A Closer Look at Few-shot Classification

Wei-Yu Che, Carnegie Mellon University

In ICLR 2019

2020-09-07
MI2RL
Kyuri Kim

Introduction

- Few-Shot Learning and Meta-learning

Few-shot learning algorithms aim to Learn to recognize novel classes with a limited amount of labeled examples. And few-shot classification aims to learn a classifier to recognize unseen classes during training with limited labeled examples.

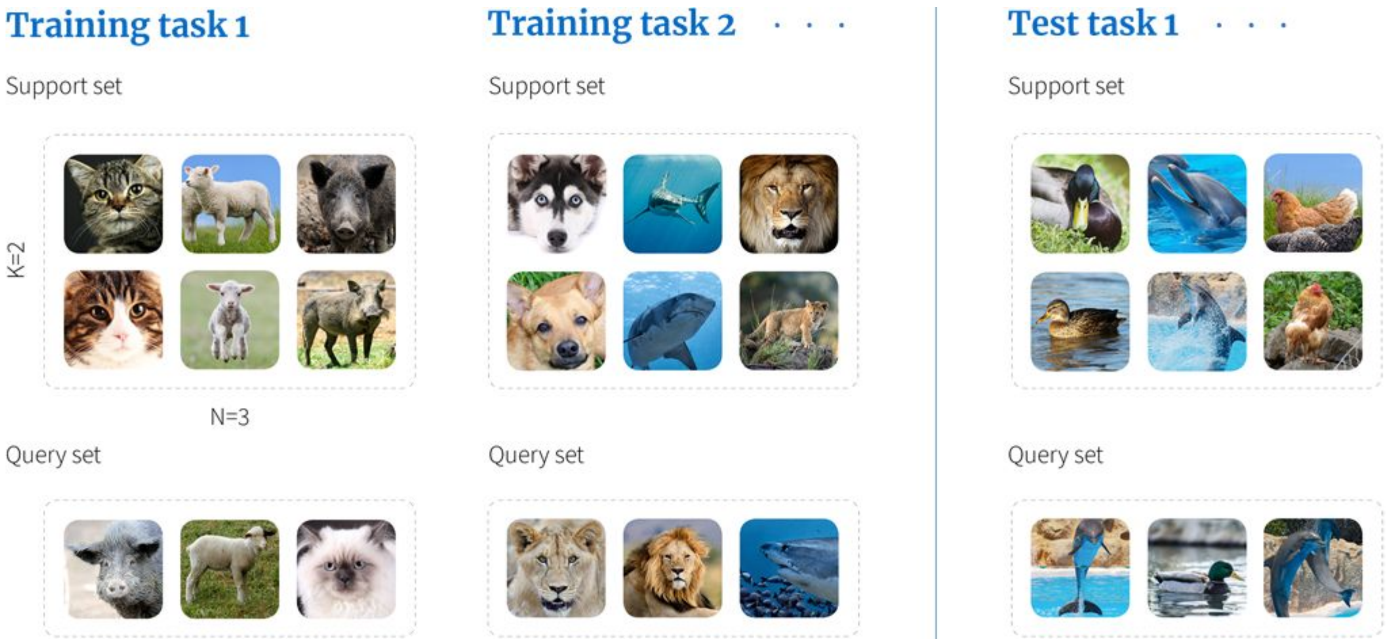
Contribution of this Paper

1. A consistent **comparative analysis** of several representative few-shot classification algorithms.
2. A modified **baseline** method that surprisingly achieves competitive performance.
3. A new experimental setting for **evaluating** the **cross-domain generalization** ability for few-shot classification algorithms.

