A CLOSER LOOK AT FEW-SHOT CLASSIFICATION

Wei-Yu Chen, Yen-Cheng Liu & Zsolt Kira, Yu-Chiang Frank Wang, Jia-Bin Huang, ICLR, 2019

Wonhee Cho

Vision and Learning Laboratory
School of Computer Science and Engineering, Chung-Ang University, Seoul, Republic of Korea
Emails : wonhee4274@cau.ac.kr
Site : github.com/WonHee94

Index

Background; Few-shot Learning(FSL)

Q. Can a machine learn a new task with few-shot data like a human?

- Deep learning model relies on training a network with **abundant labeled instances**.
- Humans can distinguish a knife only by few-shot images.
- The human **annotation cost** as well as **the scarcity of data in some classes** (e.g., rare species) significantly limit.

- N-way K-shot classification
    - [Support set] vs [Query set]
    - (Intra-Support set) N: #class; K: #support data for each class

- Few-shot classification aim
    - Learn a classifier to recognize unseen classes during training with limited labeled examples.

Background; Meta-Learning

Q) What is Meta-learning?

- Few-shot classification method as meta-learning if the prediction is conditioned on a small support set S.[1]
- It makes the training procedure explicitly learn to learn from a given small support set.

- Episodic Training method
  - Through N-way k-shot classification training, It is a method to derive the learning rules of the model by learning a variety of tasks. [Learning to learn]

[1] Vinyals et al., Matching Networks for one shot learning, NIPS, 2016

Background; Meta-Learning

Meta-learning Problem

- **A Simple View**

$$\theta^* = \underset{\theta}{\operatorname{argmin}} \mathbf{E}_{D_{Task} \sim p(D)}[L_\theta(D_{Task})]$$

  - Minimize loss function $L_\theta(D_{Task})$ from Sampled dataset D_{Task}

- **Meta learning Objective**
  - Training in the same way as testing

$$\theta^* = \underset{\theta}{\operatorname{argmax}} \mathbf{E}_{B \subset D}[\sum_{(x,y) \in B} P_\theta(y|x)] \qquad \text{; Supervised Learning object}$$

$$= \underset{\theta}{\operatorname{argmax}} \mathbf{E}_{L \subset L}[\mathbf{E}_{\mathbf{S} \subset \mathbf{D}, B \subset D}[\sum_{(x,y) \in B} P_\theta(x, y, \boldsymbol{S})]] \qquad \text{; Meta-learning object}$$

  - Maximize Classification output P_θ(y|x) from Sampled dataset point (S^i, B^i)

