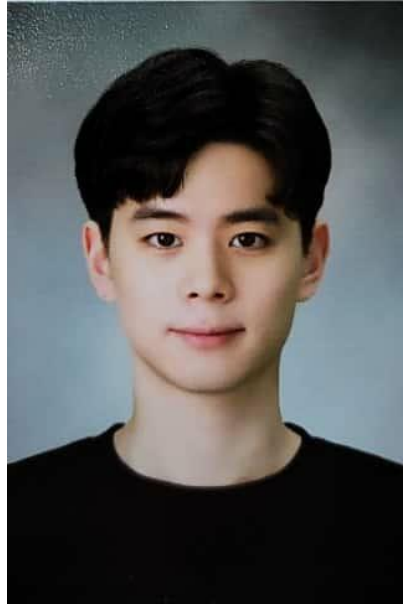


Metric-based approaches to meta-learning

목차

2020.11.06





- 목충협(Chunghyup Mok)
 - ✓ 고려대학교 산업경영공학
 - ✓ Data Mining & Quality Analytics Lab.
 - ✓ 석박통합과정(2019.03 ~)
- 관심분야
 - ✓ Machine learning / Deep learning
 - ✓ Meta-learning / Multitask learning
- E-mail : mokch@korea.ac.kr

Contents

1. Introduction

- Background

2. Few-shot learning

- Definition
- Training strategy

3. Metric-based approaches to meta-learning

- Metric learning
- Matching Networks
- Prototypical networks
- Relation network

4. Evaluations

- A closer look at few-shot classification

5. Conclusions

- Comments

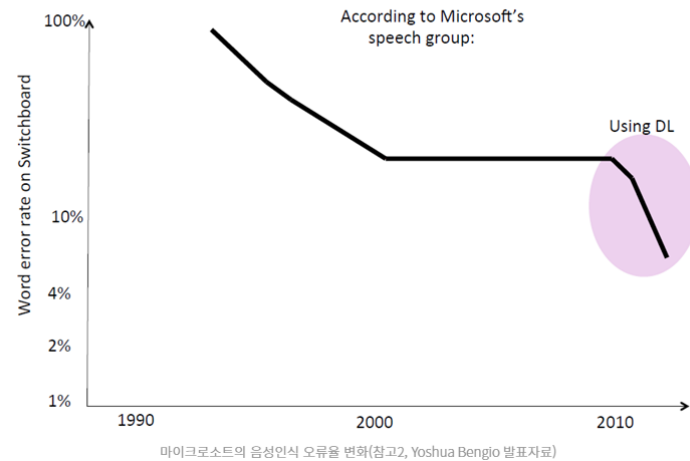
Introduction

Background

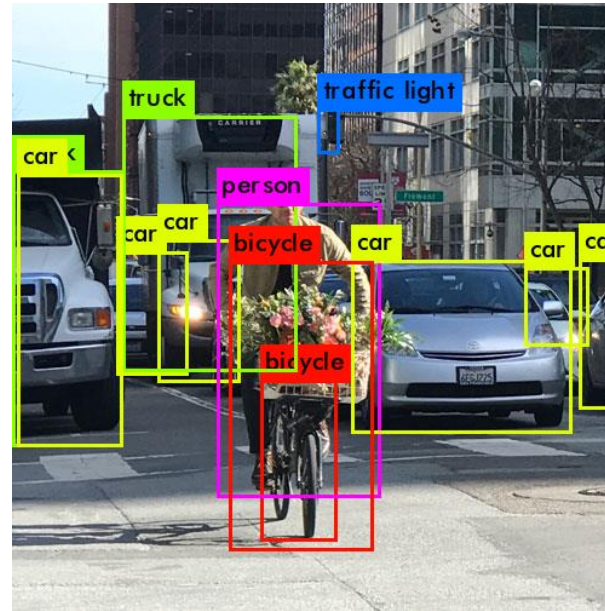
딥러닝 = 우수한 성능

음성인식

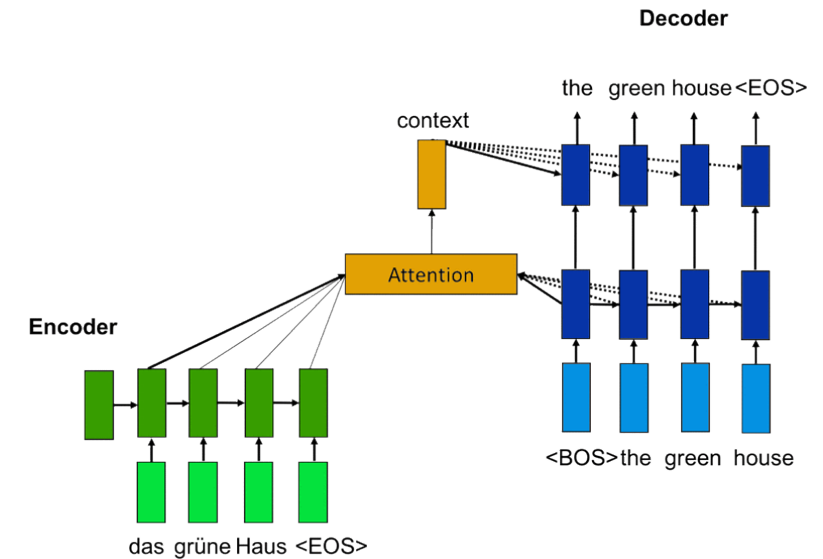
The dramatic impact of Deep Learning on Speech Recognition



이미지 처리



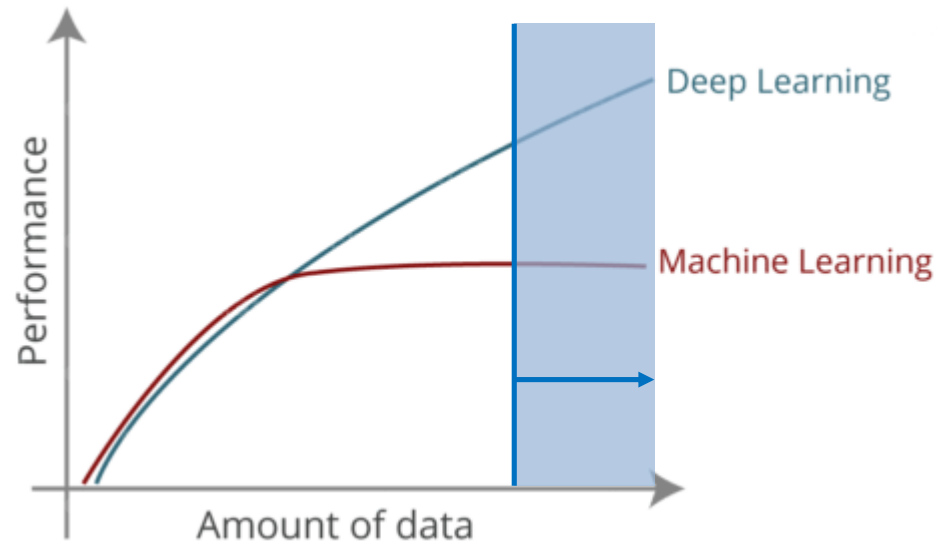
자연어 처리



Introduction

Background

딥러닝 모델의 성능을 위해서는
많은 양의 데이터가 필요!



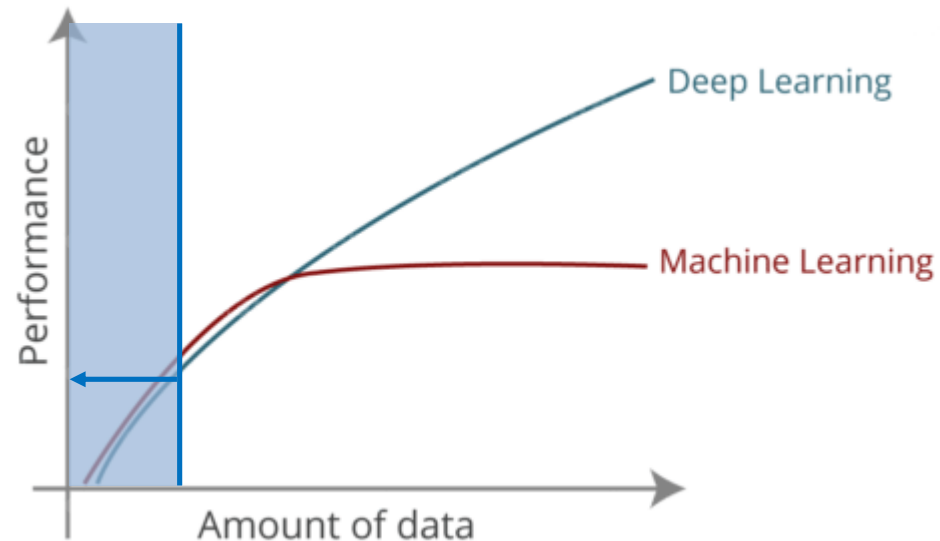
Introduction

Background

현실 데이터는 계속 변화하며,
해야 하는 일도 변화하는 경우가 대부분



충분한 양의 데이터를 사용할 수 없는 경우
좋은 성능을 보장할 수 없음
(데이터 의존도 높음)



데이터 의존도를 낮추기 위한 방법들

Domain adaptation

Semi-supervised learning

Self-supervised learning

Meta-learning

⋮



데이터 의존도를 낮추기 위한 방법들

Domain adaptation

Semi-supervised learning

Self-supervised learning

Meta-learning

⋮

Meta-learning?

학습을 위한 학습(Learning to learn)

학습 : 데이터를 이용한 일반화

학습을 잘 한다 : 적은 데이터로도 적절한 일반화가 가능하다

Meta-learning?

학습을 위한 학습(Learning to learn)

학습 : 데이터를 이용한 일반화

학습을 잘 한다 : 적은 데이터로도 적절한 일반화가 가능하다



적은 데이터로도 일반화가 가능하도록 **학습**시키자

Meta-learning

Meta-learning?

학습을 위한 학습(Learning to learn)

학습 : 데이터를 이용한 일반화

학습을 잘 한다 : 적은 데이터로도 적절한 일반화가 가능하다



적은 데이터로도 일반화가 가능하도록 학습시키자

Few-shot learning

Few-shot learning

Definition

❖ N-way, K-shot classification

- 학습 데이터의 양, N : class의 수, K : example 수



Few-shot learning

Definition

❖ N-way, K-shot classification

- 학습 데이터의 양, N : class의 수, K : example 수

N = 3, K = 2인 예시



Few-shot learning

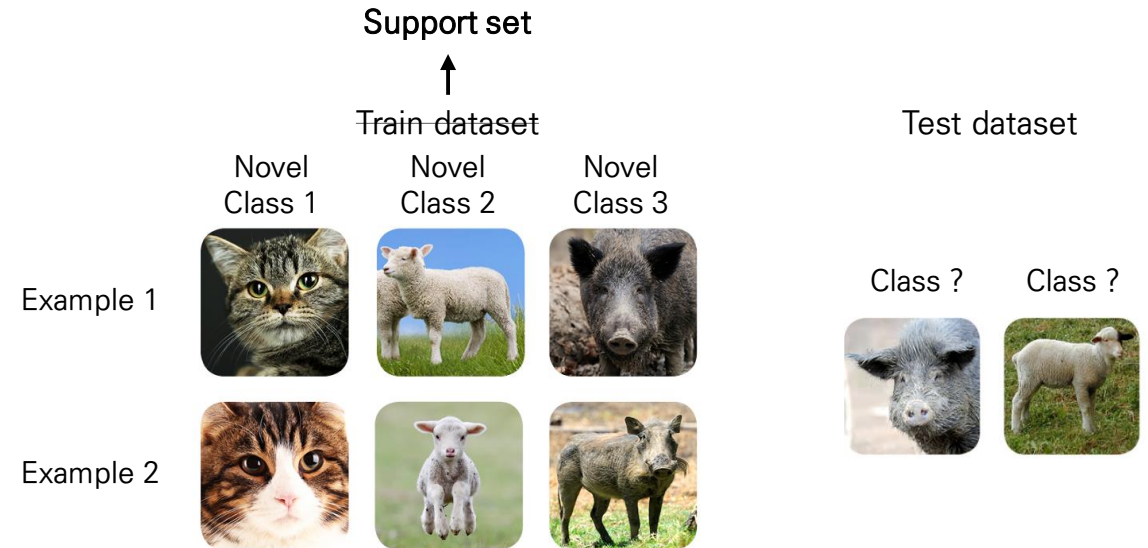
Definition

❖ N-way, K-shot classification

- 학습 데이터의 양, N : class의 수, K : example 수

❖ Support set

- 새로운 클래스에 대한 학습 데이터셋



Few-shot learning

Definition

❖ N-way, K-shot classification

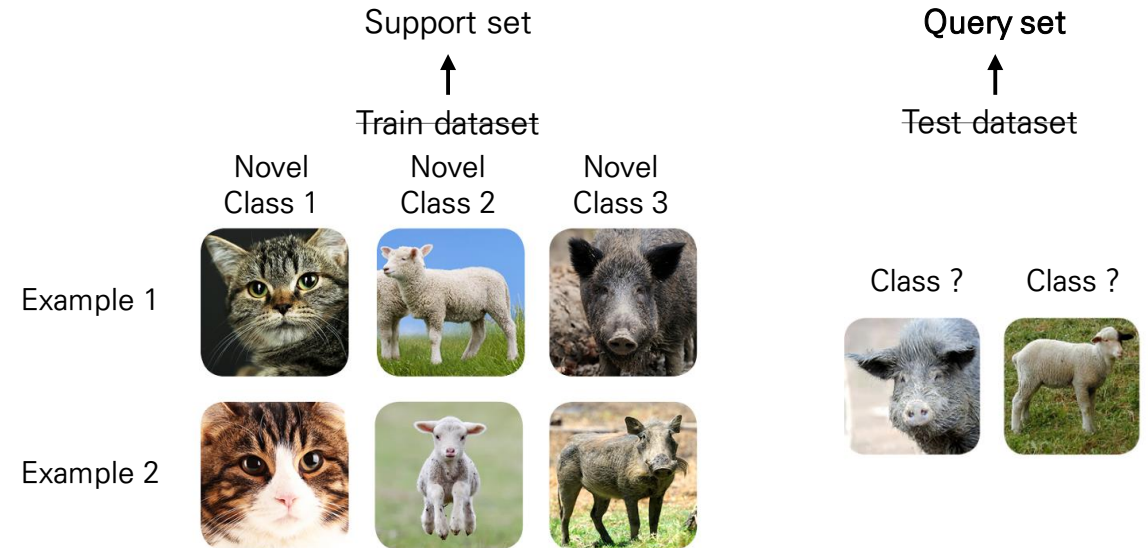
- 학습 데이터의 양, N : class의 수, K : example 수

