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

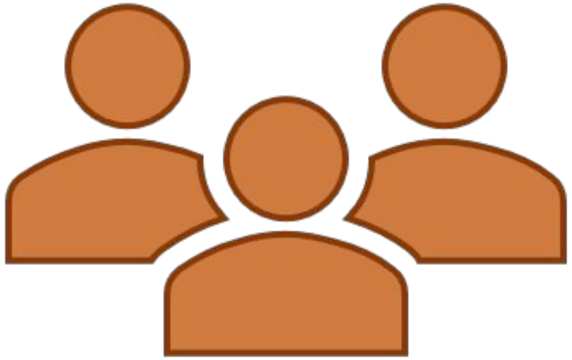
Finn, Chelsea, Pieter Abbeel, and Sergey Levine.

International Conference on Machine Learning. PMLR, 2017.

목차

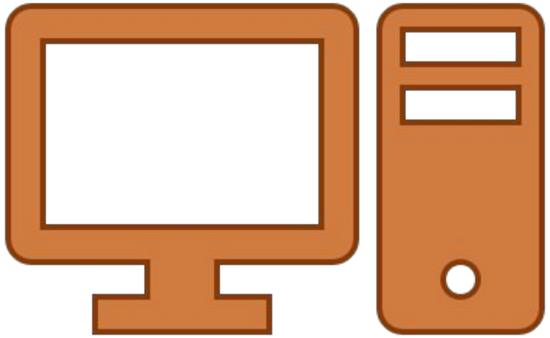
1. Meta-learning overview

1.1 Motivation



몇 분의 경험만으로 학습 가능

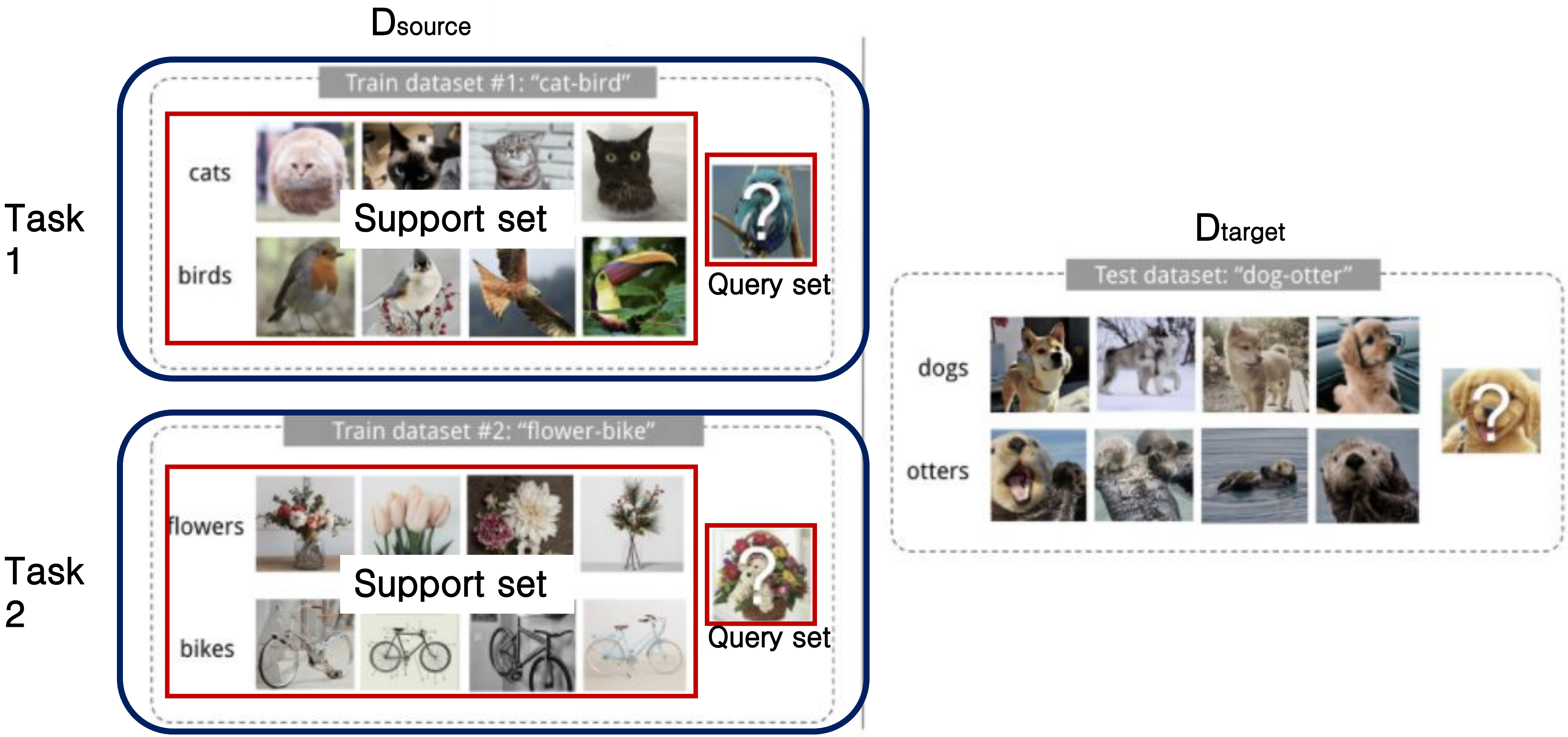