Related Work

- Few-Shot Learning and Meta-learning

Episodic Training
(N-way K-shot)



https://www.borealisai.com/en/blog/tutorial-2-few-shot-learning-and-meta-learning-i/

Related Work

- Few-Shot Learning and Meta-learning

Meta-Learning

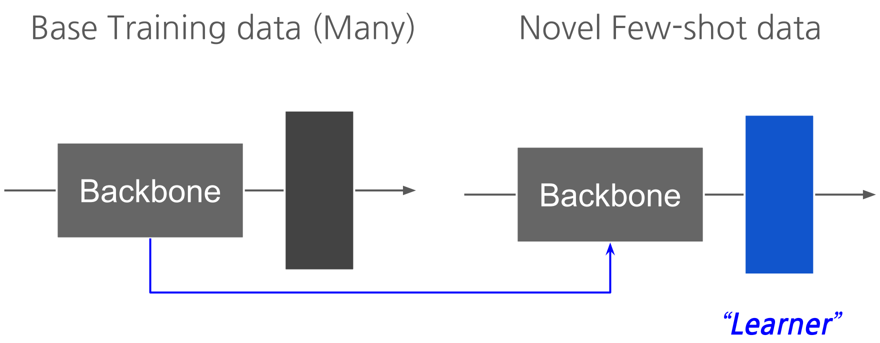
"Learn to Learn" 학습을 위해서 학습한다

Common Approaches

	Key Idea	How $P_\theta(y\|x)$ is modeled?
Model-based	RNN; Memory	$f_\theta(\mathbf{x},S)$
Metric-based	Metric Learning	$\Sigma_{(x_i,y_i)\in S}k_\theta(x,x_i)y_i$ (*)
Optimization-based	Gradient descent	$P_{g_\phi(\theta,S_L)}(y\|\mathbf{x})$

Symposium talks-NIPS 2018
done by Oriol Vinyals

Few-shot Classification (in this Paper)