❖ Support set

- 새로운 클래스에 대한 학습 데이터셋

❖ Query set

- 새로운 클래스에 대한 테스트 데이터셋



Few-shot learning

Definition

❖ N-way, K-shot classification

- 학습 데이터의 양, N : class의 수, K : example 수

❖ Support set

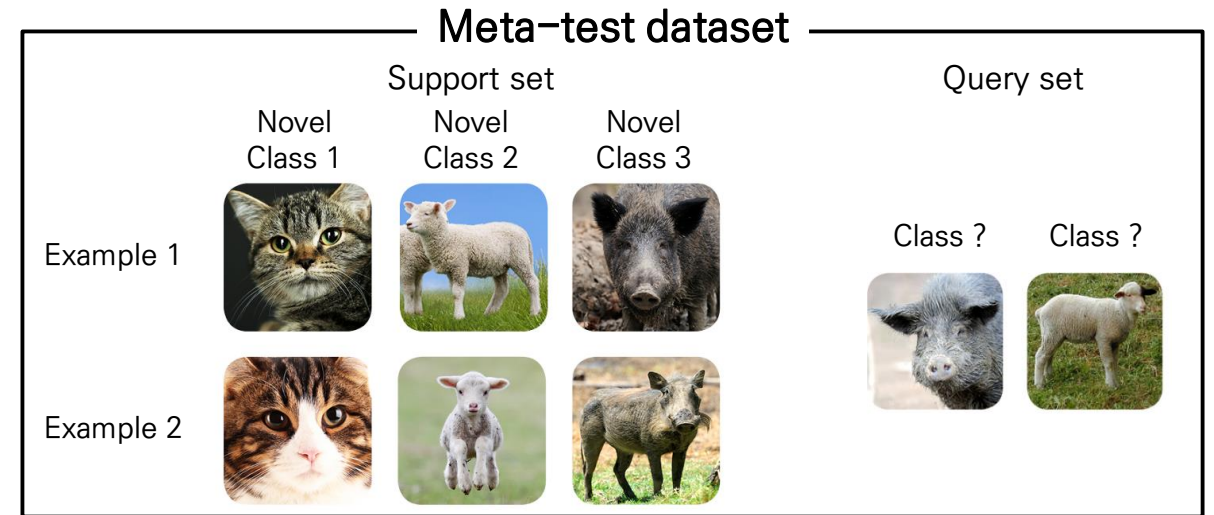
- 새로운 클래스에 대한 학습 데이터셋

❖ Query set

- 새로운 클래스에 대한 테스트 데이터셋

❖ Meta-test dataset

- 새로운 클래스에 대한 Support set + Query set



Few-shot learning

Definition

❖ N-way, K-shot classification

- 학습 데이터의 양, N : class의 수, K : example 수

❖ Support set

- 새로운 클래스에 대한 학습 데이터셋

❖ Query set

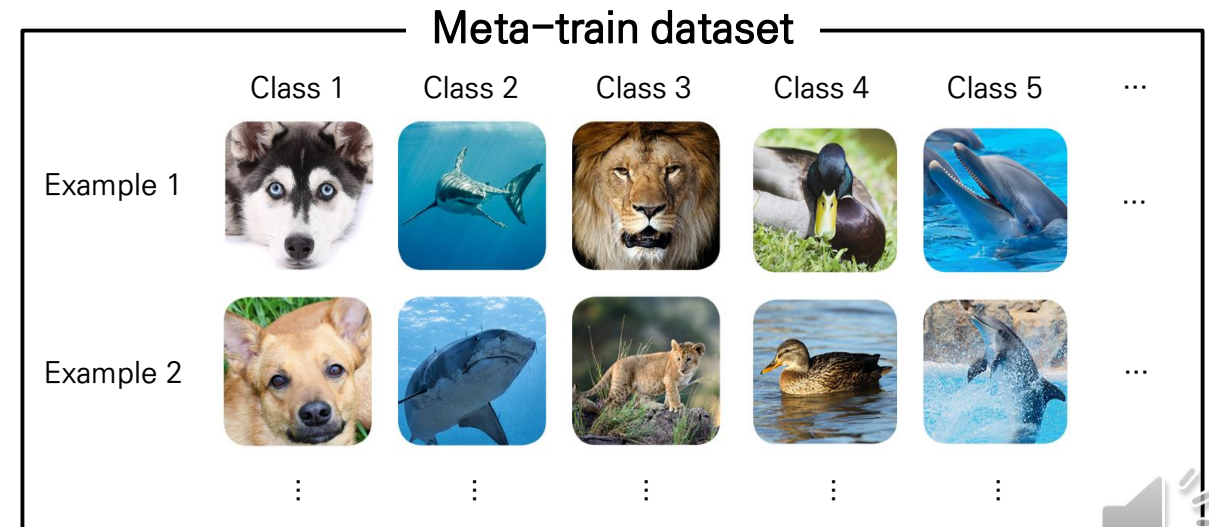
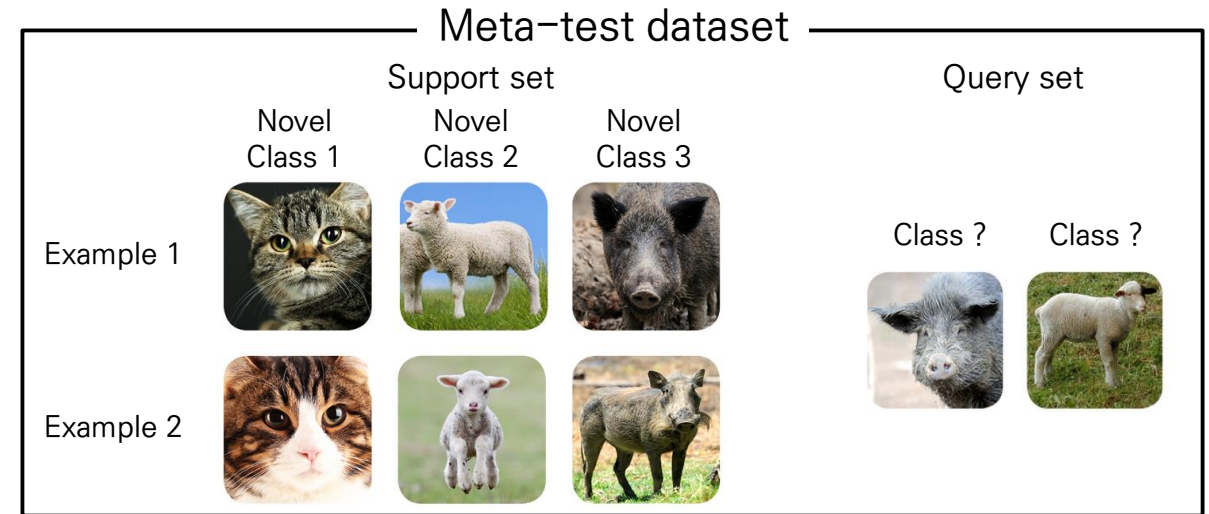
- 새로운 클래스에 대한 테스트 데이터셋

❖ Meta-test dataset

- 새로운 클래스에 대한 Support set + Query set

❖ Meta-train dataset

- Meta-learning에 사용되는 데이터셋

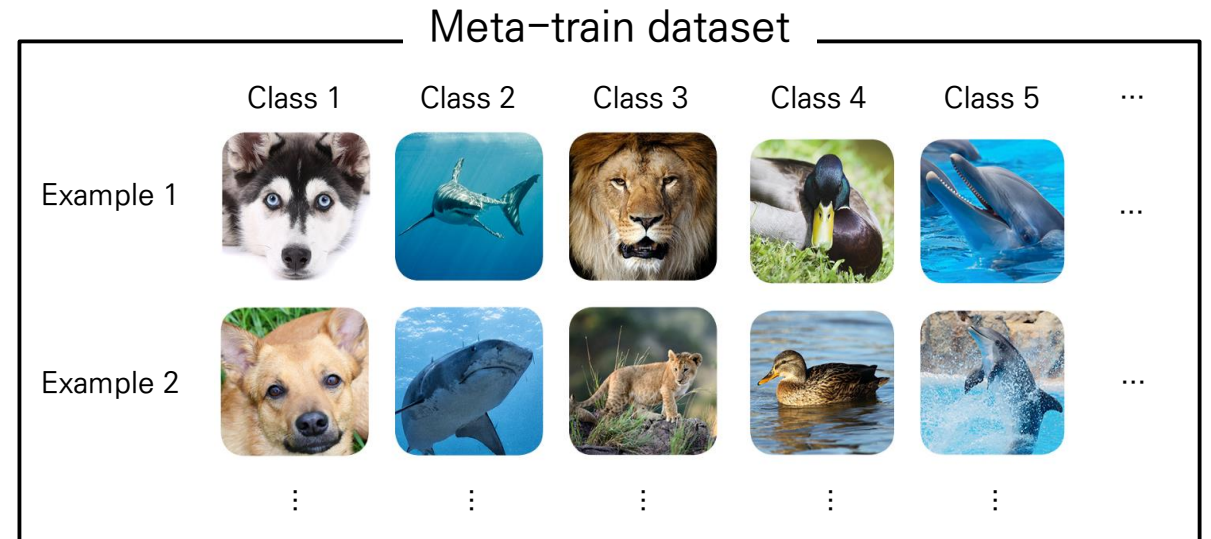


Few-shot learning

Training strategy

❖ Episode training

$$\theta = \operatorname{argmax}_{\theta} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$



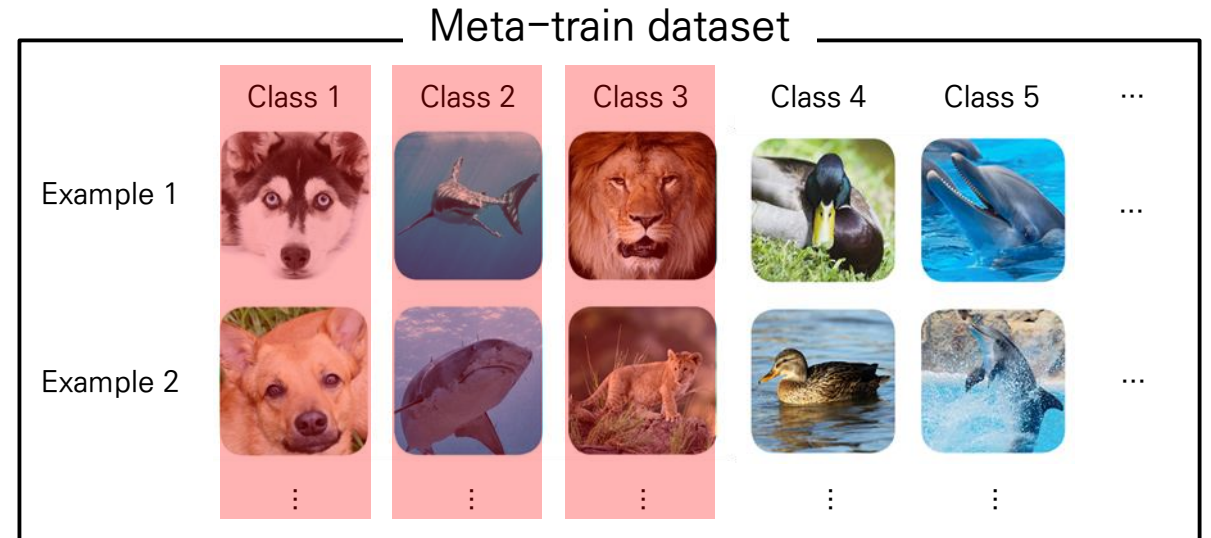
Few-shot learning

Training strategy

❖ Episode training

$$\theta = \operatorname{argmax}_{\theta} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$

- Task sampling ($E_{L \sim T}$)
 - 모든 class 중 학습시킬 N개 class 샘플링



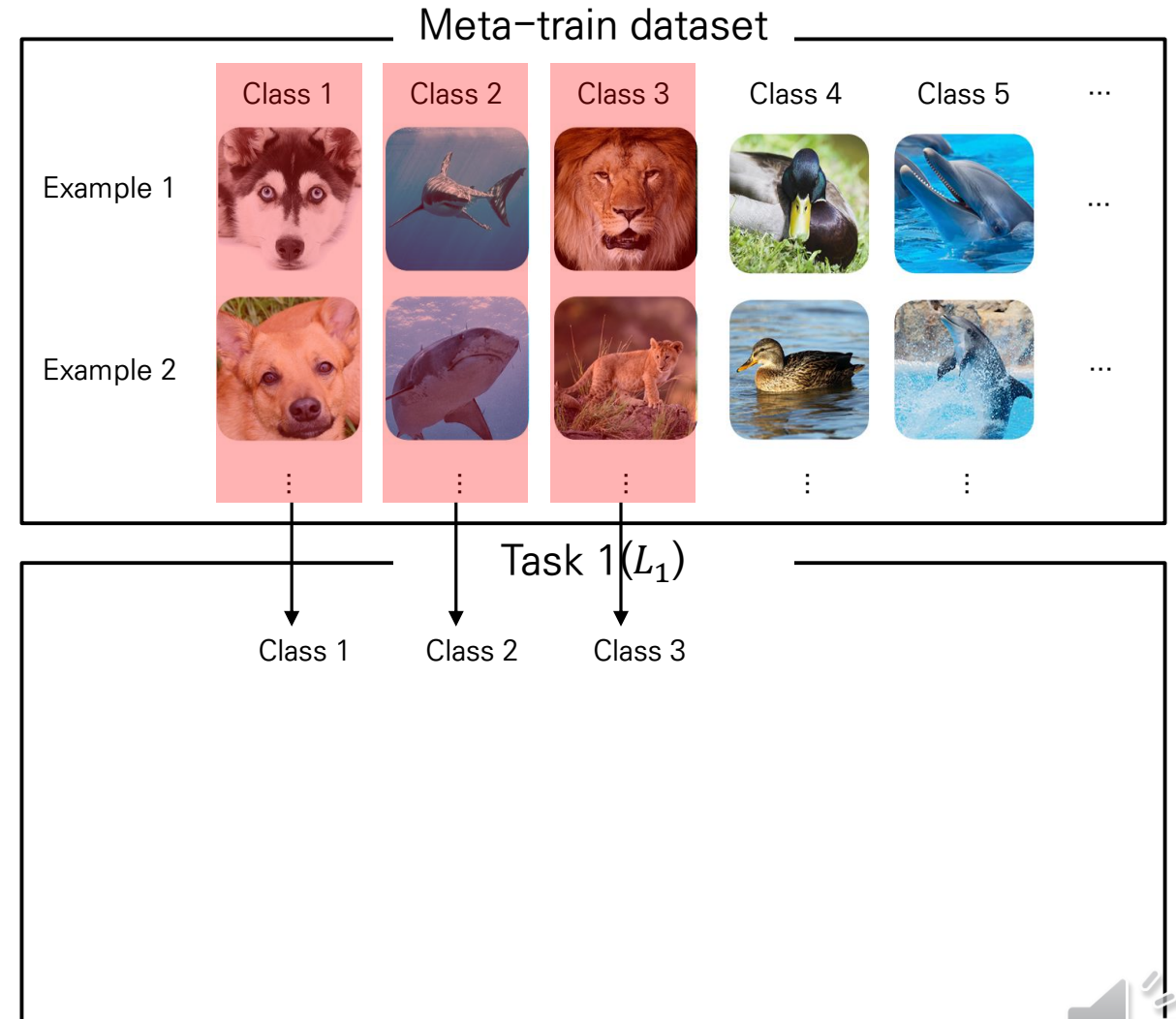
Few-shot learning

Training strategy

❖ Episode training

$$\theta = \operatorname{argmax}_{\theta} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$

