TADAM :
Task dependent adaptive metric for improved few-shot learning

2019.03.17

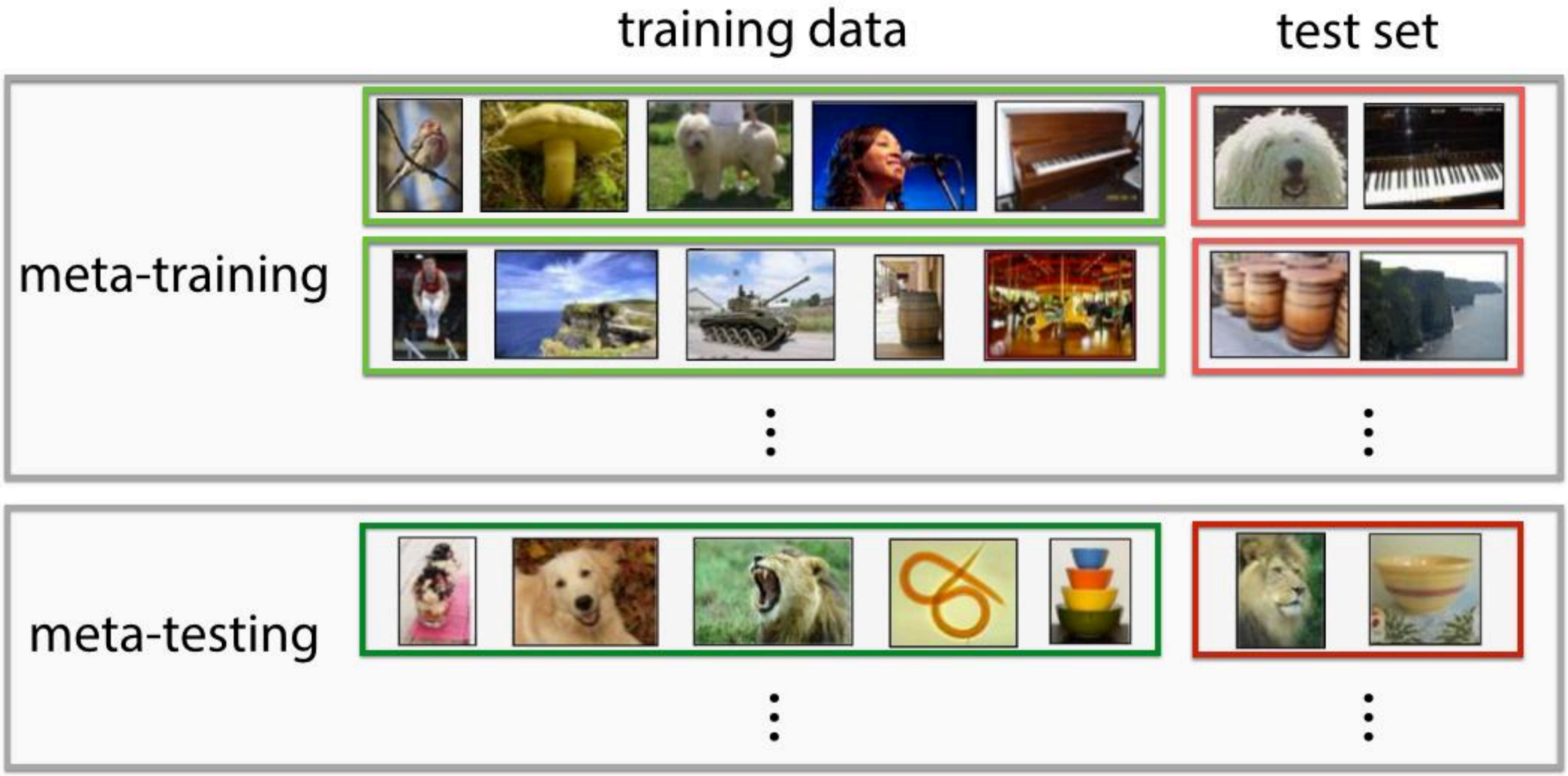
발표자 : 김용규

DAVIAN
Data and Visual Analytics Lab

KOREA UNIVERSITY

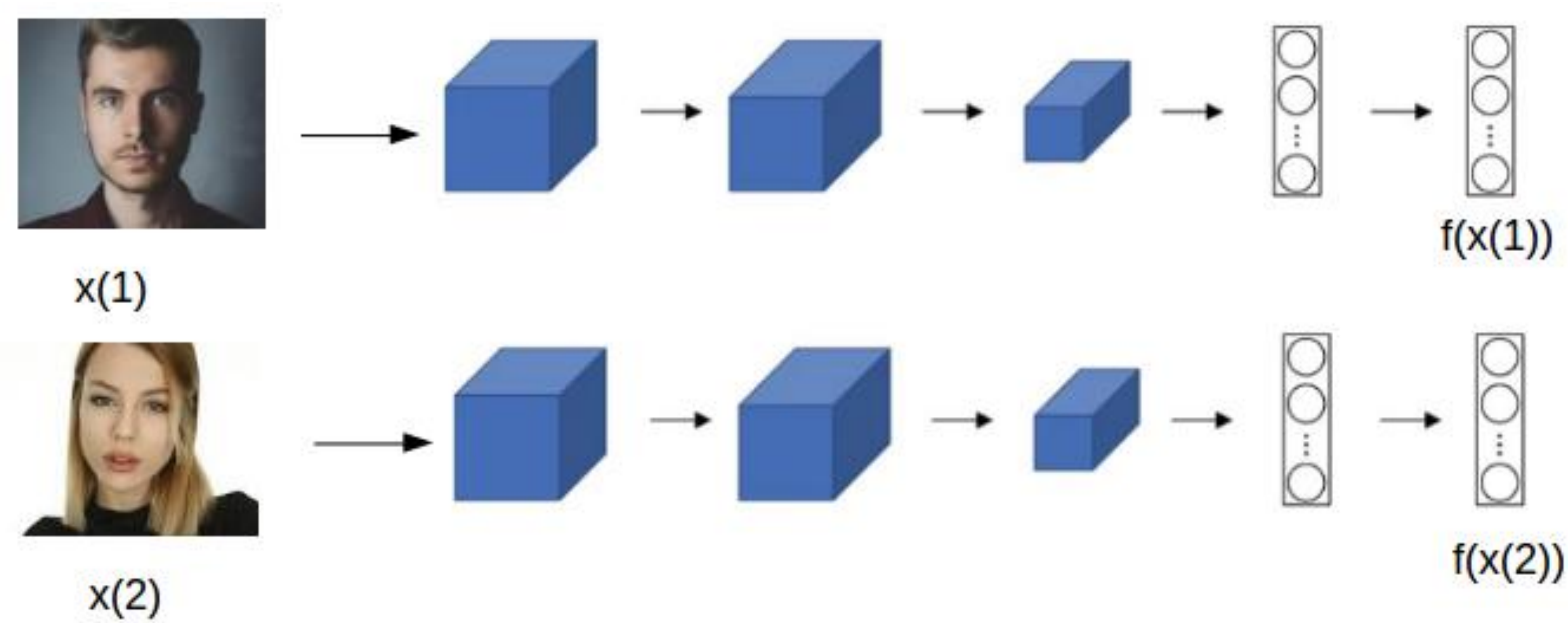