Related Work

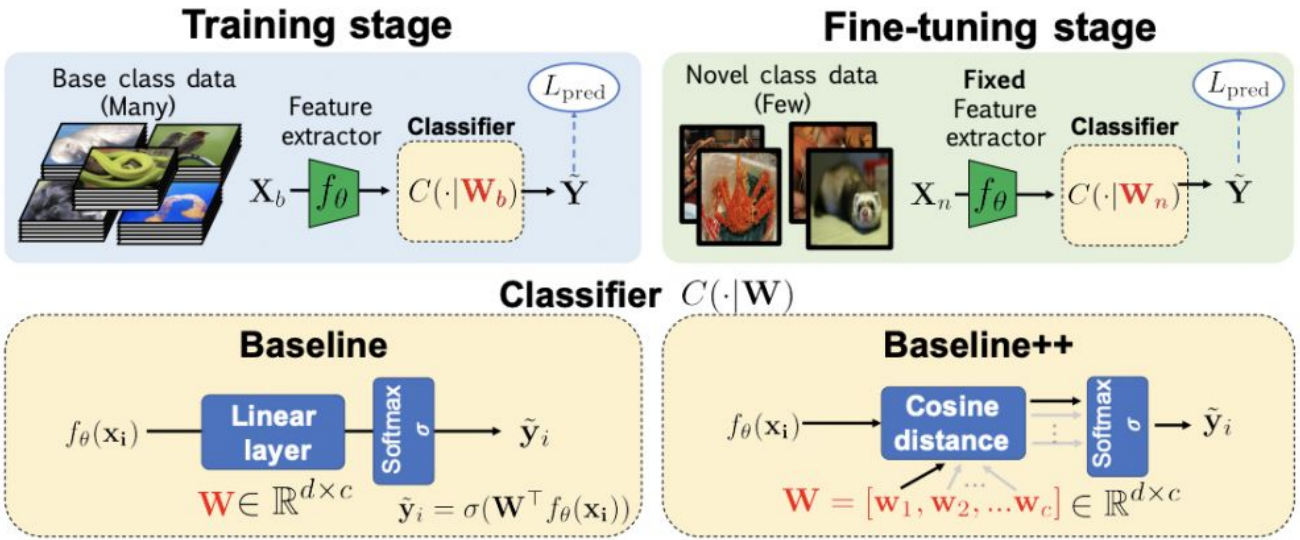
- **Mini ImageNet:**

Method	Backbone	5way-1shot	5way-5shot
MatchingNet	ConvNet	43.56±0.84	55.31± 0.73
ProtoNet	ConvNet	49.42±0.78	68.20±0.66
RelationNet	ConvNet	50.44±0.82	65.32±0.70
MAML	ConvNet	48.70±1.84	55.31±0.73
GNN	ConvNet	50.33±0.36	66.41±0.63
TPN	ConvNet	55.51±0.86	69.86±0.65
Edge-label	ConvNet	59.63±0.52	76.34±0.48
DPGN	**ConvNet**	**66.01±0.36**	**82.83±0.41**
LEO	WRN	61.76±0.08	77.59±0.12
wDAE	WRN	61.07±0.15	76.75±0.11
DPGN	**WRN**	**67.24±0.51**	**83.72±0.44**
CloserLook	ResNet18	51.75±0.80	74.27±0.63
CTM	ResNet18	62.05±0.55	78.63±0.06
DPGN	**ResNet18**	**66.63±0.51**	**84.07±0.42**
MetaGAN	ResNet12	52.71±0.64	68.63±0.67
SNAIL	ResNet12	55.71±0.99	68.88±0.92
TADAM	ResNet12	58.50±0.30	76.70±0.30
Shot-Free	ResNet12	59.04±0.43	77.64±0.39
Meta-Transfer	ResNet12	61.20±1.80	75.53±0.80
FEAT	ResNet12	62.96±0.02	78.49±0.02
MetaOptNet	ResNet12	62.64±0.61	78.63±0.46
DPGN	**ResNet12**	**67.77±0.32**	**84.60±0.43**