- Task sampling ($E_{L \sim T}$)
 - 모든 class 중 학습시킬 N개 class 샘플링



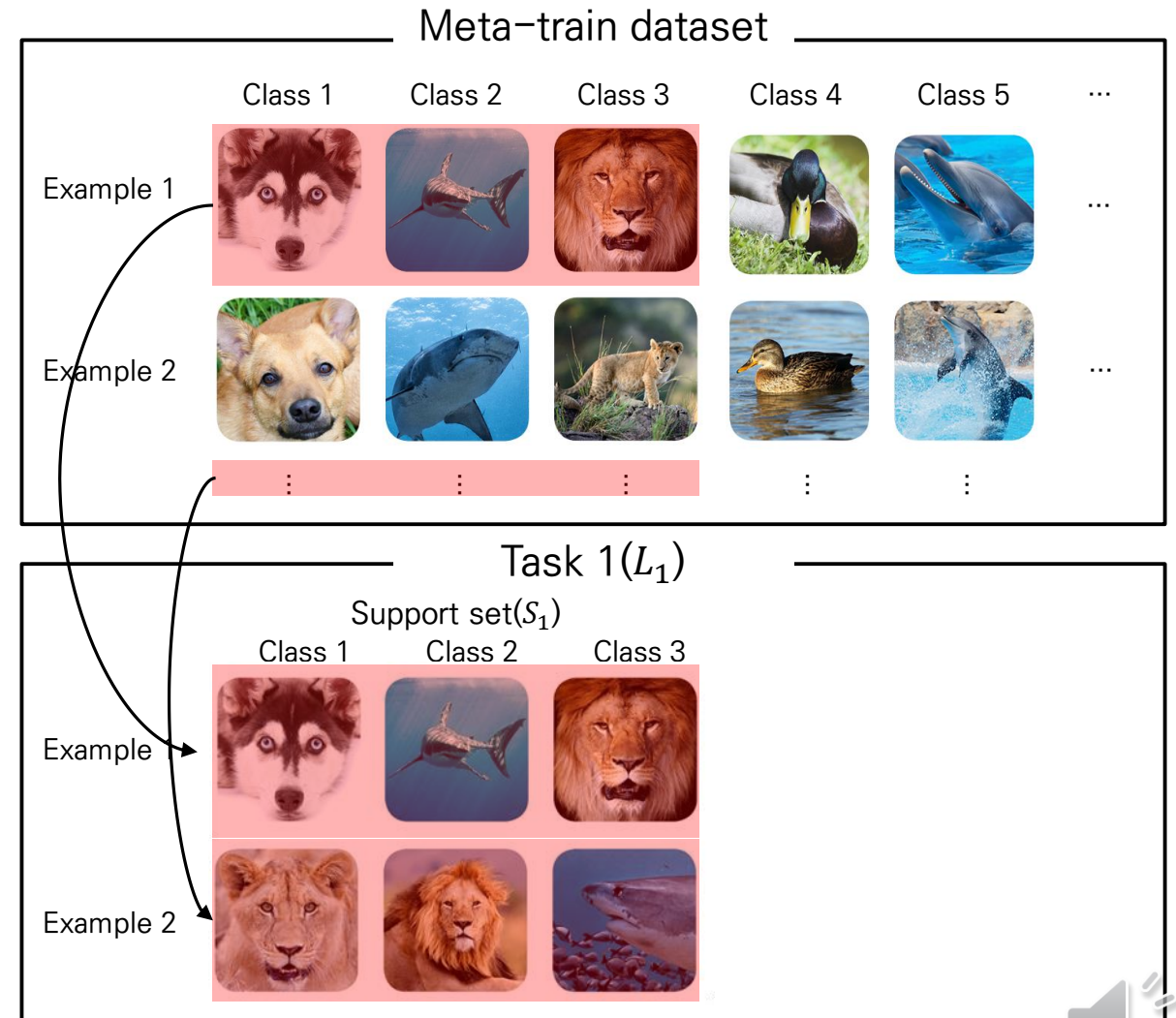
Few-shot learning

Training strategy

❖ Episode training

$$\theta = \operatorname{argmax}_{\theta} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$

- Task sampling ($E_{L \sim T}$)
 - 모든 class 중 학습시킬 N개 class 샘플링
- Support set sampling ($E_{S \sim L}$)
 - N개 class 별 각각 t개의 examples 샘플링



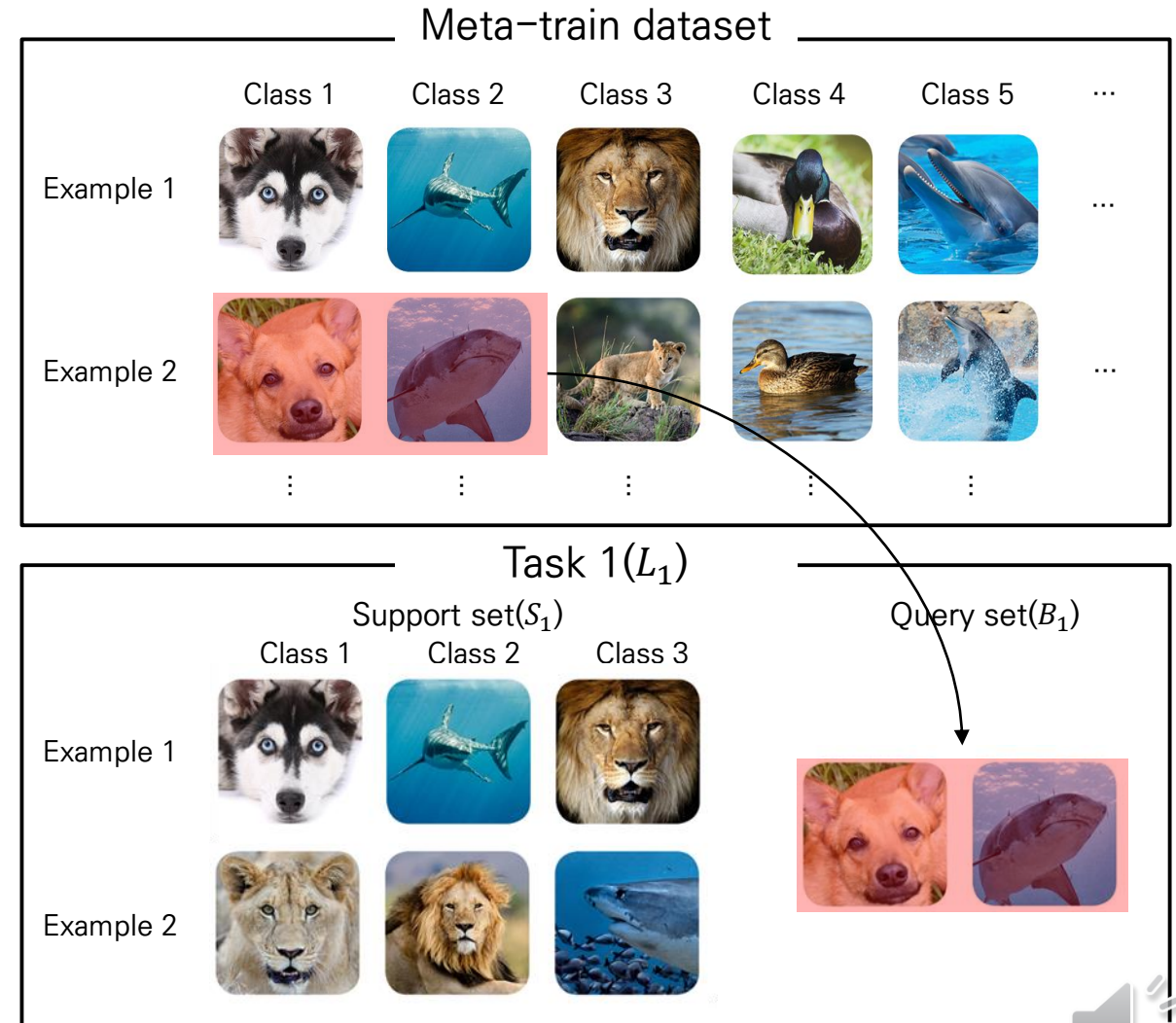
Few-shot learning

Training strategy

❖ Episode training

$$\theta = \operatorname{argmax}_{\theta} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$

- Task sampling ($E_{L \sim T}$)
 - 모든 class 중 학습시킬 N개 class 샘플링
- Support set sampling ($E_{S \sim L}$)
 - N개 class 별 각각 t개의 examples 샘플링
- Query set sampling ($E_{B \sim L}$)
 - N개 class 별 각각 u개의 examples 샘플링



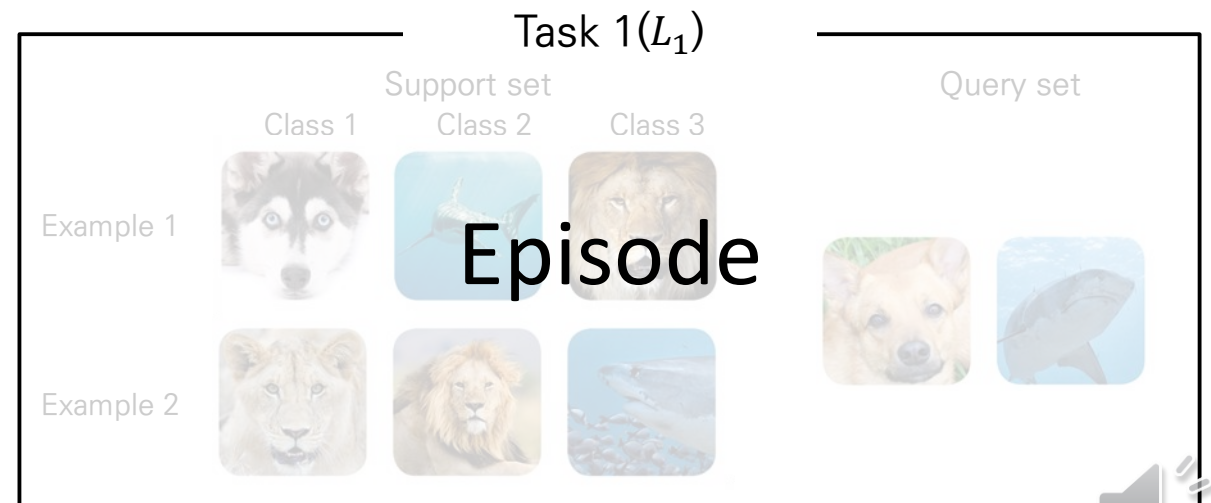
Few-shot learning

Training strategy

❖ Episode training

$$\theta = \operatorname{argmax}_{\theta} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$

- Task sampling ($E_{L \sim T}$)
 - 모든 class 중 학습시킬 N개 class 샘플링
- Support set sampling ($E_{S \sim L}$)
 - N개 class 별 각각 t개의 examples 샘플링
- Query set sampling ($E_{B \sim L}$)
 - N개 class 별 각각 u개의 examples 샘플링



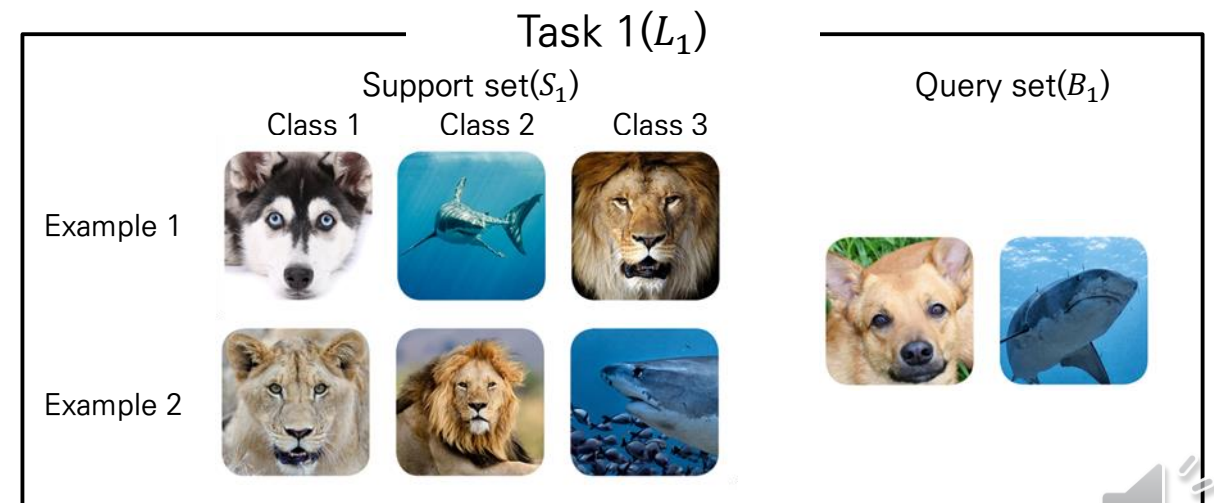
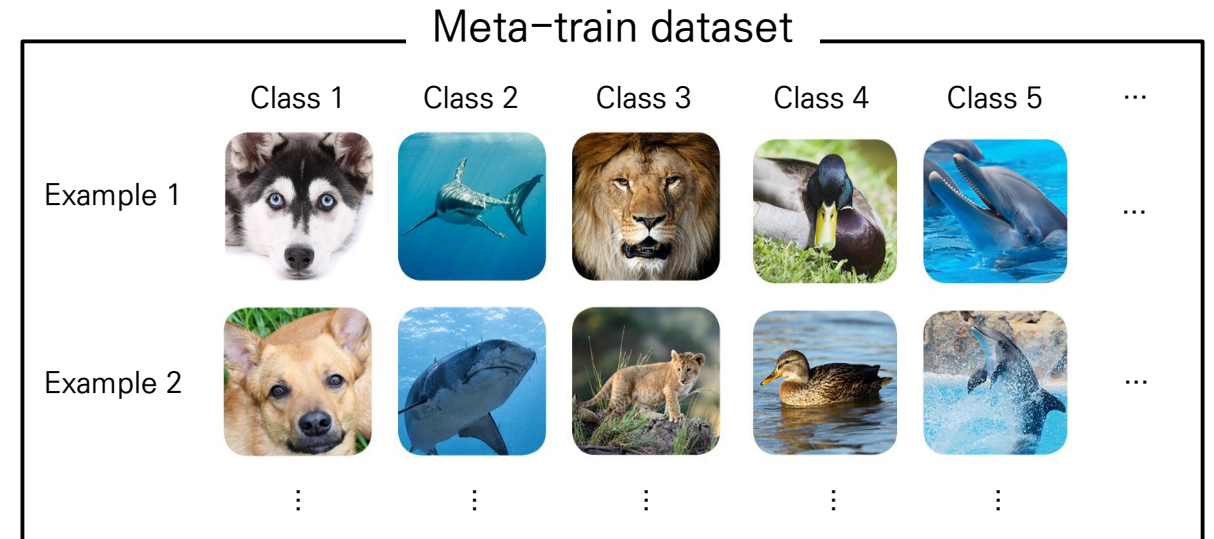
Few-shot learning

Training strategy

❖ Episode training

$$\theta = \underset{\theta}{\operatorname{argmax}} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x,y) \in B} \log P_{\theta}(y|x, S) \right] \right]$$

- Task sampling ($E_{L \sim T}$)
 - 모든 class 중 학습시킬 N개 class 샘플링
- Support set sampling ($E_{S \sim L}$)
 - N개 class 별 각각 t개의 examples 샘플링
- Query set sampling ($E_{B \sim L}$)
 - N개 class 별 각각 u개의 examples 샘플링
- Query set의 class를 가장 잘 맞추는 파라미터 θ 학습