1 Introduction

Few-shot learning



1 Introduction

Metric Learning



2 Abstract

Metric Learning

- Metric scaling

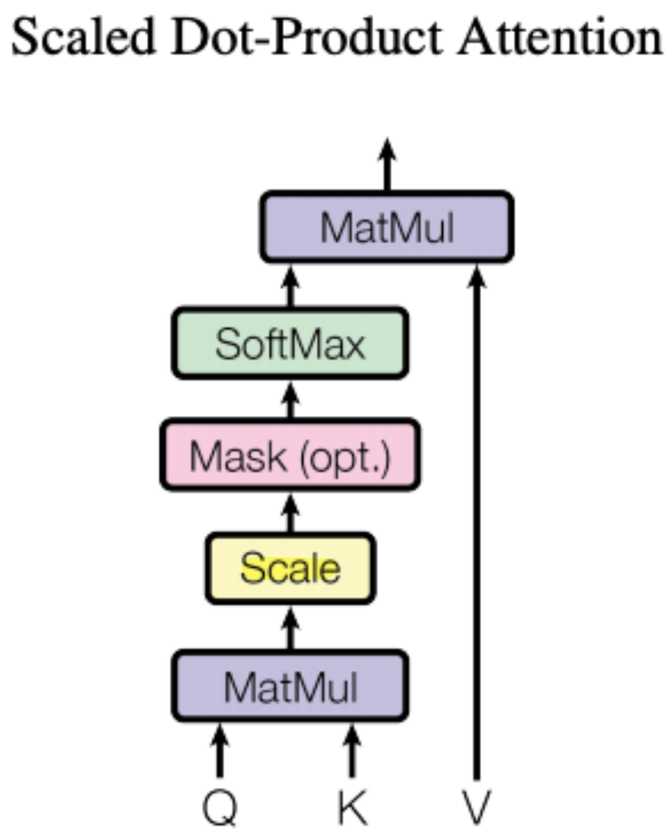
- Metric task conditioning

DAVIAN
Data and Visual Analytics Lab