적은 수의 예시만으로 사물 인식



오랜 시간 학습 해야 함

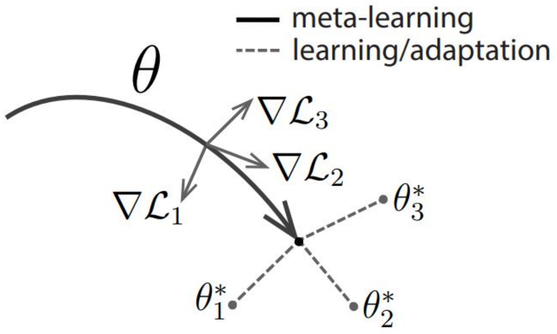
방대한 양의 data를 이용해 학습 해야 사물 인식 가능

1.2 Meta-learning overview : Learning to learn



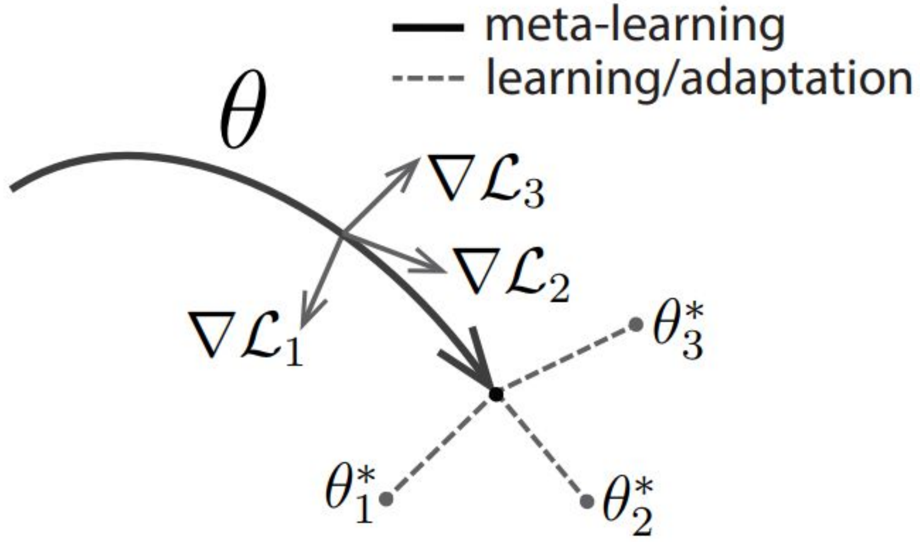
1. Meta-learning Algorithms

1. Model-based methods
2. Optimization-based Inference
   - MAML(Model-Agnostic Meta-Learning)
3. Non-parametric approach



