning and Siamese Networks in Keras. https://sorenbouma.github.io/blog/oneshot/

5. Learning About Algorithms That Learn to Learn. https://towardsdatascience.com/learning-about-algorithms-that-learn-to-learn-9022f2fa3dd5

6. Learning to Learn. http://bair.berkeley.edu/blog/2017/07/18/learning-to-learn/

7. Reptile: A Scalable Meta-Learning Algorithm. https://blog.openai.com/reptile/