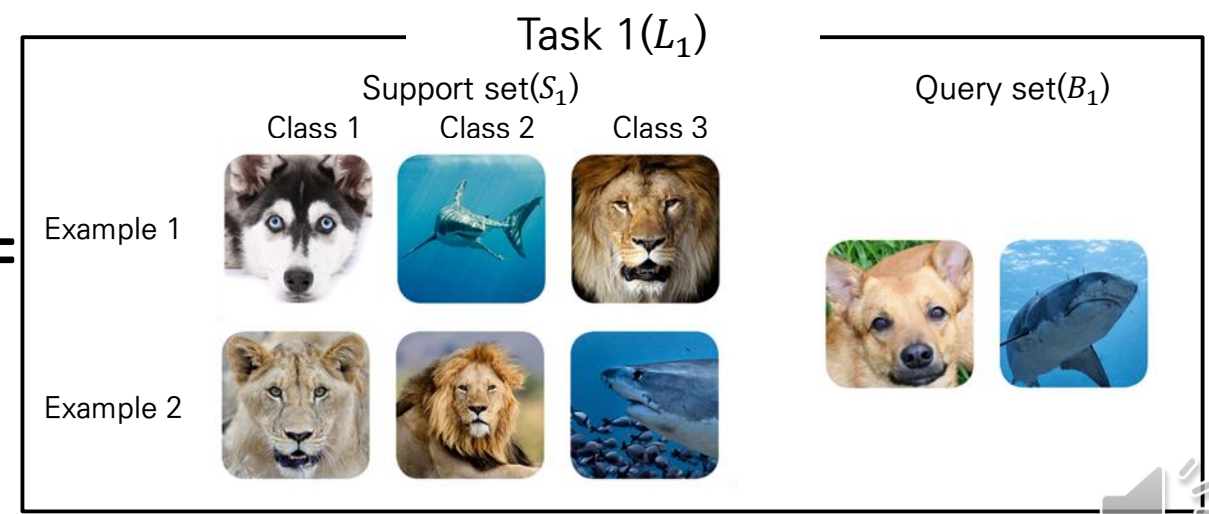
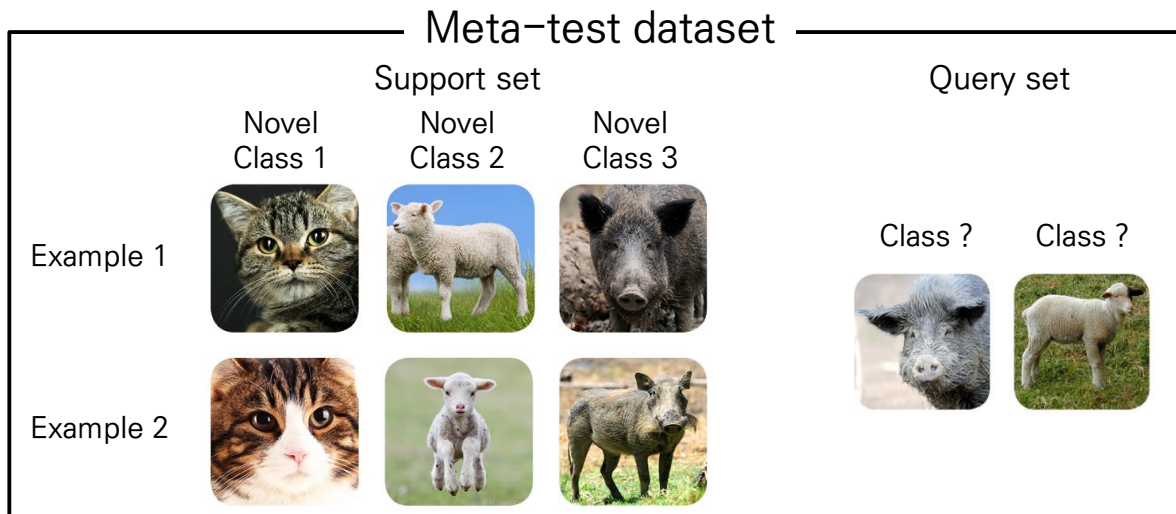
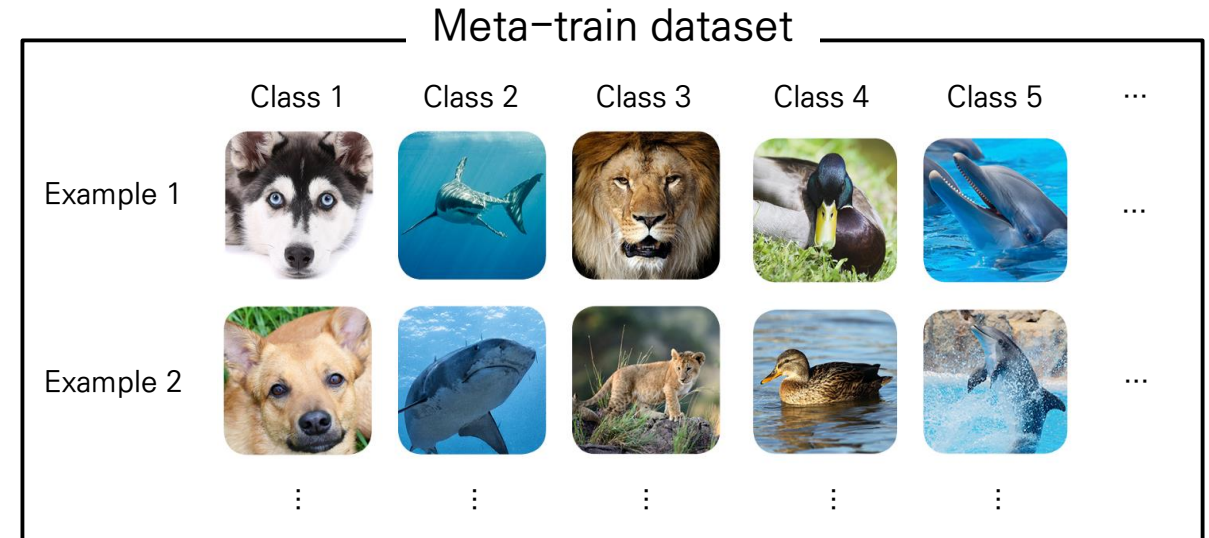


Few-shot learning

Training strategy

❖ Episode training

- Meta-test와 동일한 학습 환경
- 동일한 환경에서 좋은 성능을 내도록 학습



Metric-based approaches to meta-learning

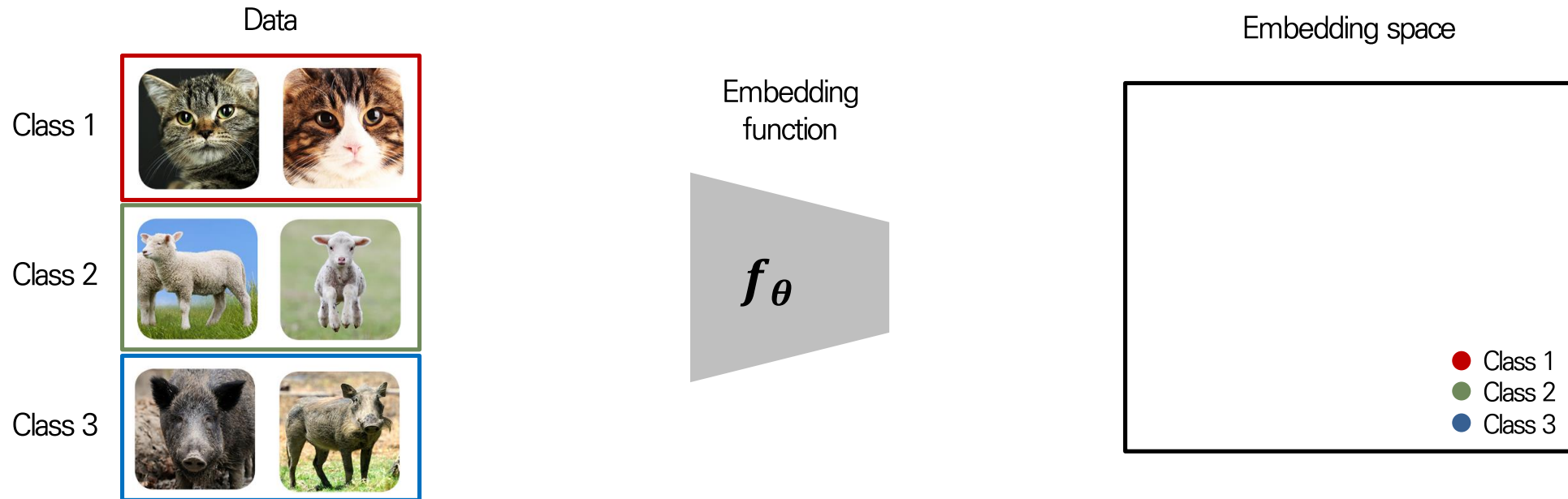
Metric learning

❖ Embedding function

- 데이터를 저차원으로 임베딩하는 함수

❖ Distance

- 임베딩 공간에서 데이터간 거리



Metric-based approaches to meta-learning

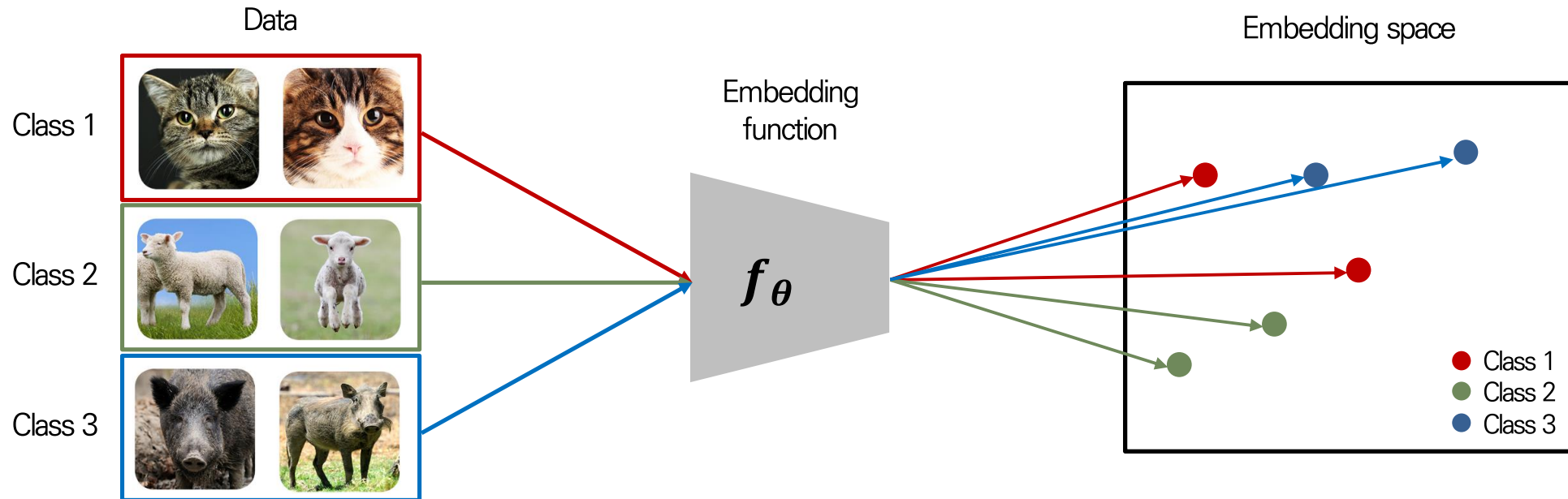
Metric learning

❖ Embedding function

- 데이터를 저차원으로 임베딩하는 함수

❖ Distance

- 임베딩 공간에서 데이터간 거리



Metric-based approaches to meta-learning

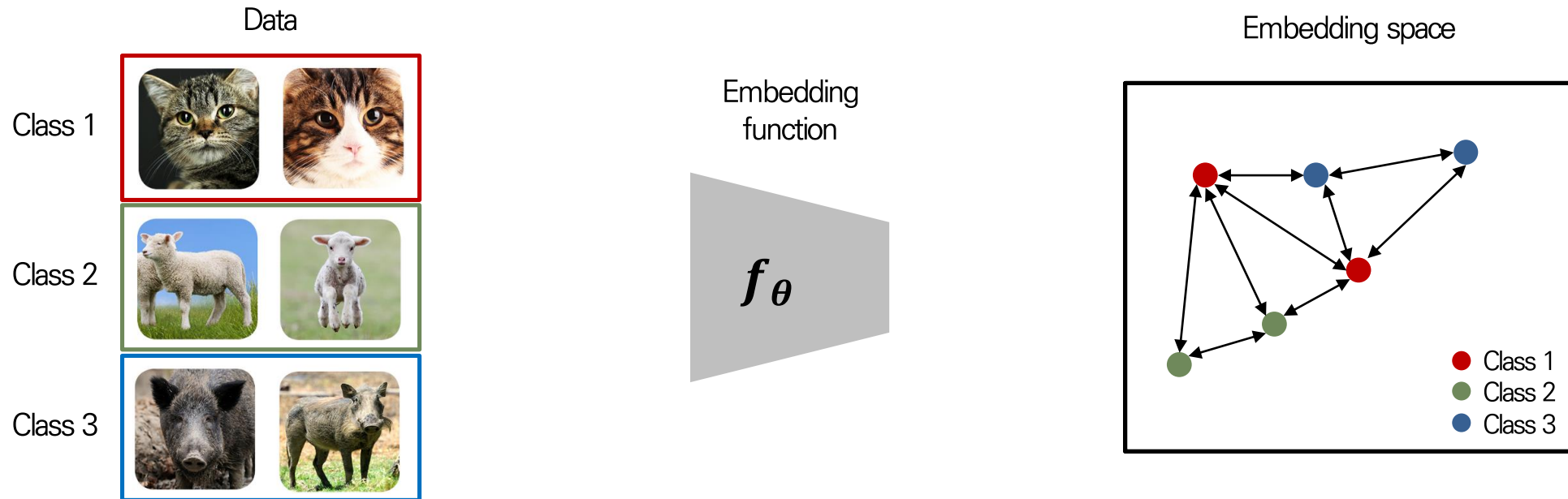
Metric learning

❖ Embedding function

- 데이터를 저차원으로 임베딩하는 함수

❖ Distance

- 임베딩 공간에서 데이터간 거리



Metric-based approaches to meta-learning

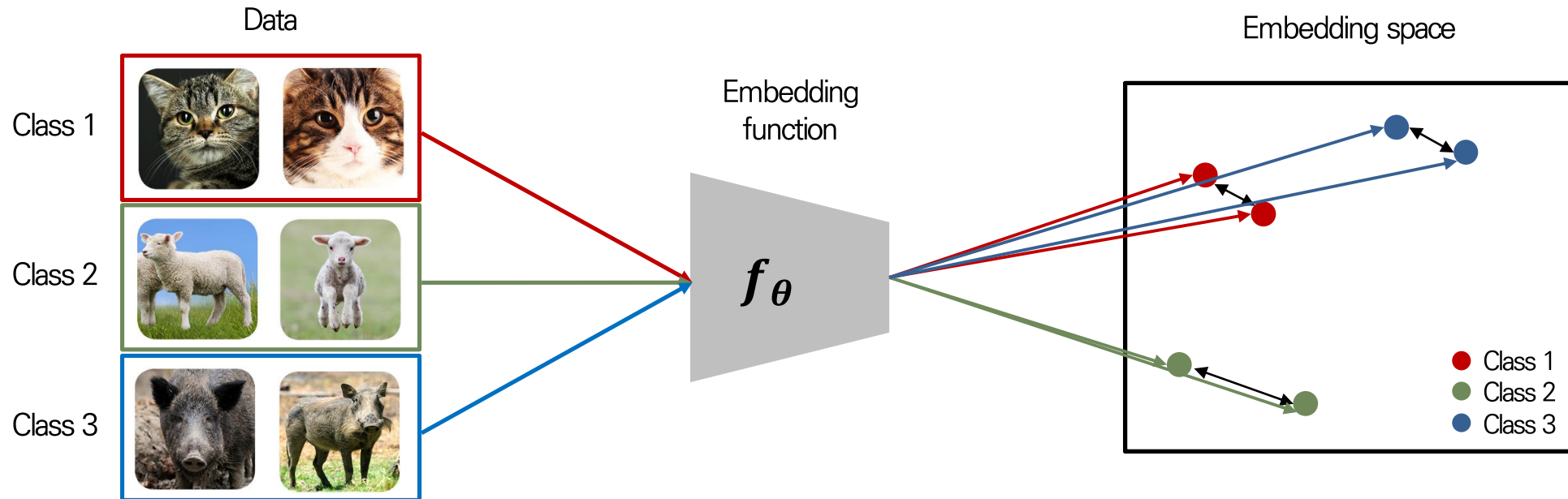
Metric learning

❖ Embedding function

- 데이터를 저차원으로 임베딩하는 함수

❖ Distance

- 임베딩 공간에서 데이터간 거리



Metric-based approaches to meta-learning

Matching networks

- Advances in Neural Information Processing Systems 29 (NIPS 2016)
- Google DeepMind
- 1967회 인용(2020.11.05 기준)