https://github.com/megvii-research/DPGN

Method

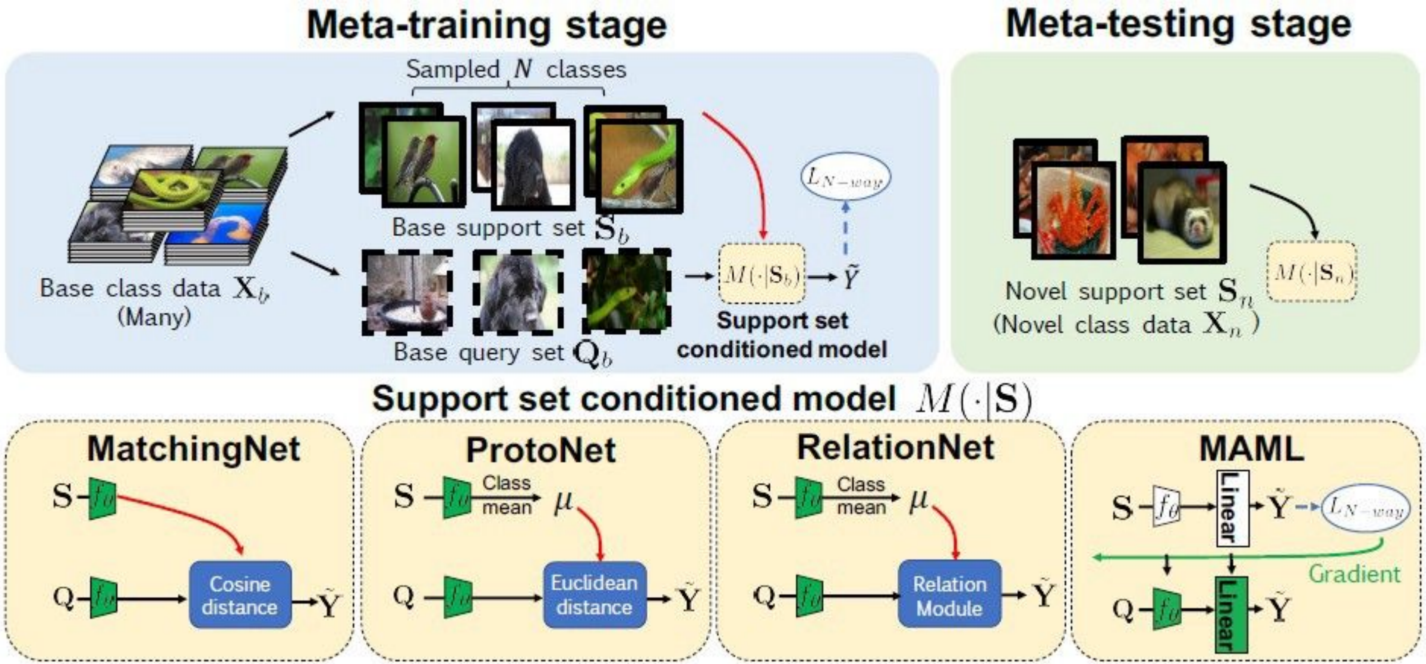
- Baseline & Baseline++



- **Fine-tuning stage:** Fix the network parameters θ in the feature extractor fθ and train a new classifier C(.|Wn) with the given labeled examples in novel classes.
- The **Baseline++ method** differs from the baseline model in the use of cosine distances between the input feature and the weight vector for each class that aims to **reduce intra-class variations.**

Method

- **Comparing with Meta-learning Algorithms**

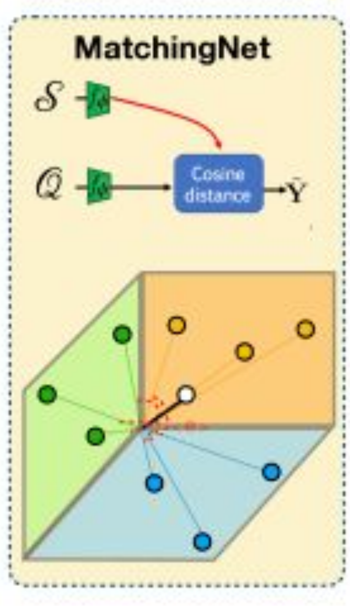


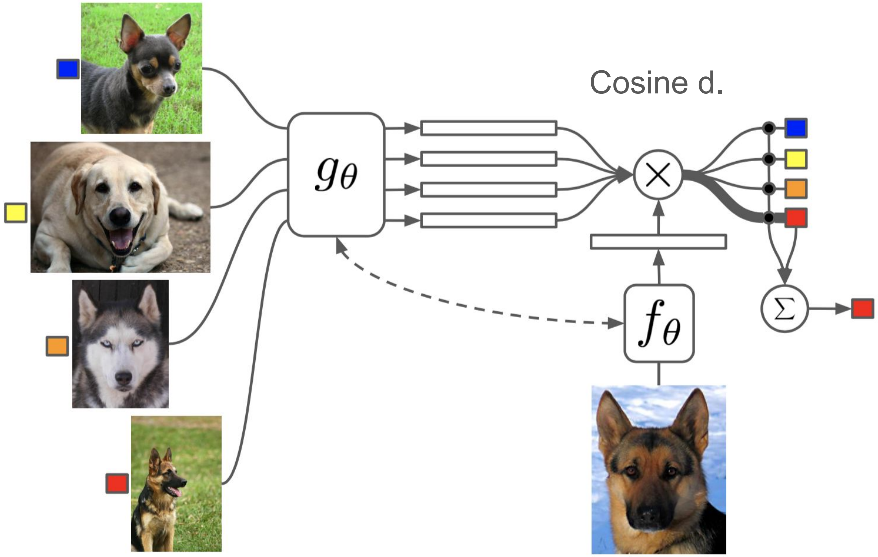
(More details about meta-learning) https://lilianweng.github.io/lil-log/2018/11/30/meta-learning.html