Background; Meta-Learning

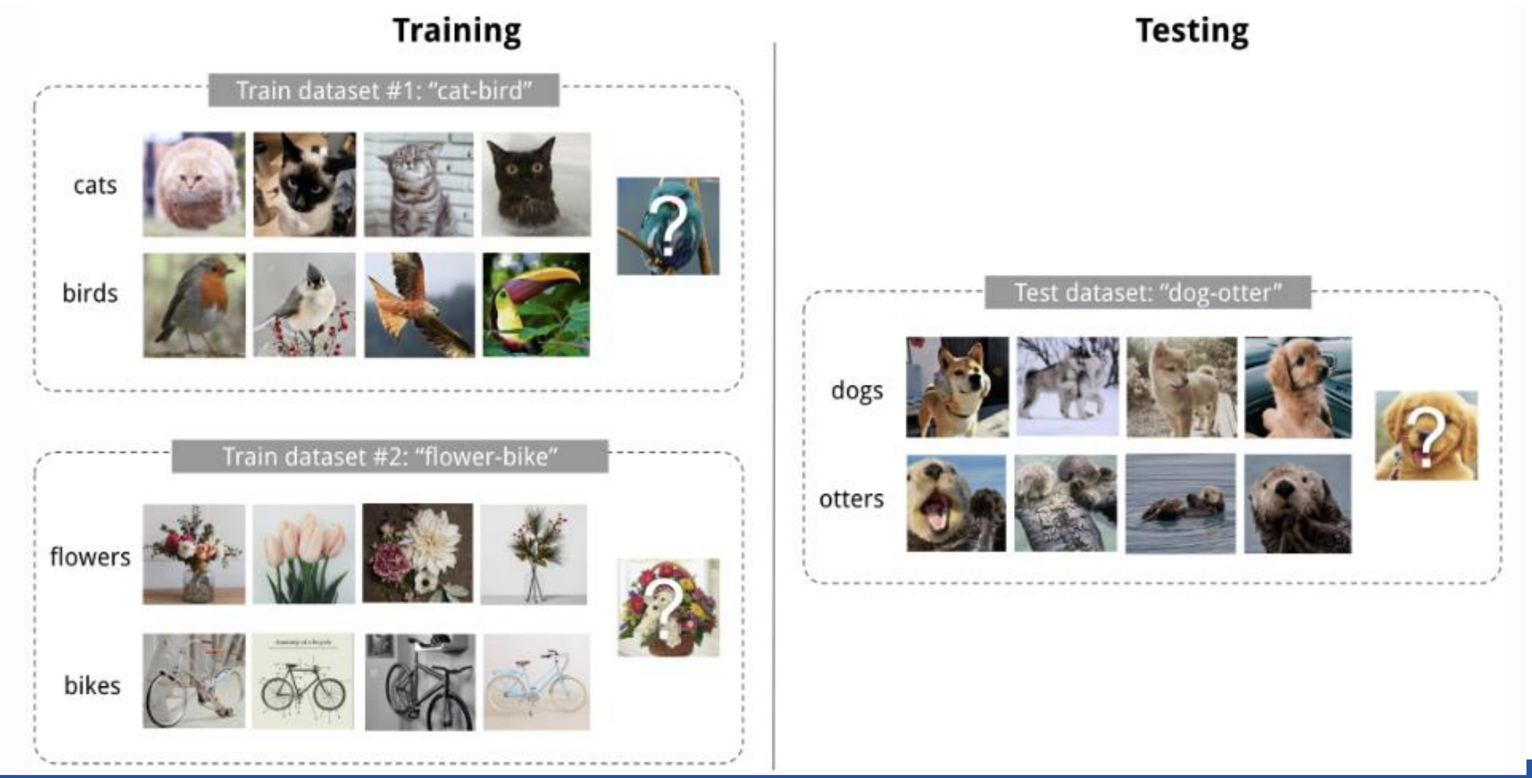
Meta-learning Intuition

- **Aim**

- Reducing prediction errors of unenabled data by using some given labeling sets for fast learning.

- **Intuition**

- In training, we give a little fake to the dataset, making it similar to the test environment. => 'Fast learning'



Related Work

- Initialization based Methods:
  - Learning Good Model initialization
    - Finn, Model Agnostic Meta-Learning
    - Nichol & Schulman, Reptile
  - Learning Optimizer
    - Ravi & Larochelle, Optimization as a model for few-shot learning

Related Work

- Distance metric learning based Methods: " learning to compare"
  - Cosine similarity
    - Vinyals, Matching Networks for one shot learning
  - Euclidean distance
    - Snell, Prototypical Networks for Few-shot Learning
  - Relation module
    - Sung, Learning to Compare: Relation network for few-shot learning
  - Graph neural network
    - Garcia & Bruna, Few-shot learning with Graph Neural Networks
- Hallucination based methods: "learning to augment"
- Domain adaptation: Reduce the domain shifht between source and target domain

Related Work

- Hallucination based methods: “learning to augment”
  - Hariharan & Girshick, Low-shot visual recognition by shrinking and hallucinating features
  - Antoniou, Data augmentation generative adversarial networks
  - Wang, Low-shot learning from imaginary data

Limitations

- Two main challenges of few-shot classification

1. The discrepancy of the implementation details among multiple FSL algorithms obscures the relative performance gain.

2. While the current evaluation focuses on recognizing novel class with limited training examples

$\Rightarrow$ Novel classes are sampled from the same dataset.

$\Rightarrow$ The lack of domain shift between the base and novel classes makes the evaluation scenarios unrealistic.

Contribution

1. A consistent comparative analysis of several representative few-shot classification algorithms.

⇒ Deeper backbones significantly reduce the performance differences among methods on datasets with limited domain differences.