2. Model-Agnostic Meta-Learning

2.1 A Model-Agnostic Meta-Learning Algorithm



$$\theta_i' = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta).$$

$$\min_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'}) = \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)})$$

$$\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta_i'})$$

2.1 A Model-Agnostic Meta-Learning Algorithm

Algorithm 1 Model-Agnostic Meta-Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ with respect to K examples
6: Compute adapted parameters with gradient descent: $\theta'_i = \theta - \alpha \nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$
7: **end for**
8: Update $\theta \leftarrow \theta - \beta \nabla_\theta \sum_{\mathcal{T}_i \sim p(\mathcal{T})} \mathcal{L}_{\mathcal{T}_i}(f_{\theta'_i})$
9: **end while**

1. 랜덤한 θ 값 할당
2. 완료될 때 까지 반복
3. Batch : 한 번 업데이트할 때 사용하는 샘플 묶음
4. 모든 Ti에 대해 반복
5. θ일 때, i번째 Task에 대한 Loss값을 θ 에 대해 미분
   => Loss의 기울기를 구함.
6. θ값 업데이트 한 θ'값 구함.
   (α : 업데이트 크기 조절 변수)
8. θ'을 이용한 모든 Task들의 Loss값을
   θ에 대해 미분하여 최적의 θ 값 얻음

2.2 MAML for Supervised Learning

$$\mathcal{L}_{\mathcal{T}_i}(f_\phi) = \sum_{\mathbf{x}^{(j)},\mathbf{y}^{(j)}\sim\mathcal{T}_i} \mathbf{y}^{(j)} \log f_\phi(\mathbf{x}^{(j)}) + (1-\mathbf{y}^{(j)}) \log(1-f_\phi(\mathbf{x}^{(j)})) \quad (3)$$

Algorithm 2 MAML for Few-Shot Supervised Learning

Require: $p(\mathcal{T})$: distribution over tasks
Require: α, β: step size hyperparameters
1: randomly initialize θ
2: **while** not done **do**
3: Sample batch of tasks $\mathcal{T}_i \sim p(\mathcal{T})$
4: **for all** $\mathcal{T}_i$ **do**
5: Sample K datapoints $\mathcal{D} = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$
6: Evaluate $\nabla_\theta \mathcal{L}_{\mathcal{T}_i}(f_\theta)$ using $\mathcal{D}$ and $\mathcal{L}_{\mathcal{T}_i}$ in Equation (2) or (3)
7: Compute adapted parameters with gradient descent: $\theta_i' = \theta - \alpha\nabla_\theta\mathcal{L}_{\mathcal{T}_i}(f_\theta)$
8: Sample datapoints $\mathcal{D}_i' = \{\mathbf{x}^{(j)}, \mathbf{y}^{(j)}\}$ from $\mathcal{T}_i$ for the meta-update
9: **end for**
10: 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
11: **end while**

$$\mathcal{L}_{\mathcal{T}_i}(f_\phi) = \sum_{\mathbf{x}^{(j)},\mathbf{y}^{(j)}\sim\mathcal{T}_i} \| f_\phi(\mathbf{x}^{(j)}) - \mathbf{y}^{(j)} \|_2^2, \quad (2)$$

5. K개의 D = {x,y}값 선택

6. D값을 이용하는 i번째 Task에 대한 대한 Loss값을 θ에 대해 미분(식 2 또는 3 이용)

식 2) 평균 제곱 오차(MSE) for regression

식 3) for classification

7. θ값 업데이트 한 θ‘값 구함.