Method

- Comparing with Meta-learning Algorithms

① Metric-based: Matching Network



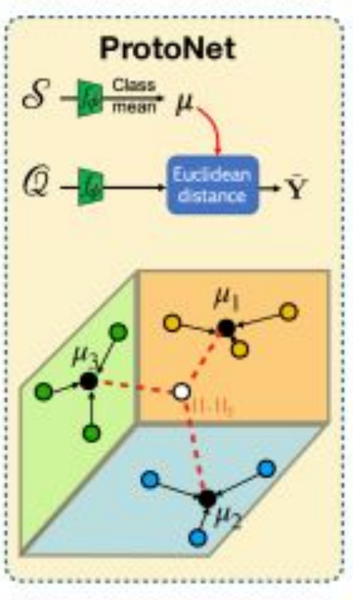


https://arxiv.org/pdf/1606.04080.pdf Matching Networks for One Shot Learning

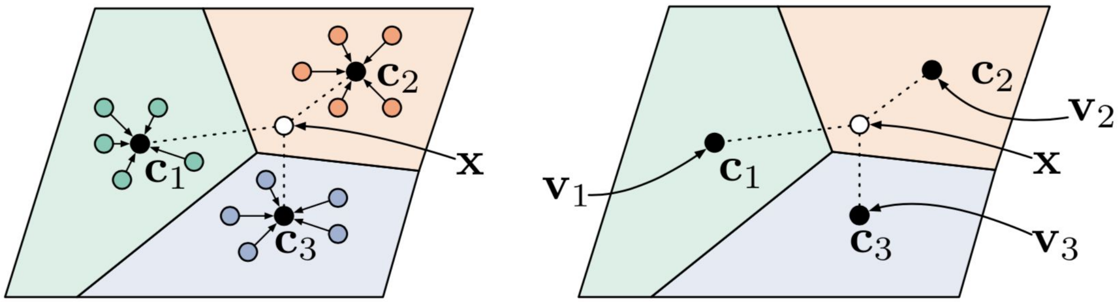
Method

- Comparing with Meta-learning Algorithms

② Metric-based: Prototypical Network



3-way 5-shot



(a) Few-shot

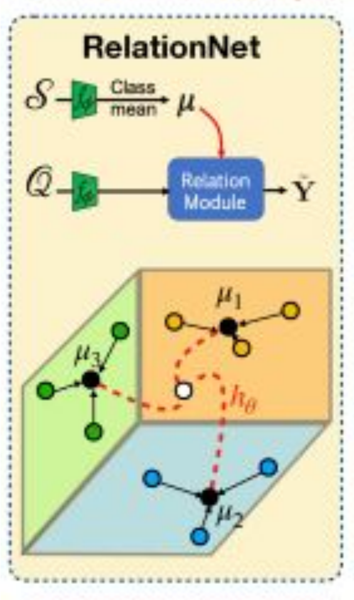
(b) Zero-shot

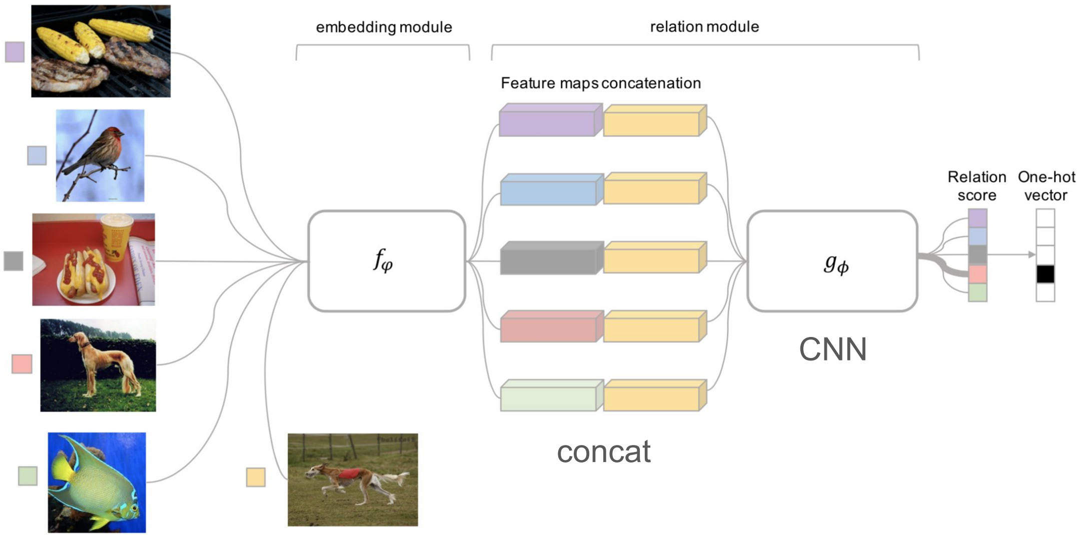
https://arxiv.org/pdf/1703.05175.pdf Prototypical Networks for Few-shot Learning