Matching Networks for One Shot Learning

Oriol Vinyals
Google DeepMind
vinyals@google.com

Charles Blundell
Google DeepMind
cblundell@google.com

Timothy Lillicrap
Google DeepMind
countzero@google.com

Koray Kavukcuoglu
Google DeepMind
korayk@google.com

Daan Wierstra
Google DeepMind
wierstra@google.com

Abstract

Learning from a few examples remains a key challenge in machine learning. Despite recent advances in important domains such as vision and language, the



Metric-based approaches to meta-learning

Matching networks

❖ Embedding function

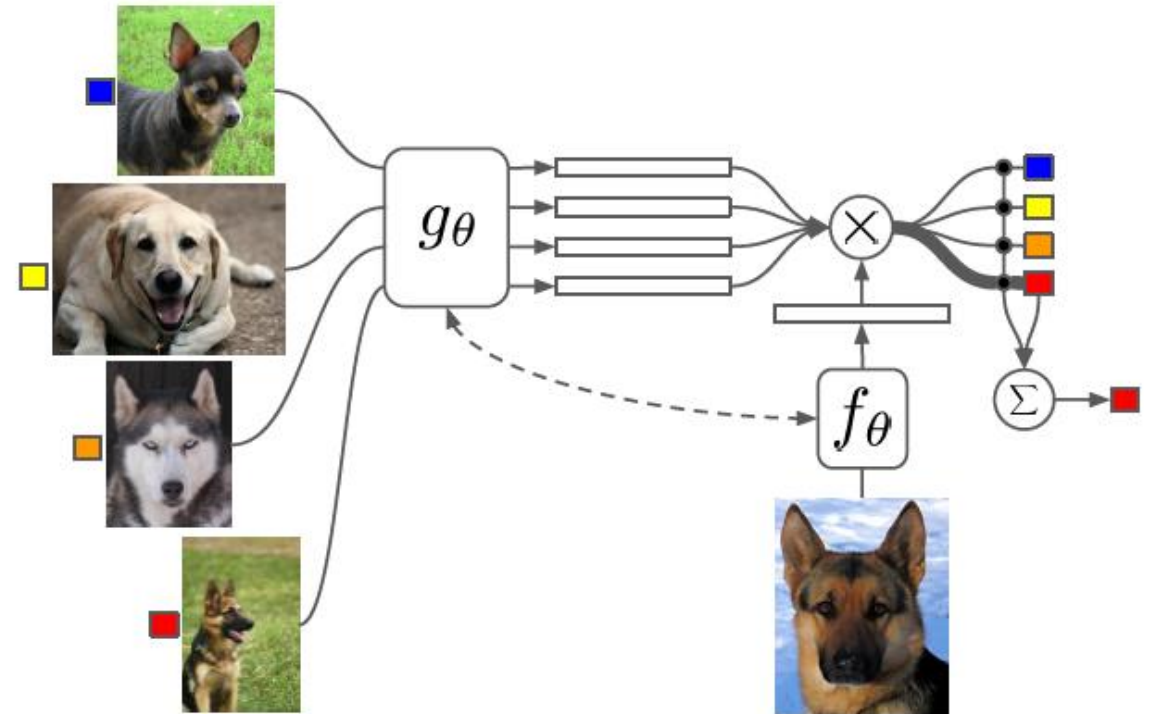


Figure 1: Matching Networks architecture

Metric-based approaches to meta-learning

Matching networks

❖ Embedding function

- g_θ : Support set의 임베딩 함수
- f_θ : Query set의 임베딩 함수

1. Simple embedding ($g_\theta = f_\theta$)

Conv-4

2. Full contextual embeddings ($g_\theta \neq f_\theta$)

g_θ : bi-LSTM

f_θ : attLSTM

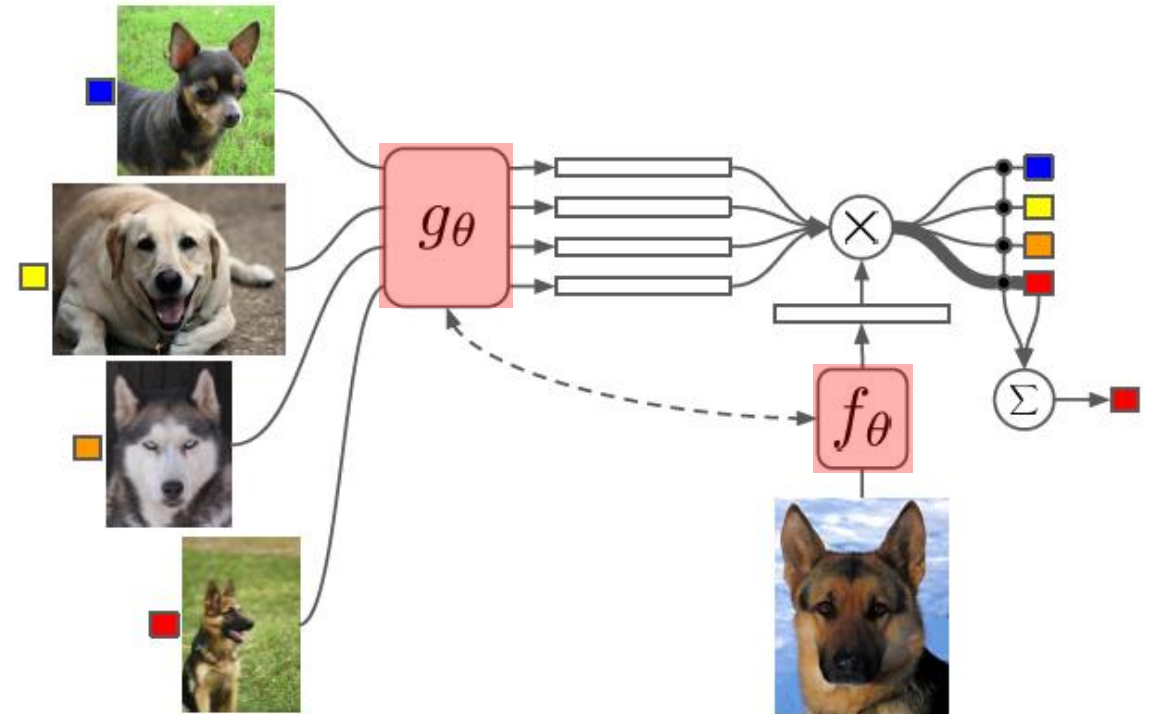


Figure 1: Matching Networks architecture

Metric-based approaches to meta-learning

Matching networks

❖ Embedding function

- g_θ : Support set의 임베딩 함수
- f_θ : Query set의 임베딩 함수

1. Simple embedding($g_\theta = f_\theta$)

Conv-4

2. Full contextual embeddings($g_\theta \neq f_\theta$)

g_θ : bi-LSTM

f_θ : attLSTM

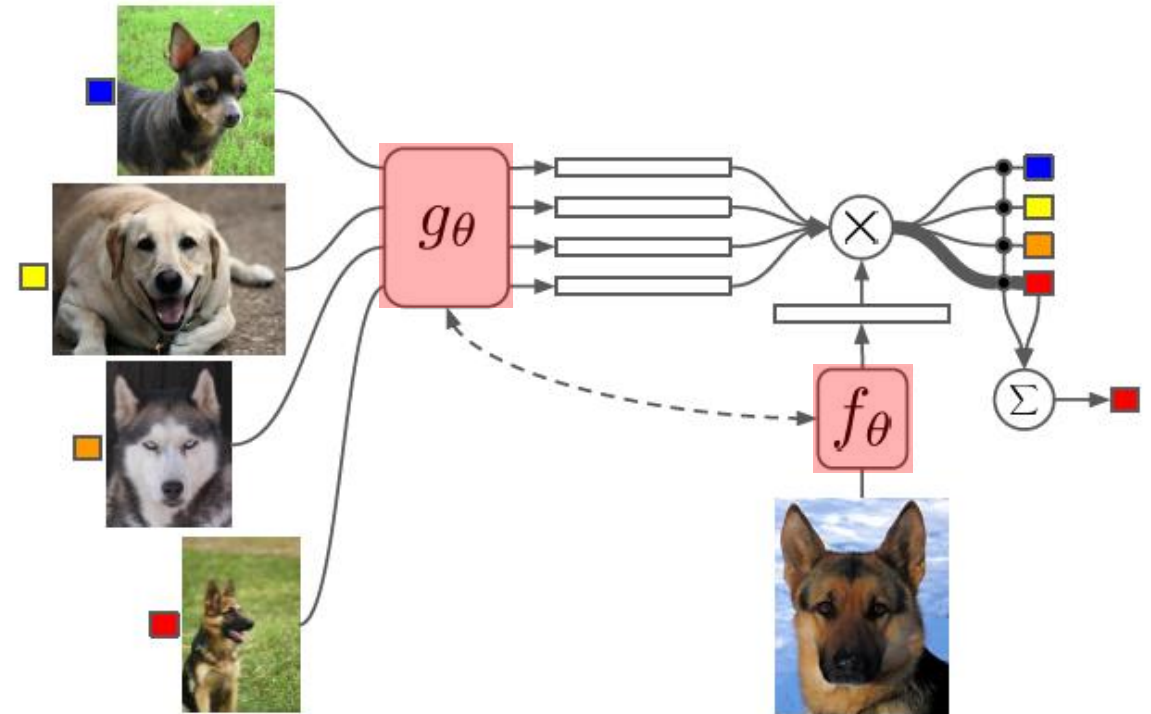


Figure 1: Matching Networks architecture

Metric-based approaches to meta-learning

Matching networks

❖ Distance

- Cosine similarity(attention)

$$a(x, x_i) = \frac{\exp(\text{cosine}(f(x), g(x_i)))}{\sum_{j=1}^k (\text{cosine}(f(x), g(x_j)))}$$

❖ Training strategy

- Episode training

$$\theta = \underset{\theta}{\operatorname{argmax}} E_{L \sim T} \left[E_{S \sim L, B \sim L} \left[\sum_{(x, y) \in B} \log P_{\theta}(y | x, S) \right] \right]$$

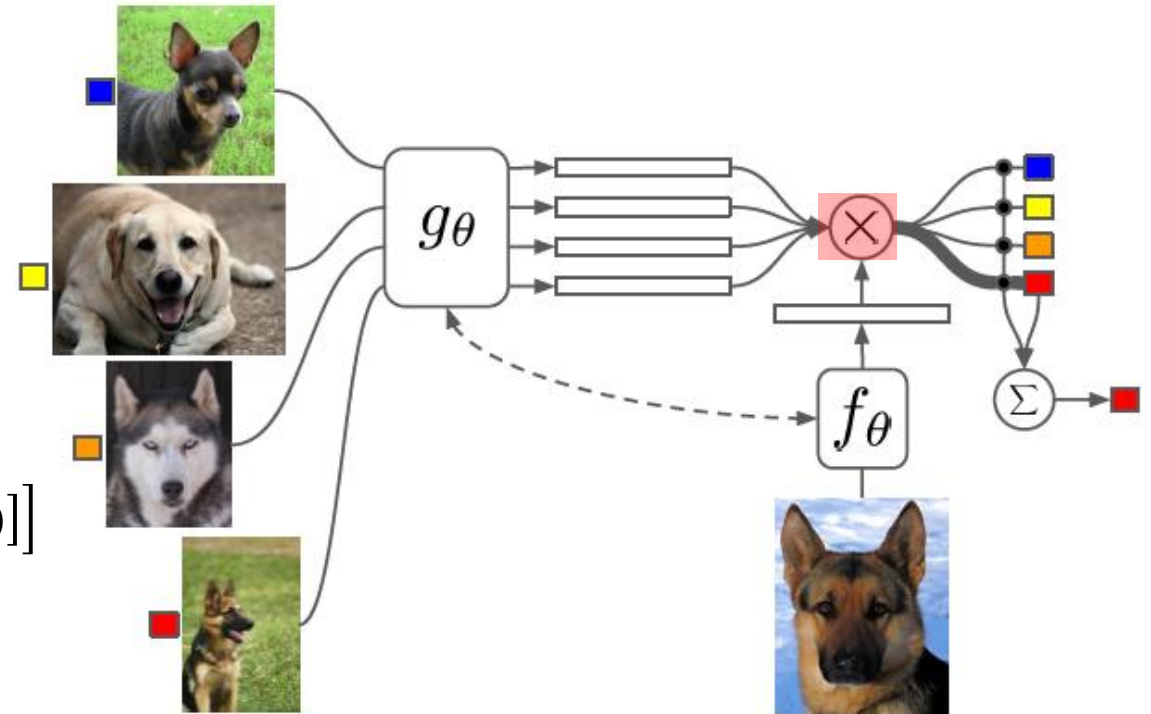


Figure 1: Matching Networks architecture

Metric-based approaches to meta-learning

Prototypical networks

- Advances in Neural Information Processing Systems 30 (NIPS 2017)
- 1632회 인용(2020.11.05 기준)

Prototypical Networks for Few-shot Learning

Jake Snell
University of Toronto*
Vector Institute

Kevin Swersky
Twitter

Richard Zemel
University of Toronto
Vector Institute
Canadian Institute for Advanced Research

Abstract

We propose *Prototypical Networks* for the problem of few-shot classification, where a classifier must generalize to new classes not seen in the training set, given only a small number of examples of each new class. Prototypical Networks learn a metric space in which classification can be performed by computing distances to prototype representations of each class. Compared to recent approaches for few-shot learning, they reflect a simpler inductive bias that is beneficial in this limited-data regime, and achieve excellent results. We provide an analysis showing that some simple design decisions can yield substantial improvements over recent