(α : 업데이트 크기 조절 변수)

8. 최종 θ를 얻을때 사용할 D값 할당

10. θ’과 D’을 이용한 모든 Task들의 Loss값을 θ에 대해 미분하여 최적의 θ 값 얻음

3. Experimental Evaluation

3. Experimental Evaluation

(1) MAML이 새로운 과제의 빠른 학습을 가능하게 할 수 있는가?

Can MAML enable fast learning of new tasks?

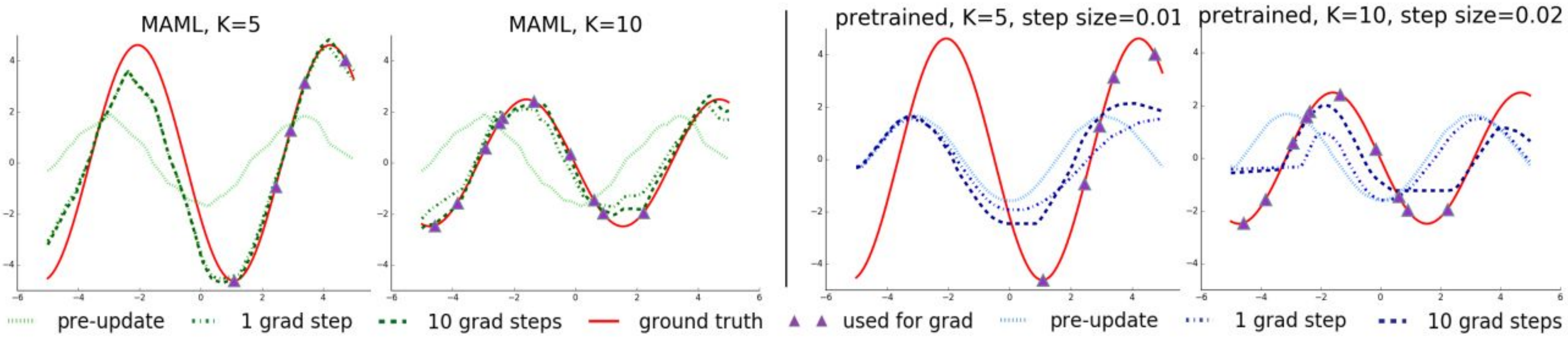
(2) 감독 회귀, 분류 및 강화 학습을 포함한 여러 다른 도메인에서 MAML을 메타 학습에 사용할 수 있는가?

Can MAML be used for meta-learning in multiple different domains, including supervised regression, classification, and reinforcement learning?

(3) MAML로 학습한 모델이 추가 경사도 업데이트 및/또는 예제를 통해 지속적으로 개선될 수 있는가?