Method

- Comparing with Meta-learning Algorithms

③ Metric-based: Relation Network



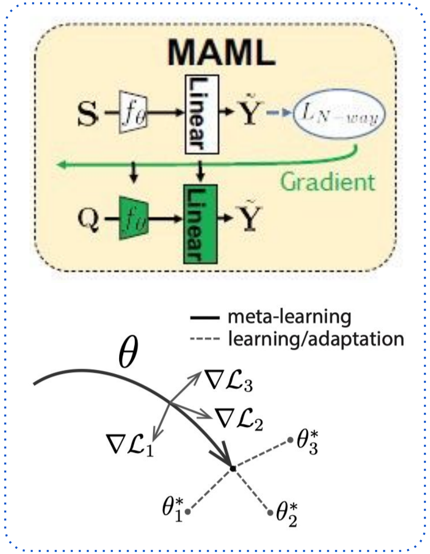


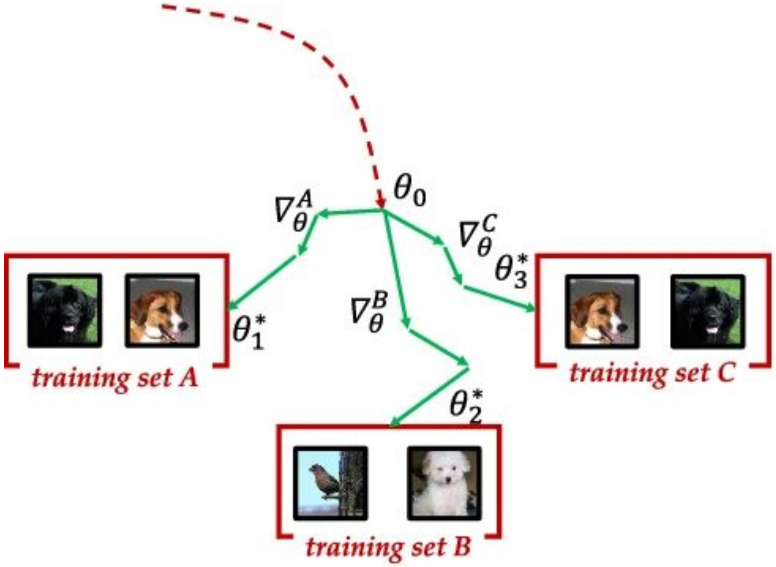
https://arxiv.org/pdf/1711.06025.pdf Learning to Compare: Relation Network for Few-Shot Learning

Method

- **Comparing with Meta-learning Algorithms**

④ Optimization-based: Model-Agnostic Meta-Learning





https://arxiv.org/pdf/1703.03400.pdf Model-Agnostic Meta-Learning for Fast Adaptation of Deep Networks

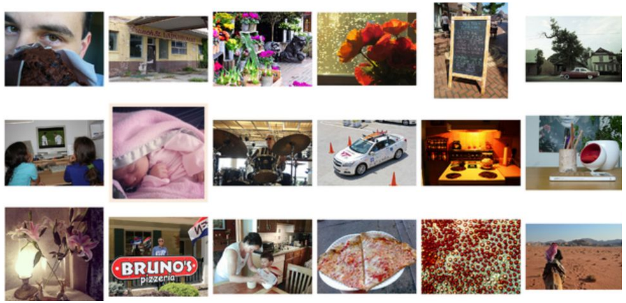
Experiments

We address the few-shot classification problem under three scenarios: **1) generic object recognition**, **2) fine-grained image classification**, and **3) cross-domain adaptation.**

- 5-way k shot classification for support set
- 5-way 16 shot classification for query set