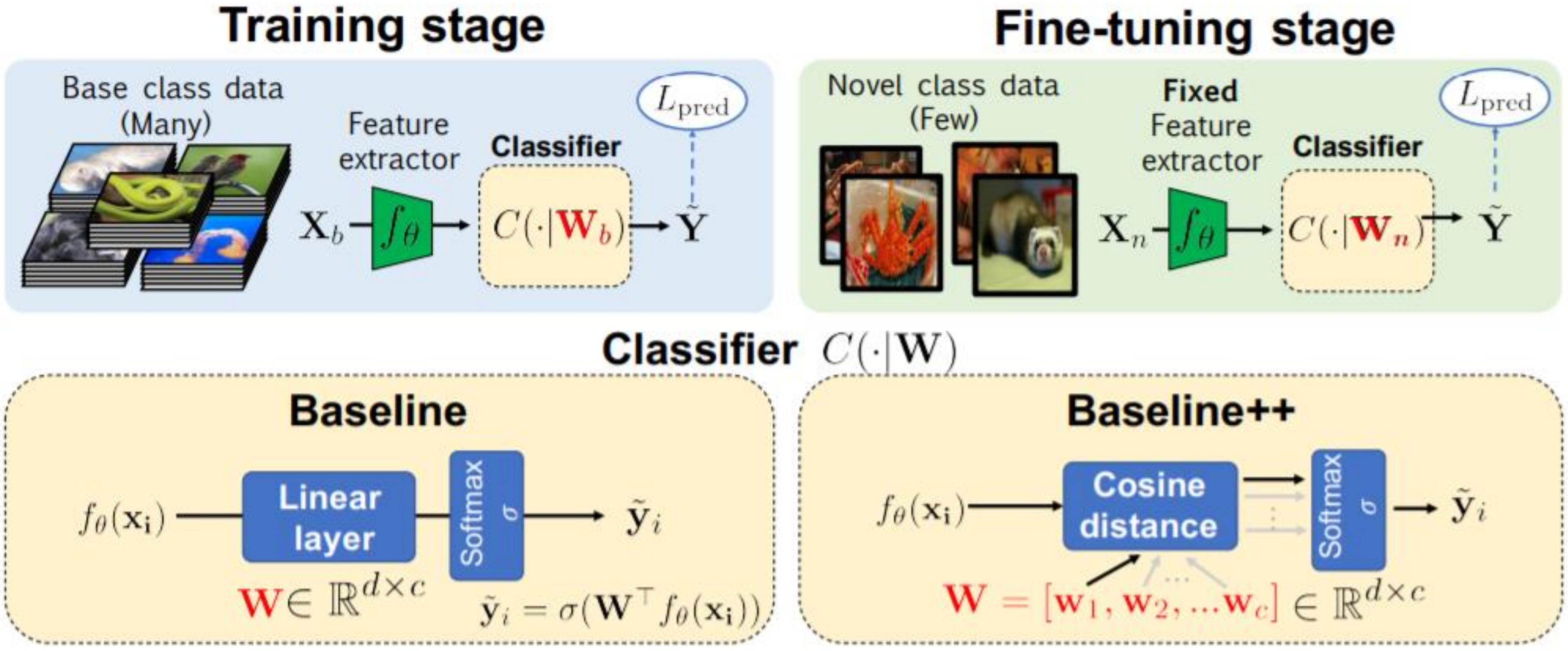
2. Modified baseline method that achieves competitive performance when compared with the SOTA.

3. A new experimental setting for evaluating the cross-domain generalization ability for few-shot classification.

⇒ Reducing intra-class variation is an important when the feature backbone is shallow.

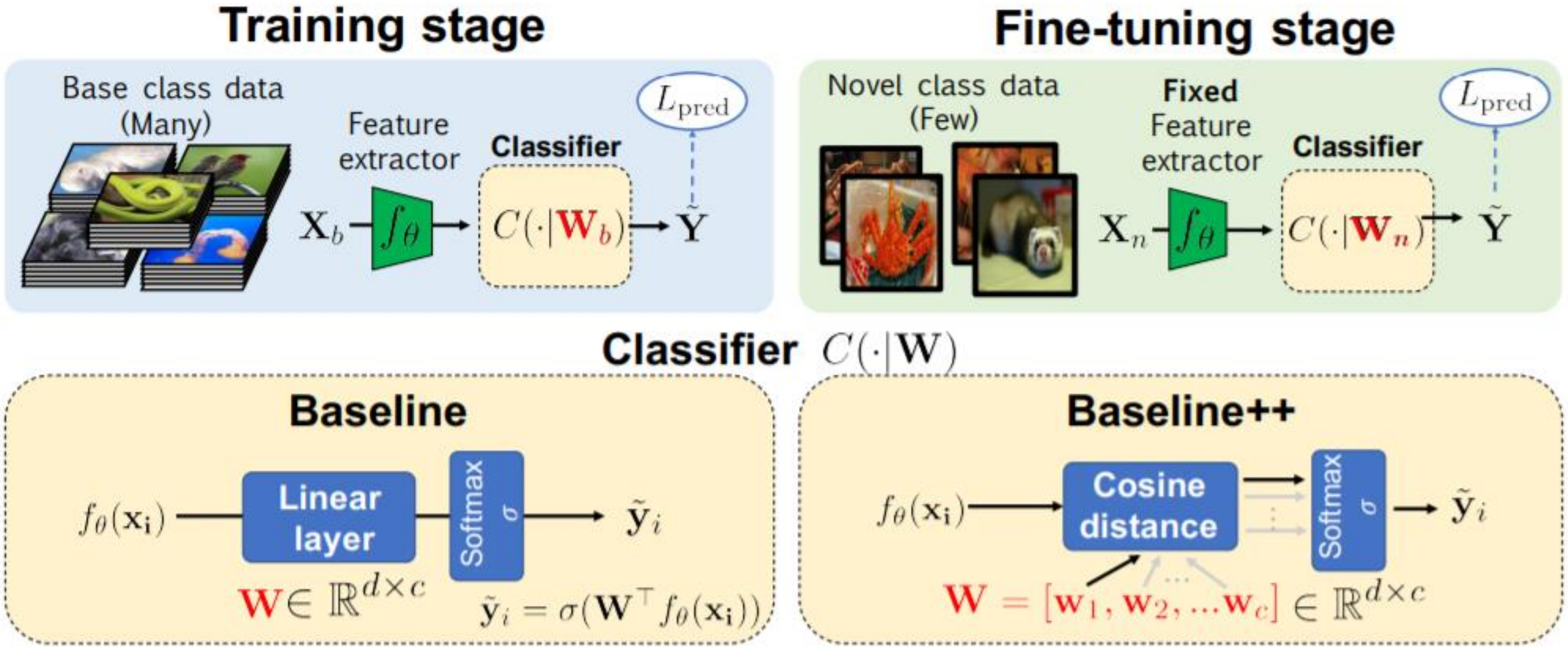
⇒ But not critical when using deeper backbones.