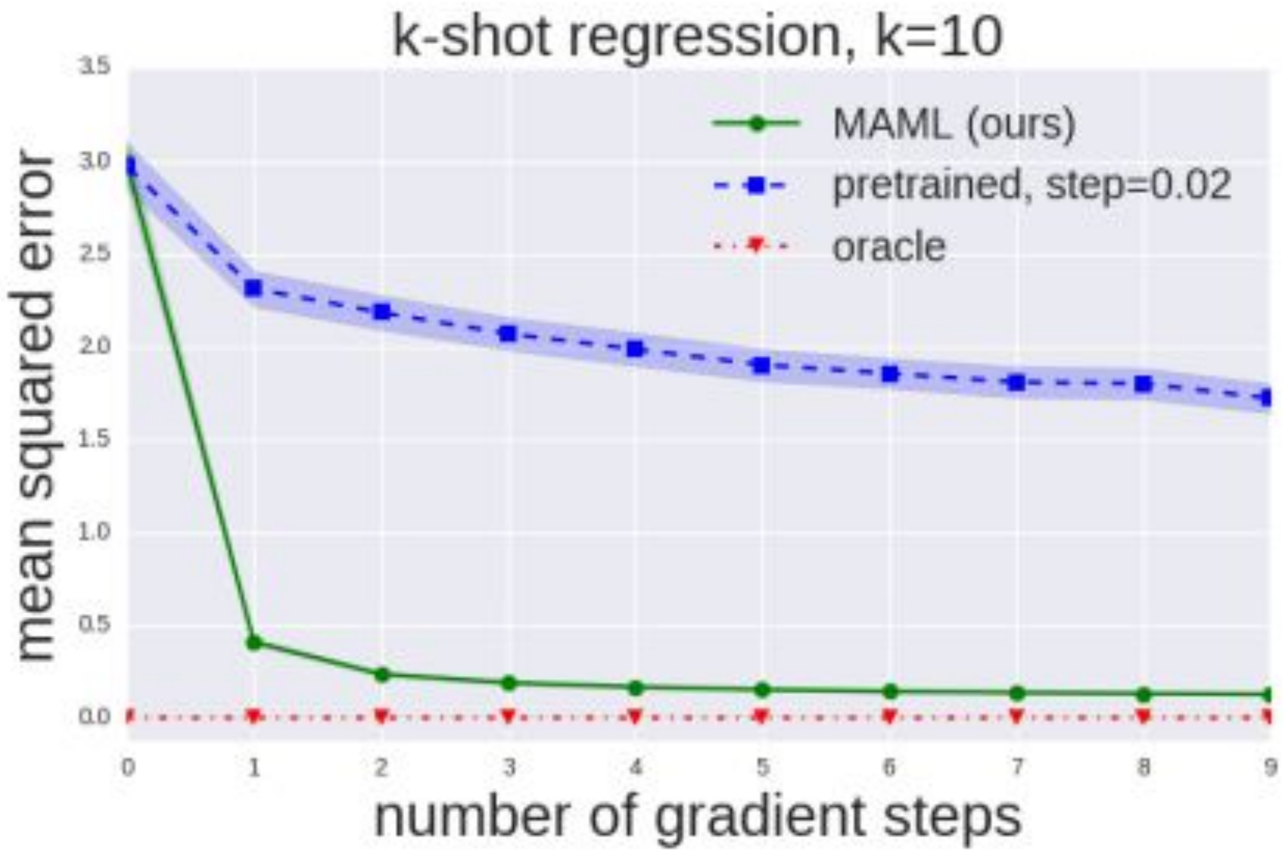
Can model learned with MAML continue to improve with additional gradient updates and/or examples?

3.1 Regression



[Sine wave fitting
]

3.1 Regression



3.2 Classification

Omniglot (Lake et al., 2011)	5-way Accuracy 1-shot	5-way Accuracy 5-shot	20-way Accuracy 1-shot	20-way Accuracy 5-shot
MANN, no conv (Santoro et al., 2016)	82.8%	94.9%	–	–
MAML, no conv (ours)	**89.7 ± 1.1%**	**97.5 ± 0.6%**	–	–
Siamese nets (Koch, 2015)	97.3%	98.4%	88.2%	97.0%
matching nets (Vinyals et al., 2016)	98.1%	98.9%	93.8%	98.5%
neural statistician (Edwards & Storkey, 2017)	98.1%	99.5%	93.2%	98.1%
memory mod. (Kaiser et al., 2017)	98.4%	99.6%	95.0%	98.6%
MAML (ours)	**98.7 ± 0.4%**	**99.9 ± 0.1%**	**95.8 ± 0.3%**	**98.9 ± 0.2%**

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

[N-way classification
]

4. Advanced researches

4. Advanced researches

> Meta-SGD: Learning to learn quickly for few-shot learning, Li et al, Sep 2017

> Recasting gradient-based meta-learning as hierarchical Bayes, Grant et al, ICLR 2018

> Gradient-based meta-learning with learned layerwise metric and and subspace, Lee et al, ICML 2018

> Probabilistic Model-agnostic meta learning, Finn et al, Jun 2018

> Bayesian Model-Agnostic Meta-Learning, Kim et al, Jun 2018

⭐Thank you ⭐