Metric-based approaches to meta-learning

Prototypical networks

❖ Embedding function

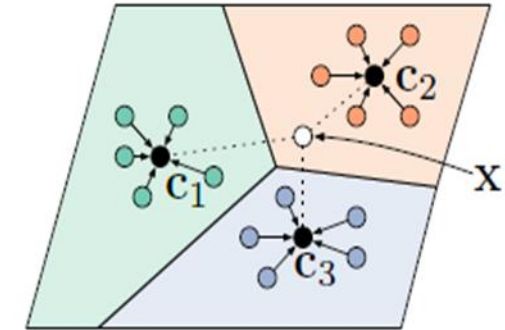
- f_{ϕ} : support, query embedding function, Conv-4 사용

❖ Prototype

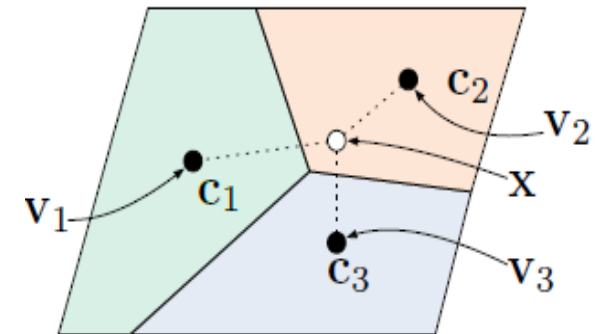
- $c_k = \frac{1}{|S_k|} \sum_{(x_i, y) \in S_k} f_{\phi}(x_i) \rightarrow c_k$ 는 class k 의 prototype
- 모든 Support embedding vector와 거리계산하지 않아도 됨
- Zero-shot learning에도 사용할 수 있음

❖ Distance

- Euclidean distance 사용
- Cosine similarity 보다 Euclidean distance가 더 좋음
- 미분가능한 모든 distance 사용 가능



(a) Few-shot



(b) Zero-shot

Metric-based approaches to meta-learning

Relation Network

- Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition(CVPR 2018)
- 852회 인용(2020.11.05 기준)

Learning to Compare: Relation Network for Few-Shot Learning

Flood Sung¹ Yongxin Yang³ Li Zhang² Tao Xiang¹ Philip H.S. Torr² Timothy M. Hospedales³
¹Queen Mary University of London ²University of Oxford ³The University of Edinburgh

floodsung@gmail.com t.xiang@qmul.ac.uk

{lz, phst}@robots.ox.ac.uk {yongxin.yang, t.hospedales}@ed.ac.uk

Abstract

We present a conceptually simple, flexible, and general framework for few-shot learning, where a classifier must learn to recognise new classes given only few examples from each. Our method, called the Relation Network (RN), is trained end-to-end from scratch. During meta-learning, it learns to learn a deep distance metric to compare a small number of images within episodes, each of which is designed to simulate the few-shot setting. Once trained, a RN is able to classify images of new classes by computing relation scores between query images and the few examples of each new class without further updating the network. Be-

zero-shot [11, 3, 24, 45, 25, 31] learning.

Few-shot learning aims to recognise novel visual categories from very few labelled examples. The availability of only one or very few examples challenges the standard ‘fine-tuning’ practice in deep learning [10]. Data augmentation and regularisation techniques can alleviate overfitting in such a limited-data regime, but they do not solve it. Therefore contemporary approaches to few-shot learning often decompose training into an auxiliary meta learning phase where transferrable knowledge is learned in the form of good initial conditions [10], embeddings [36, 39] or optimisation strategies [29]. The target few-shot learning problem is then learned by fine-tuning [10] with the learned op-

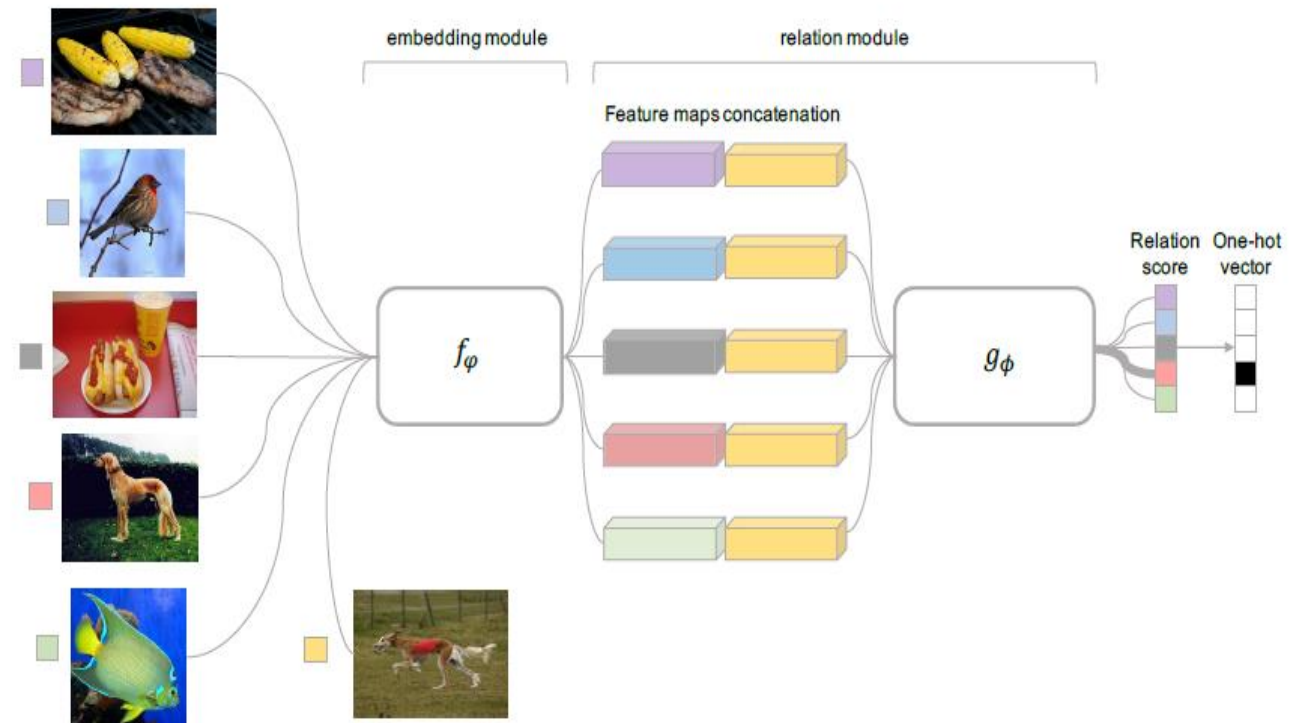


Metric-based approaches to meta-learning

Relation Network

❖ Embedding function

- f_ϕ : support, query embedding function
- Conv-4 사용



Metric-based approaches to meta-learning

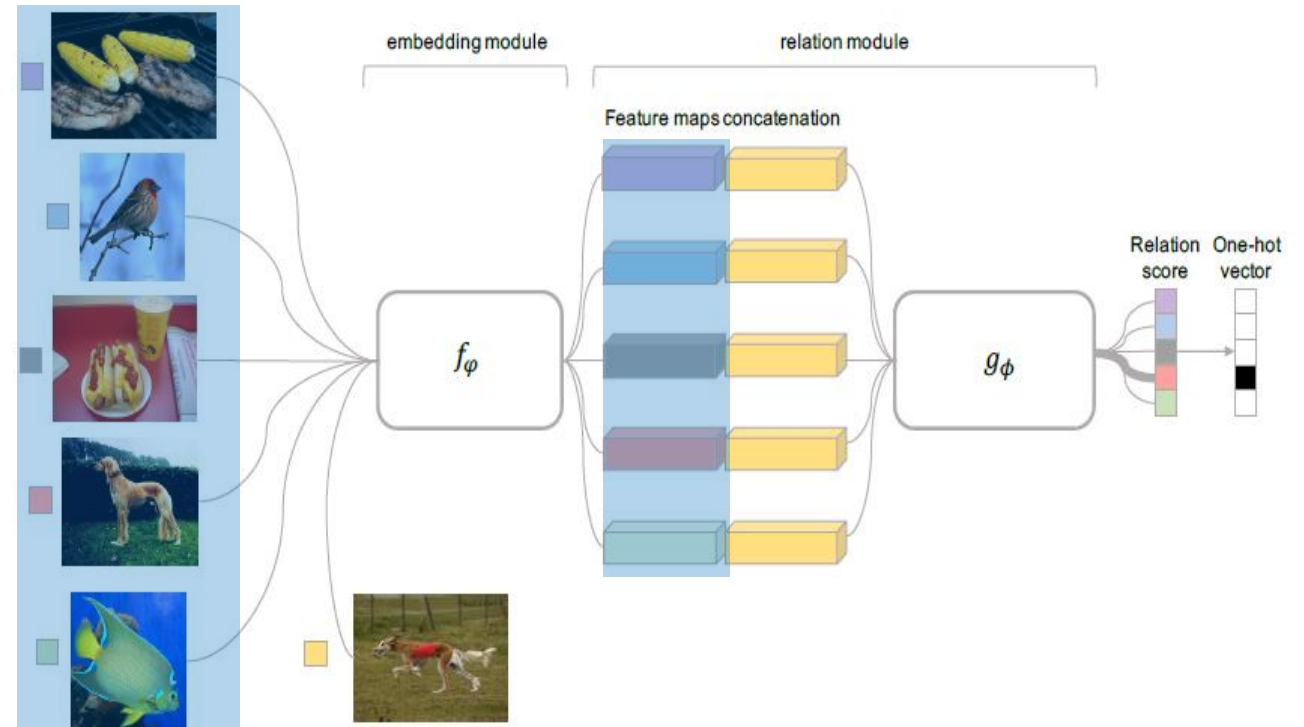
Relation Network

❖ Embedding function

- f_ϕ : support, query embedding function
- Conv-4 사용

❖ Distance

- Relation module
 1. Support input의 feature map과 query input의 feature map을 붙임
 2. 합쳐진 featuremap으로 relation score(r_{ij})를 예측하는 모델 학습



$$\phi, \emptyset = \underset{\phi, \emptyset}{\operatorname{argmin}} \sum_{i=1}^m \sum_{j=1}^n (r_{ij} - 1(y_i == y_j))^2$$

Metric-based approaches to meta-learning

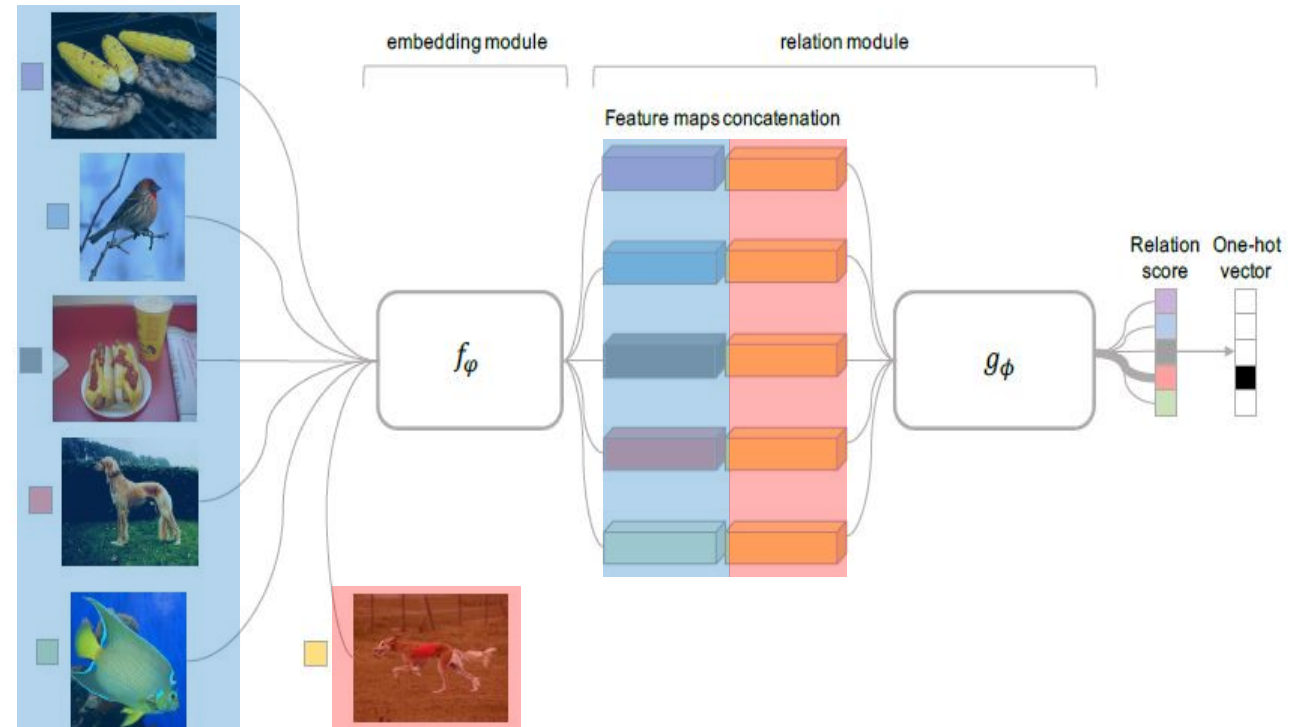
Relation Network

❖ Embedding function

- f_ϕ : support, query embedding function
- Conv-4 사용

❖ Distance

- Relation module
 1. Support input의 feature map과 query input의 feature map을 붙임
 2. 합쳐진 featuremap으로 relation score(r_{ij})를 예측하는 모델 학습



$$\varphi, \phi = \underset{\varphi, \phi}{\operatorname{argmin}} \sum_{i=1}^m \sum_{j=1}^n (r_{ij} - 1(y_i == y_j))^2$$

Evaluations

A closer look at few-shot classification

- 7th International Conference on Learning Representations(ICLR 2019)
- 852회 인용(2020.11.05 기준)

A CLOSER LOOK AT FEW-SHOT CLASSIFICATION

Wei-Yu Chen
Carnegie Mellon University
weiyuc@andrew.cmu.edu

Yen-Cheng Liu & Zsolt Kira
Georgia Tech
{ycliu, zkira}@gatech.edu

Yu-Chiang Frank Wang
National Taiwan University
ycwang@ntu.edu.tw

Jia-Bin Huang
Virginia Tech
jbhuang@vt.edu

ABSTRACT

Few-shot classification aims to learn a classifier to recognize unseen classes during training with limited labeled examples. While significant progress has been made, the growing complexity of network designs, meta-learning algorithms, and differences in implementation details make a fair comparison difficult. In this paper, we present 1) a consistent comparative analysis of several representative few-shot classification algorithms, with results showing that deeper backbones significantly reduce the performance differences among methods on datasets with limited domain differences, 2) a modified baseline method that surprisingly achieves competitive performance when compared with the state-of-the-art on both the *mini-*



Evaluations

A closer look at few-shot classification

❖ Models

- Baseline
- Baseline++
- MatchingNet
- ProtoNet
- MAML
- RelationNet



Evaluations

A closer look at few-shot classification

❖ Models

- Baseline
- Baseline++
- MatchingNet
- ProtoNet
- MAML
- RelationNet



Evaluations

A closer look at few-shot classification

❖ Models

- Baseline
- Baseline++

- MAML



Evaluations

A closer look at few-shot classification

❖ Models


- Baseline
- Baseline++
- **MAML**


종료


Model-agnostic meta-learning for fast adaptation of deep networks


2019. 07. 19
Data Mining & Quality Analytics Lab.
목충협

Model-agnostic meta-learning for fast ad.