Few-shot classification_Baseline



Training Stage: Train f_θ, W_b
Fine-tuning Stage: Forward f_θ, Fine-tune W_n
Loss: Cross-Entropy Loss

Few-shot classification_Baseline++

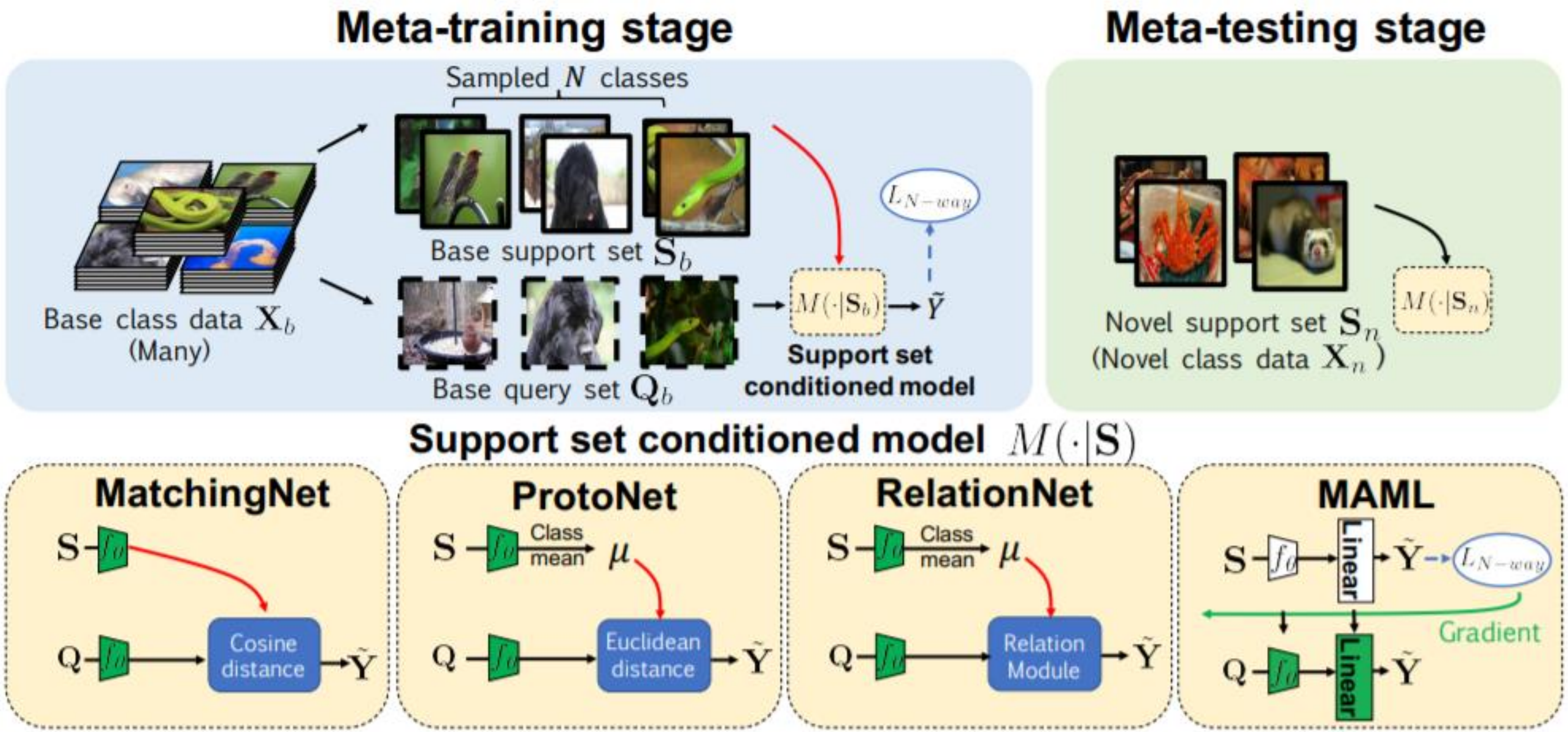


W = $[w_1, w_2, \ldots, w_c]$ for each class

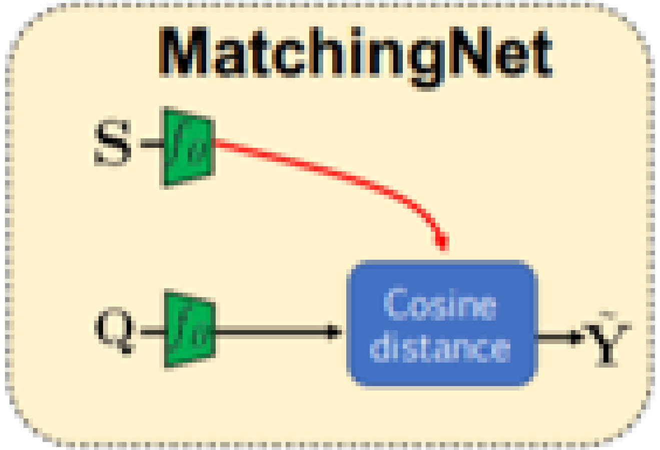
Cosine Similarity between $f_\theta(x_i)$ and w_j -> Softmax