Experiment Dataset

- mini-Image Net
  - Subset of 100 classes from ImageNet
  - Contains 600 images for each class
  - Randomly select 64 base, 16 validation, 20 novel class
- CUB
  - Contains 200 classes and 11,788 images in total
  - Randomly split dataset into 100 base, 50 validation, 50 novel class



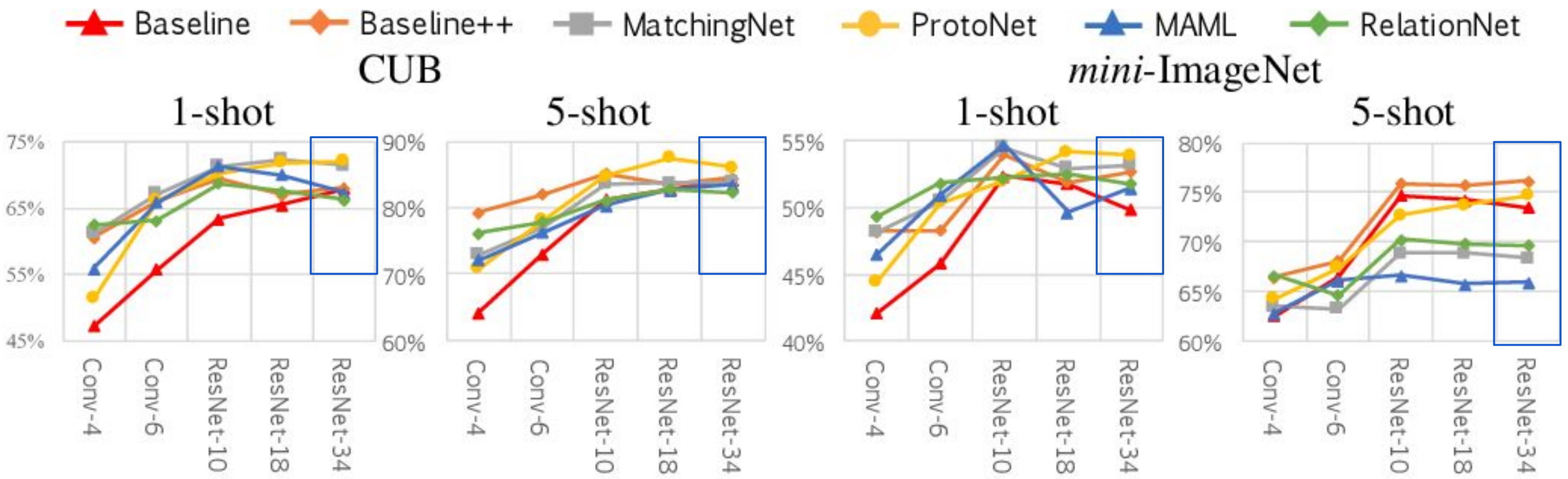
Result

- 1. Few-shot classification results for both the *mini-Imagenet* and *CUB* datasets

Method	CUB 1-shot	CUB 5-shot	*mini*-ImageNet 1-shot	*mini*-ImageNet 5-shot
Baseline	47.12 ± 0.74	64.16 ± 0.71	42.11 ± 0.71	62.53 ± 0.69
Baseline++	60.53 ± 0.83	79.34 ± 0.61	48.24 ± 0.75	66.43 ± 0.63
MatchingNet Vinyals et al. (2016)	61.16 ± 0.89	72.86 ± 0.70	48.14 ± 0.78	63.48 ± 0.66
ProtoNet Snell et al. (2017)	51.31 ± 0.91	70.77 ± 0.69	44.42 ± 0.84	64.24 ± 0.72
MAML Finn et al. (2017)	55.92 ± 0.95	72.09 ± 0.76	46.47 ± 0.82	62.71 ± 0.71
RelationNet Sung et al. (2018)	62.45 ± 0.98	76.11 ± 0.69	49.31 ± 0.85	66.60 ± 0.69