발표자:  목충협

 2019년 7월 19일

 오후 1시 ~

 고려대학교 신공학관 218호

세미나 정보 보기 →



Evaluations

A closer look at few-shot classification

❖ Models

- Baseline
- Baseline++

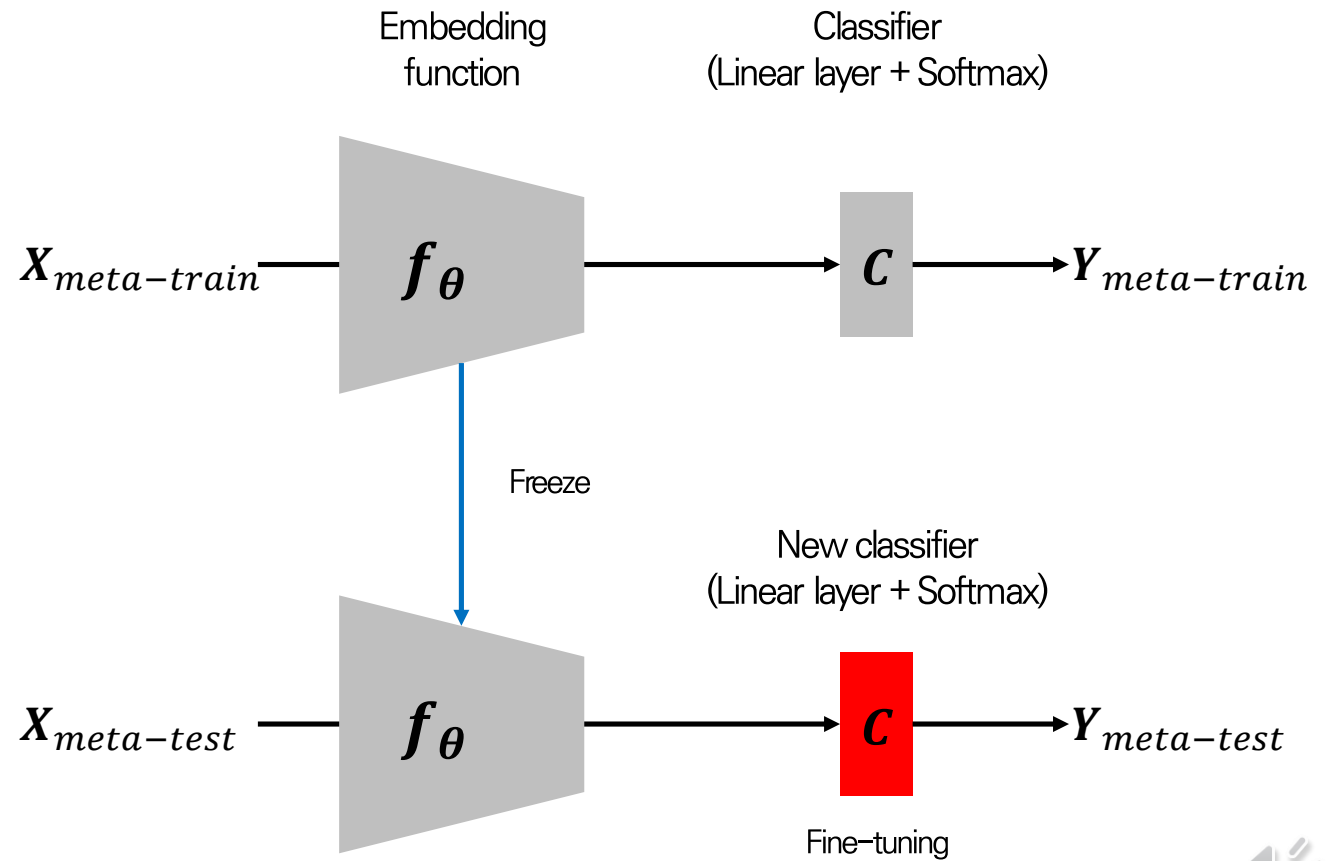
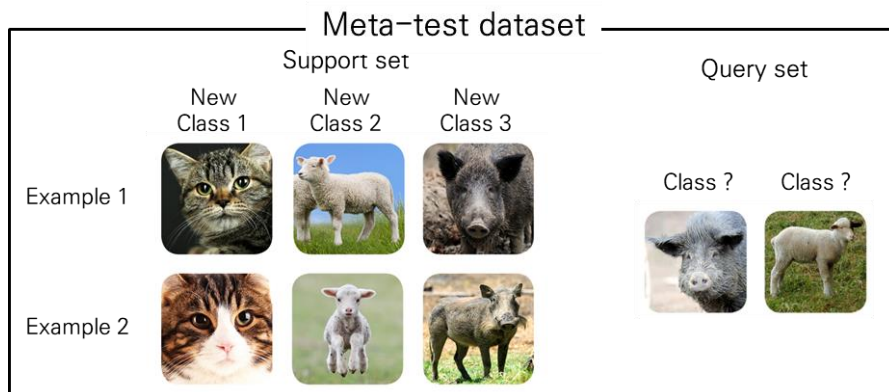
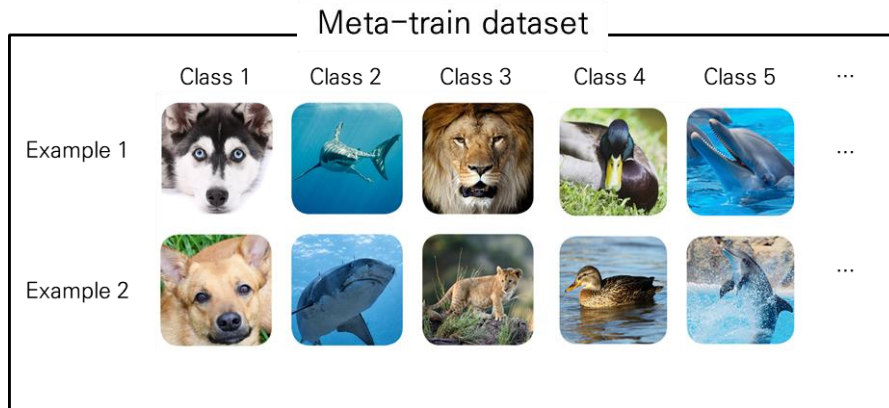


Evaluations

A closer look at few-shot classification

❖ Baseline

- Simple fine-tuning model

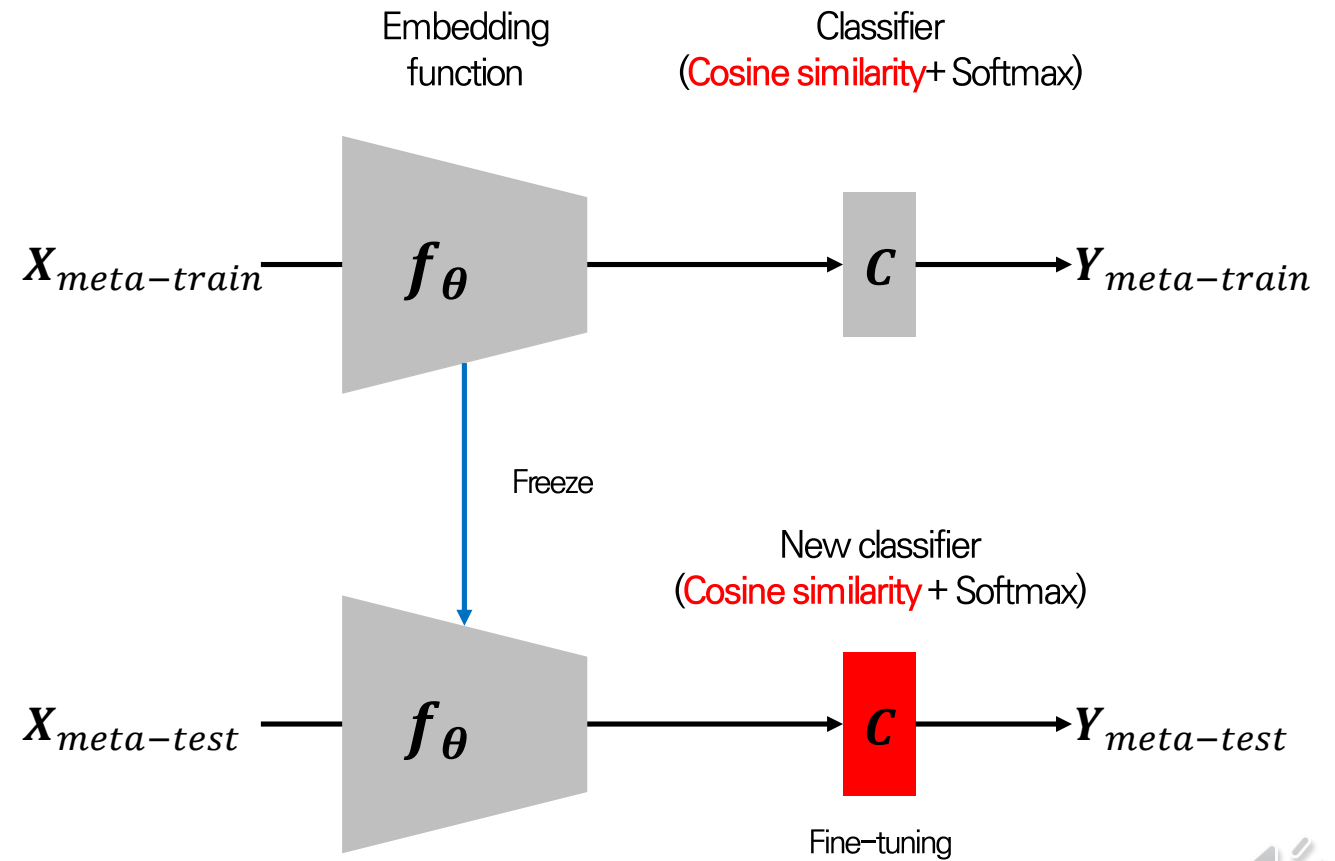
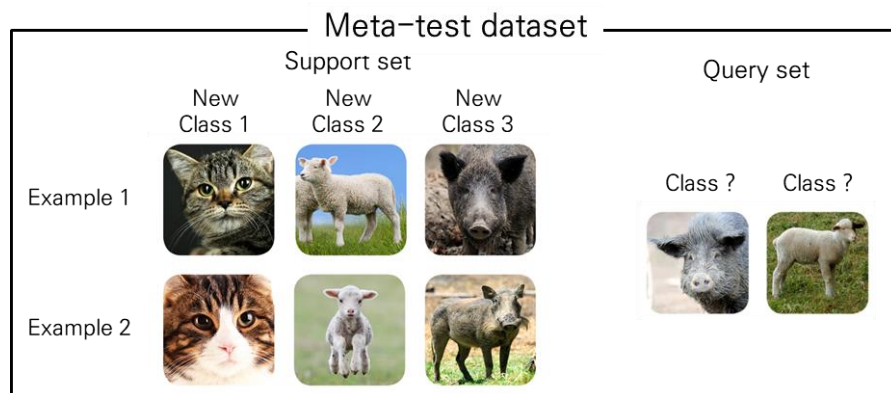
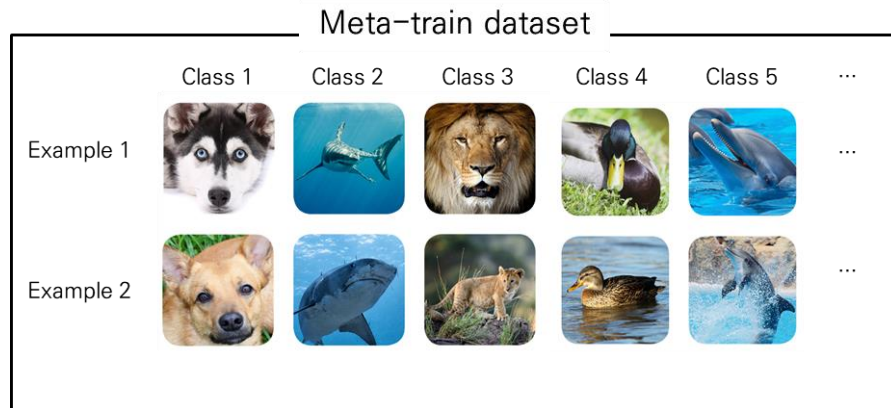


Evaluations

A closer look at few-shot classification

❖ Baseline++

- Simple fine-tuning model



Evaluations

A closer look at few-shot classification

❖ Datasets

- mini-ImageNet
 - 100 Classes (64 base, 16 validation, 20 novel)
 - 각 클래스별 600장
 - 다양한 class로 이루어져 있음
- CUB
 - 200 Classes (100 base, 50 validation, 50 novel)
 - 각 클래스별 60장
 - 새(조류)에 관한 class로 이루어져 있음



Evaluations

A closer look at few-shot classification

❖ Re-implementation

- mini-ImageNet dataset, Conv-4 사용
- * → augmentation 미적용

Method	1-shot		5-shot	
	Reported	Ours	Reported	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



Evaluations

A closer look at few-shot classification

❖ Baseline++

- Baseline / Baseline++ 의 차이는 마지막 레이어가 linear layer / cosine similarity 인지의 차이
- Baseline++ 는 논문에서 제시한 방법이며, 성능이 좋은 이유는 Intra-class variation을 줄이기 때문

Method	CUB		<i>mini-ImageNet</i>	
	1-shot	5-shot	1-shot	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

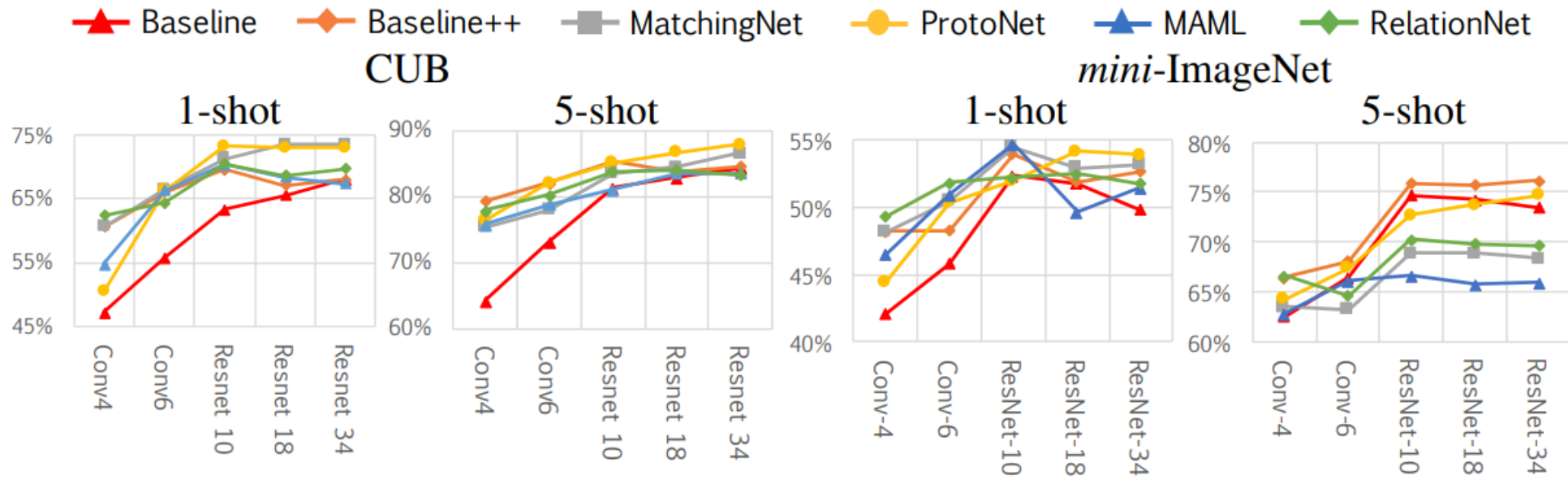


Evaluations

A closer look at few-shot classification

❖ Backbone depth

- Backbone의 깊이가 깊어질수록, 성능 차이가 줄어듦
- 데이터셋 (CUB/mini-ImageNet dataset)에 따라 메타 러닝의 효과가 차이남