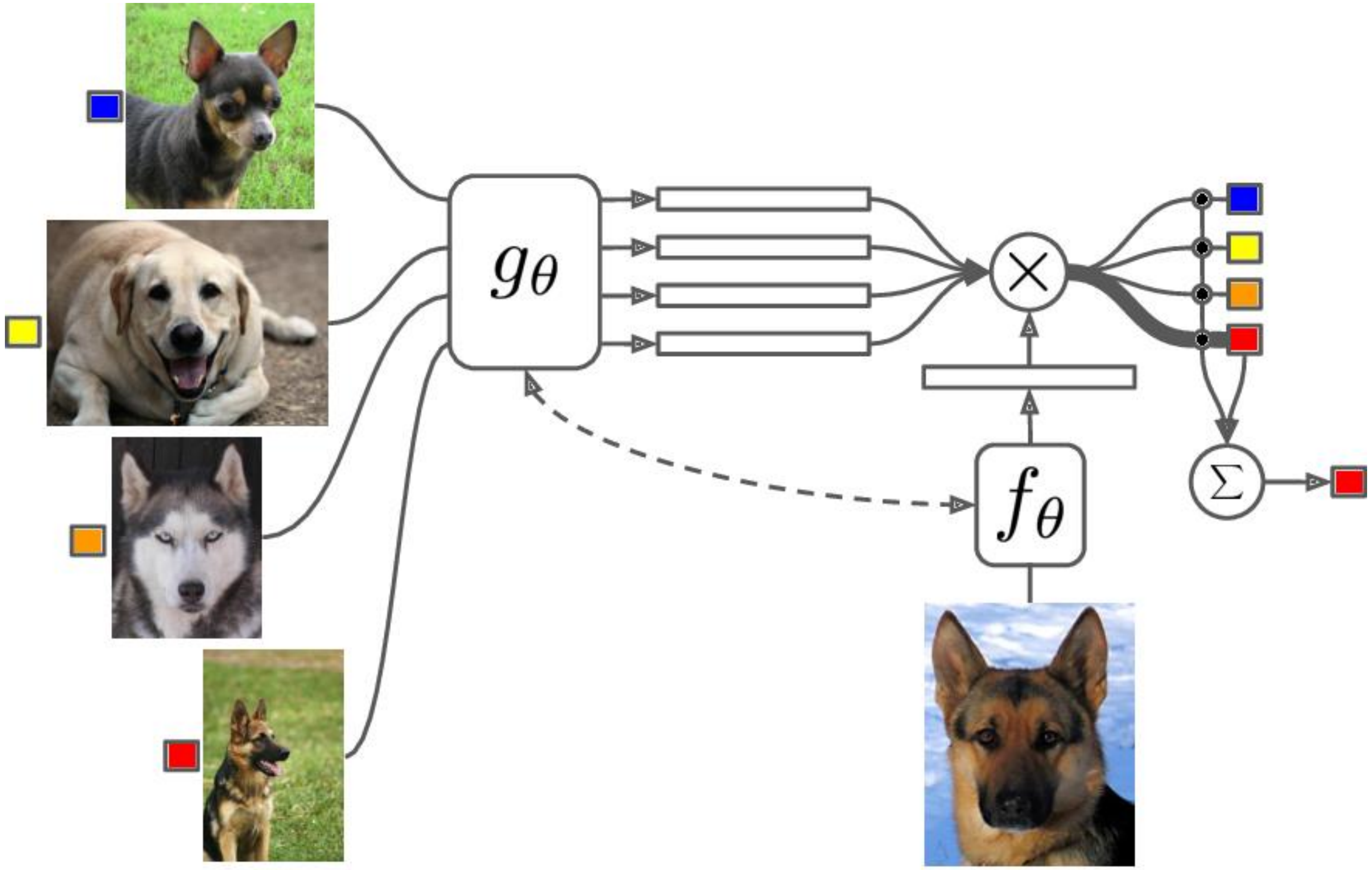
Reduce Intra-class Variations

Meta-learning based Few-shot classification

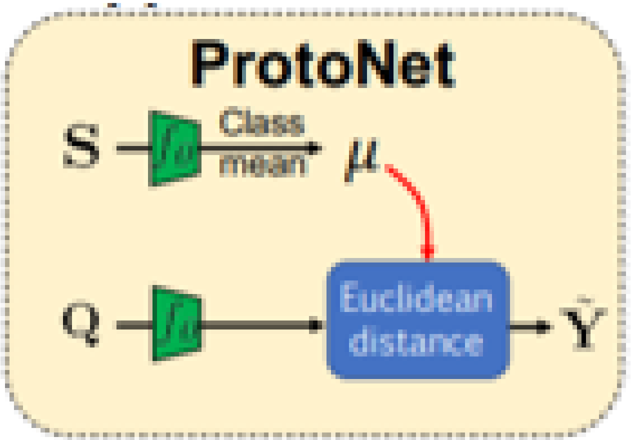


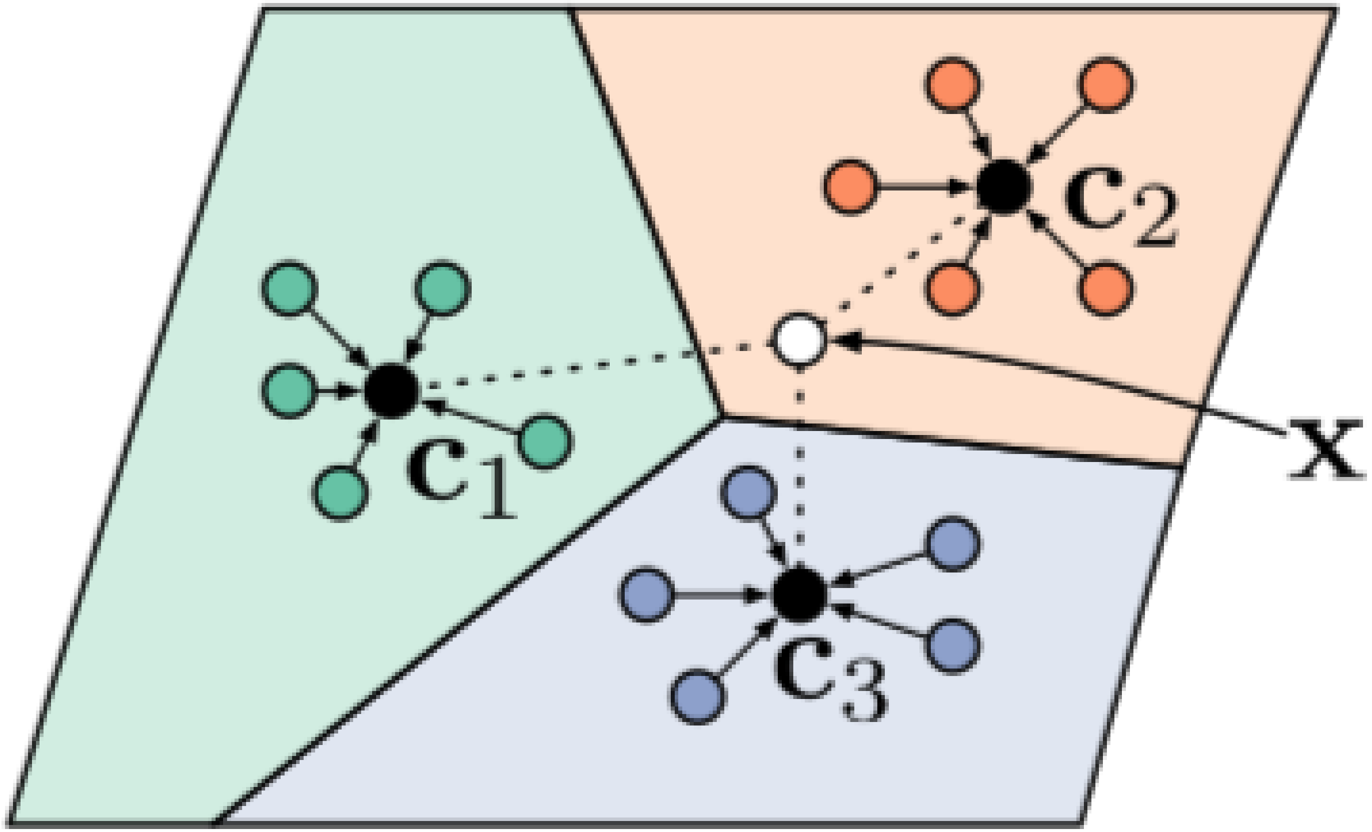
MatchingNet





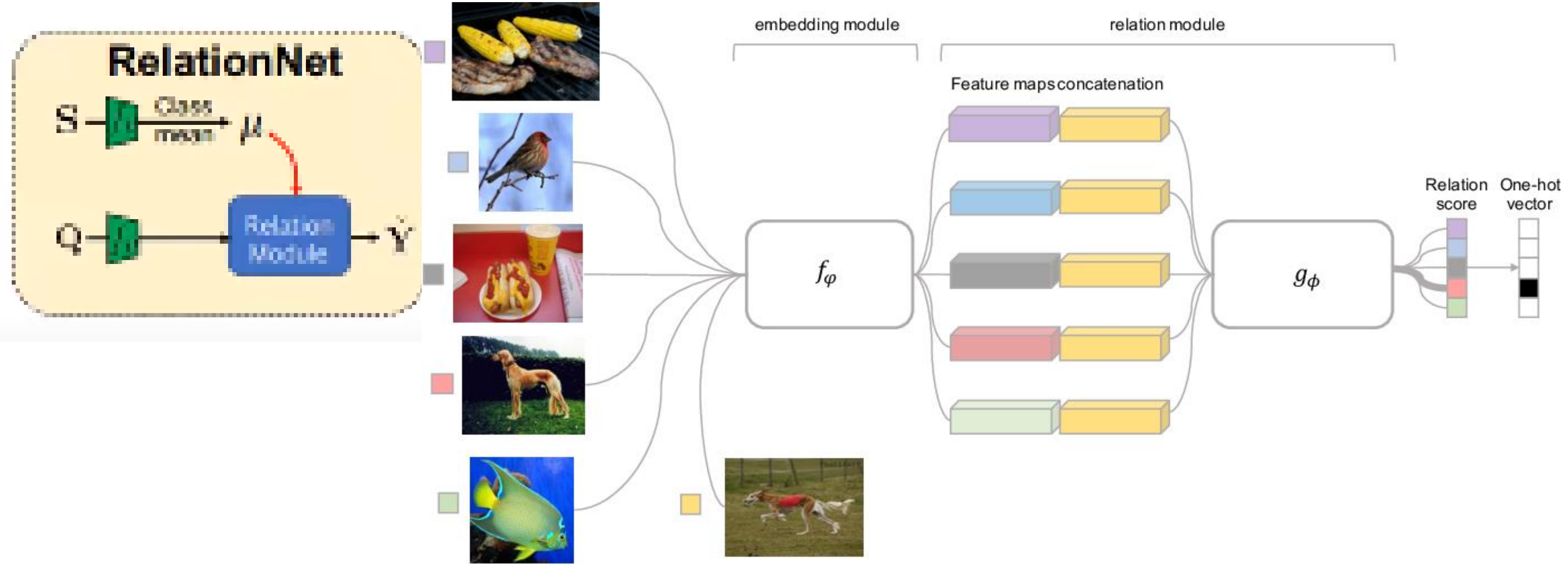
ProtoNet





(a) Few-shot

RelationNet



Experiments

- Mini-ImageNet: 600 images per class (100), 64 base class, 16 validation class, 20 novel class
- CUB-200-2011: 200 classes, 11,788 images, 100 base class, 50 validation class, 50 novel class

- Backbone: Conv-4

- Adam Optimizer

- Data Augmentation

Experimental Results

600 test episode with 95% confidence