Result

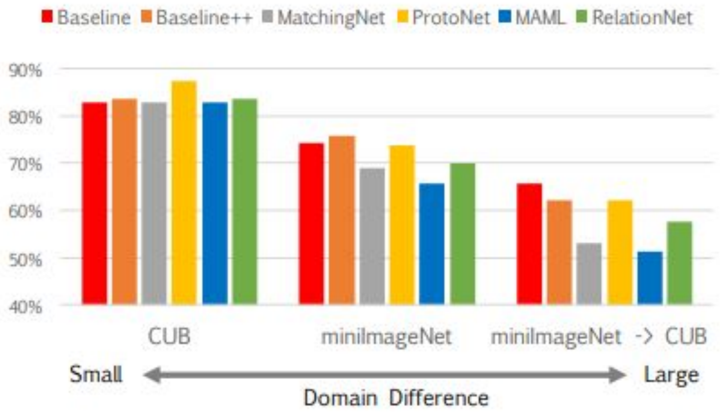
- 2. Few-shot classification accuracy vs. backbone depth



Result

- 3. Effect of Domain Difference

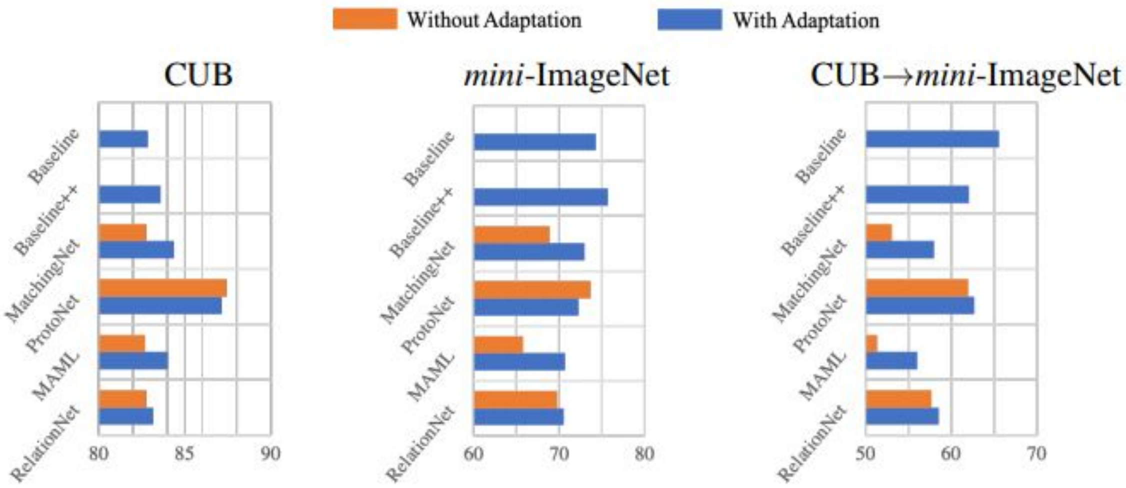
	mini-ImageNet →CUB
Baseline	**65.57±0.70**
Baseline++	62.04±0.76
MatchingNet	53.07±0.74
ProtoNet	62.02±0.70
MAML	51.34±0.72
RelationNet	57.71±0.73



- Baseline outperforms all other methods under this scenario.
- The Baseline model performs relative well with larger domain differences.
- That is, **as the domain difference grows larger, the adaptation based on a few novel class instances becomes more important.**

Result

- 4. Effect of Domain Difference



- learning how to adapt in the meta-training stage is important future direction
- **Learning to learn adaptation in the meta-training stage would be an important direction for future meta-learning research in few-shot classification.**

Conclusions

1. **Our results show that the Baseline++ model is competitive to state of art under standard conditions**
   - Baseline model achieves competitive performance with recent state-of-the-art meta-learning algorithms on both CUB and mini-ImageNet benchmark datasets when using a **deeper feature backbone**

2. **Baseline compares favorably against all the evaluated meta-learning algorithms under a realistic scenario where there exists domain shift between the base and novel classes**
   - **Learning to learn adaptation in the meta-training stage would be an important** direction for future meta-learning research in few-shot classification.