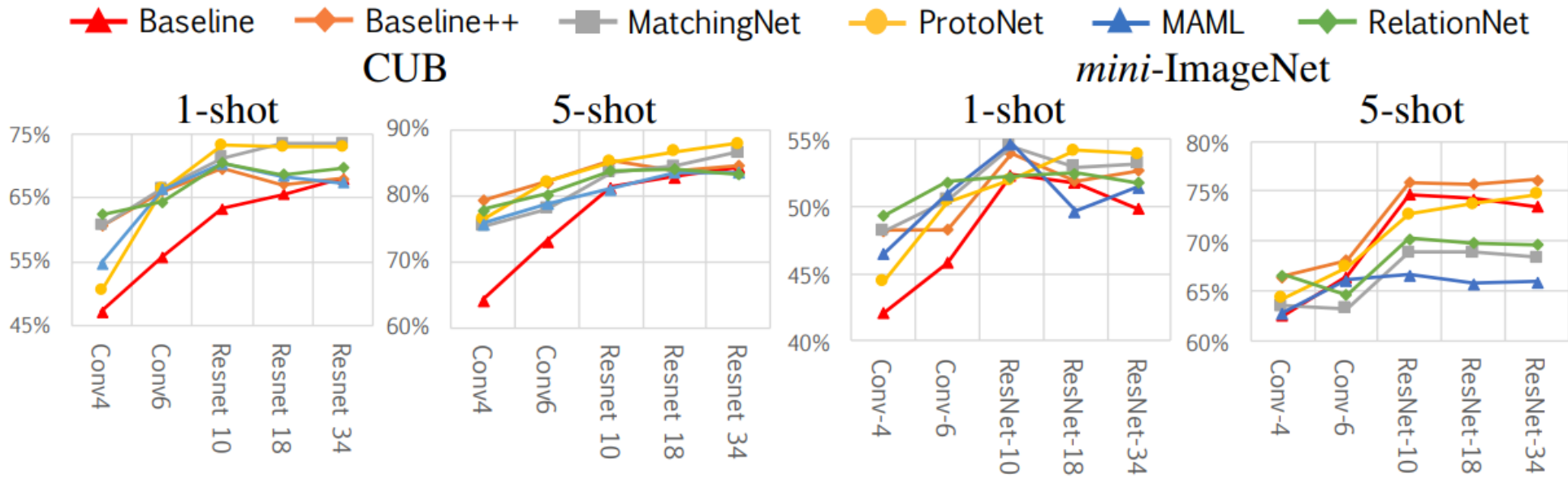


Evaluations

A closer look at few-shot classification

❖ Backbone depth

- Backbone의 깊이가 깊어질수록, 성능 차이가 줄어듦 → Intra-class variance
- 데이터셋 (CUB/mini-ImageNet dataset)에 따라 메타 러닝의 효과가 차이남 → Domain difference

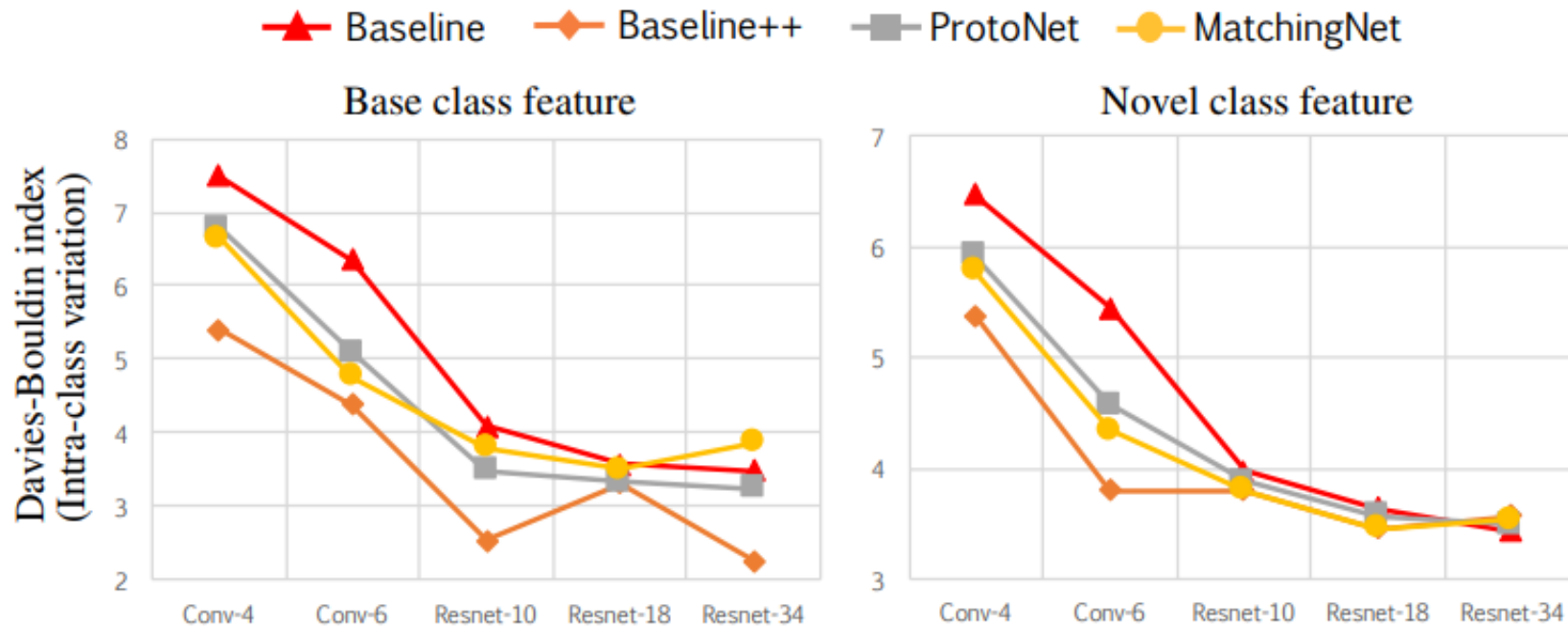


Evaluations

A closer look at few-shot classification

❖ Intra-class variance

- Backbone의 깊이가 깊어질수록, Intra-class variance가 줄어듦 → Class별로 잘 모여있음

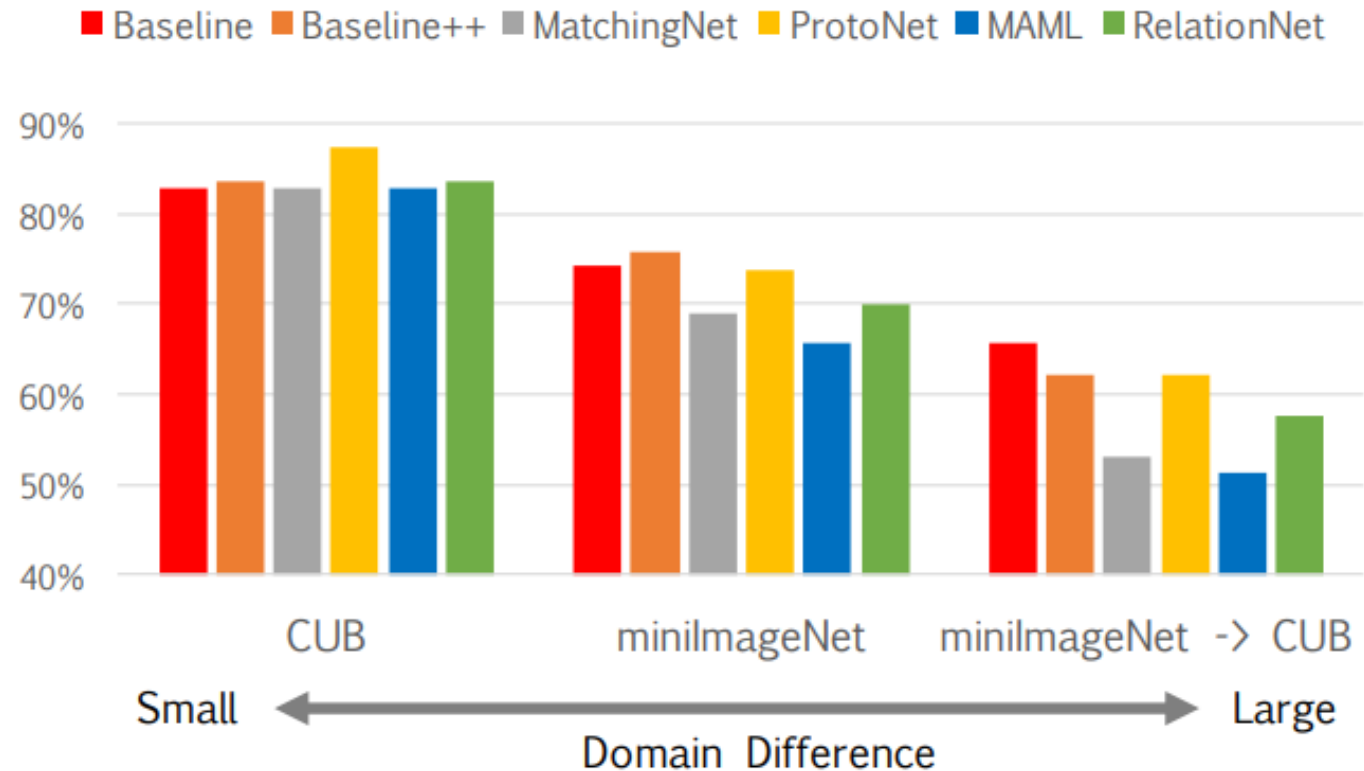


Evaluations

A closer look at few-shot classification

❖ Domain difference

- 도메인이 바뀌면, 기존의 데이터들은 학습에 큰 도움이 되지 않음

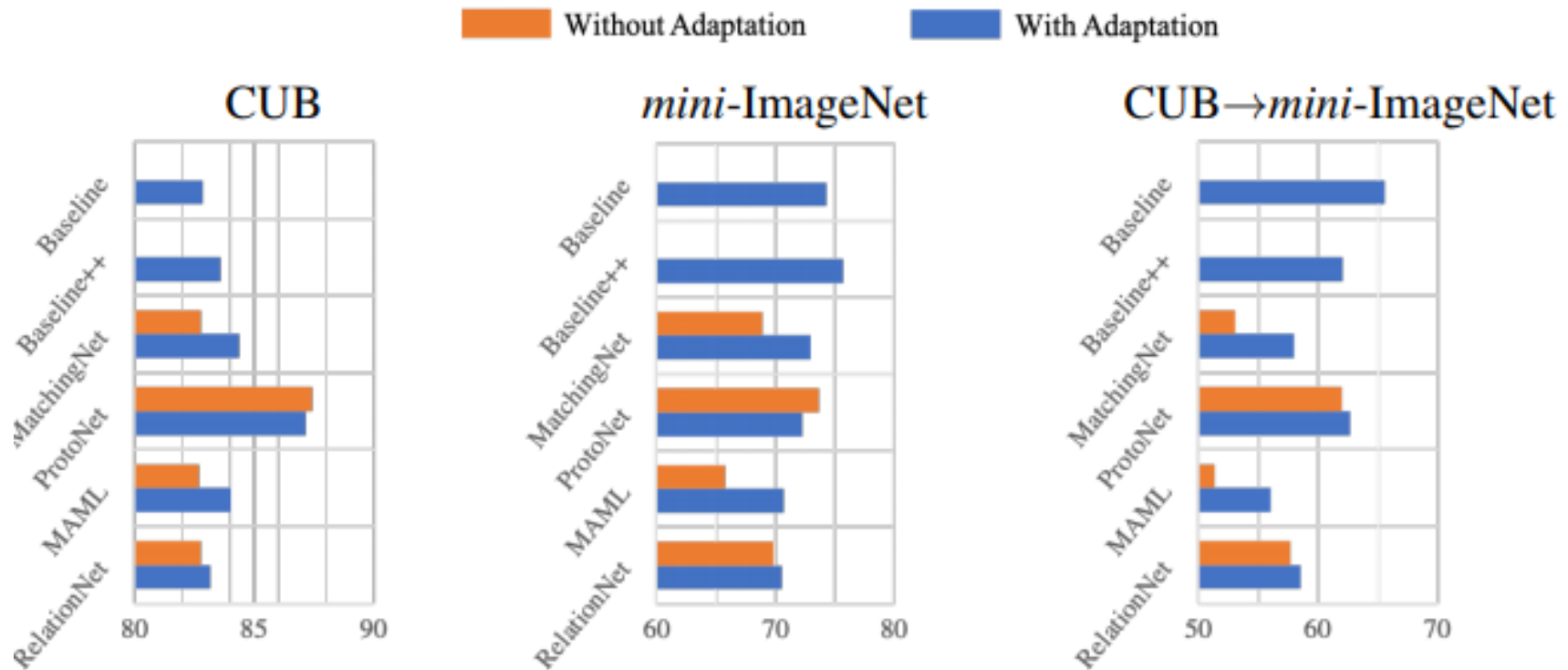


Evaluations

A closer look at few-shot classification

❖ Domain adaptation

- 마지막 layer만 fine-tuning 진행



Conclusion

❖ Metric-based approaches to meta-learning의 장점

- 단순하고 직관적이어서 이해하기 쉬움
- 임베딩을 이용하기 때문에 시각화하기 편하고, 어떤 일이 일어나는지 해석하기 쉬움
- Metric이나 Embedding space를 변경할 수 있기 때문에 도메인 지식이나 혹은 다른 정보들을 적용하기 쉬움



Conclusion

❖ Metric-based approaches to meta-learning의 단점

- Adaptation이 어렵기 때문에, Domain 변경에 취약
- Metric/Embedding이 성능에 많은 영향을 주는데, 결정하기가 쉽지 않음
- Distance를 계산하기 때문에, 데이터가 증가할수록 계산량이 많음
- Classification 외의 Task에는 적용하기 어려움



Conclusion

❖ Comments

- Few-shot learning을 넘어 Many-shot learning...
- 목적이 같은 Self-supervised / Semi-supervised 등 여러 방법론들과 공유하는 아이디어들이 많음
- 데이터의 양이 적기 때문에 Uncertainty를 잘 이용할 수 있어야함
- 딥러닝 아이디어들에 대한 해석이 매우 다양함



Thank you