* w/o data augmentation

Method	1-shot Reported	1-shot Ours	5-shot Reported	5-shot Ours
Baseline	-	42.11 ± 0.71	-	62.53 ± 0.69
Baseline*[3]	41.08 ± 0.70	36.35 ± 0.64	51.04 ± 0.65	54.50 ± 0.66
MatchingNet[3] Vinyals et al. (2016)	43.56 ± 0.84	48.14 ± 0.78	55.31 ± 0.73	63.48 ± 0.66
ProtoNet	-	44.42 ± 0.84	-	64.24 ± 0.72
ProtoNet# Snell et al. (2017)	49.42 ± 0.78	47.74 ± 0.84	68.20 ± 0.66	66.68 ± 0.68
MAML Finn et al. (2017)	48.07 ± 1.75	46.47 ± 0.82	63.15 ± 0.91	62.71 ± 0.71
RelationNet Sung et al. (2018)	50.44 ± 0.82	49.31 ± 0.85	65.32 ± 0.70	66.60 ± 0.69

Experimental Results

Table 2: **Few-shot classification results for both the *mini*-ImageNet and *CUB* datasets.** The Baseline++ consistently improves the Baseline model by a large margin and is competitive with the state-of-the-art meta-learning methods. All experiments are from 5-way classification with a Conv-4 backbone and data augmentation.

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)	60.52 ± 0.88	75.29 ± 0.75	48.14 ± 0.78	63.48 ± 0.66
ProtoNet Snell et al. (2017)	50.46 ± 0.88	76.39 ± 0.64	44.42 ± 0.84	64.24 ± 0.72
MAML Finn et al. (2017)	54.73 ± 0.97	75.75 ± 0.76	46.47 ± 0.82	62.71 ± 0.71
RelationNet Sung et al. (2018)	62.34 ± 0.94	77.84 ± 0.68	49.31 ± 0.85	66.60 ± 0.69

Experimental Results

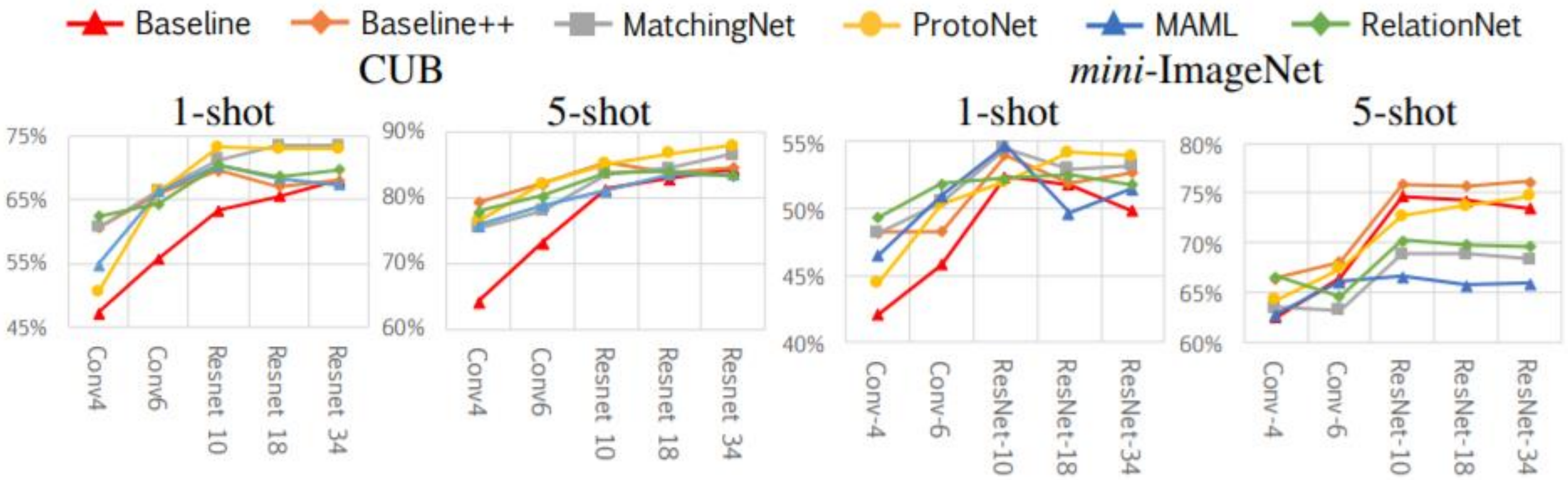


Figure 3: **Few-shot classification accuracy vs. backbone depth**. In the CUB dataset, gaps among different methods diminish as the backbone gets deeper. In *mini*-ImageNet 5-shot, some meta-learning methods are even beaten by Baseline with a deeper backbone. (Please refer to Figure A3 and Table A5 for larger figure and detailed statistics.)

Experimental Results

	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

Table 3: **5-shot accuracy under the cross-domain scenario with a ResNet-18 backbone.** Baseline outperforms all other methods under this scenario.

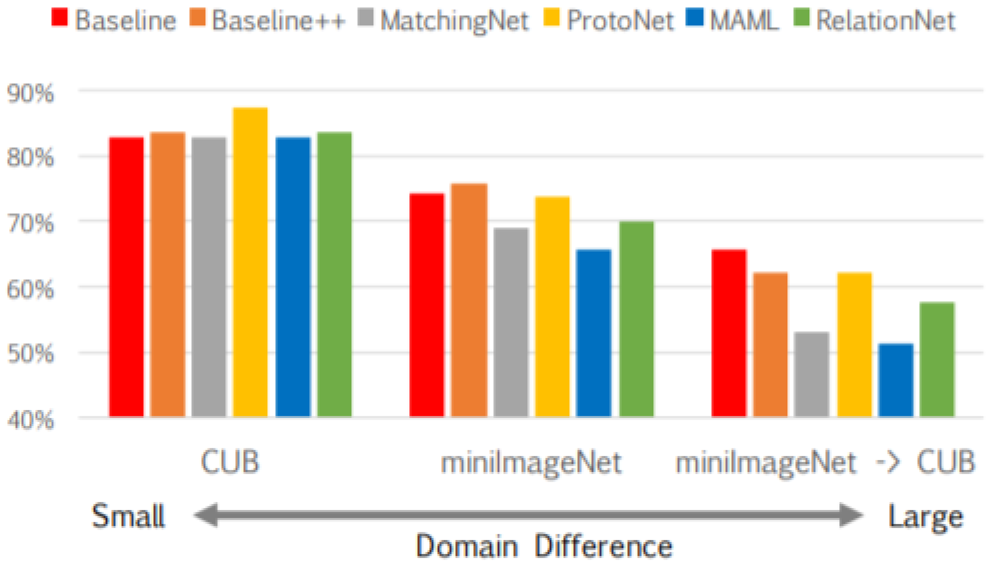


Figure 4: **5-shot accuracy in different scenarios with a ResNet-18 backbone.** The Baseline model performs relative well with larger domain differences.

Q & A

Thank You