KOREA UNIVERSITY

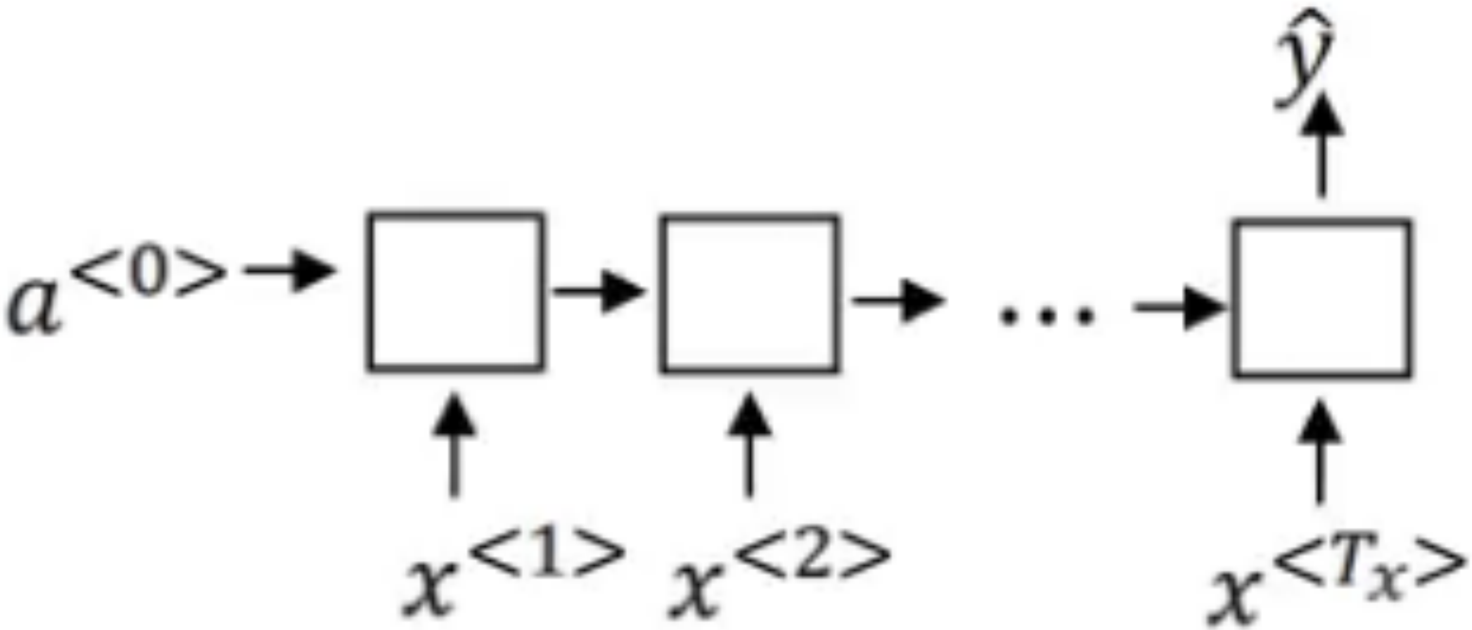
2 Abstract

Metric scaling



2 Abstract

Metric task conditioning



3 Contribution

Task-dependent metric space

- Learner를 task sample set에 한정해서 task-dependent metric space를 학습

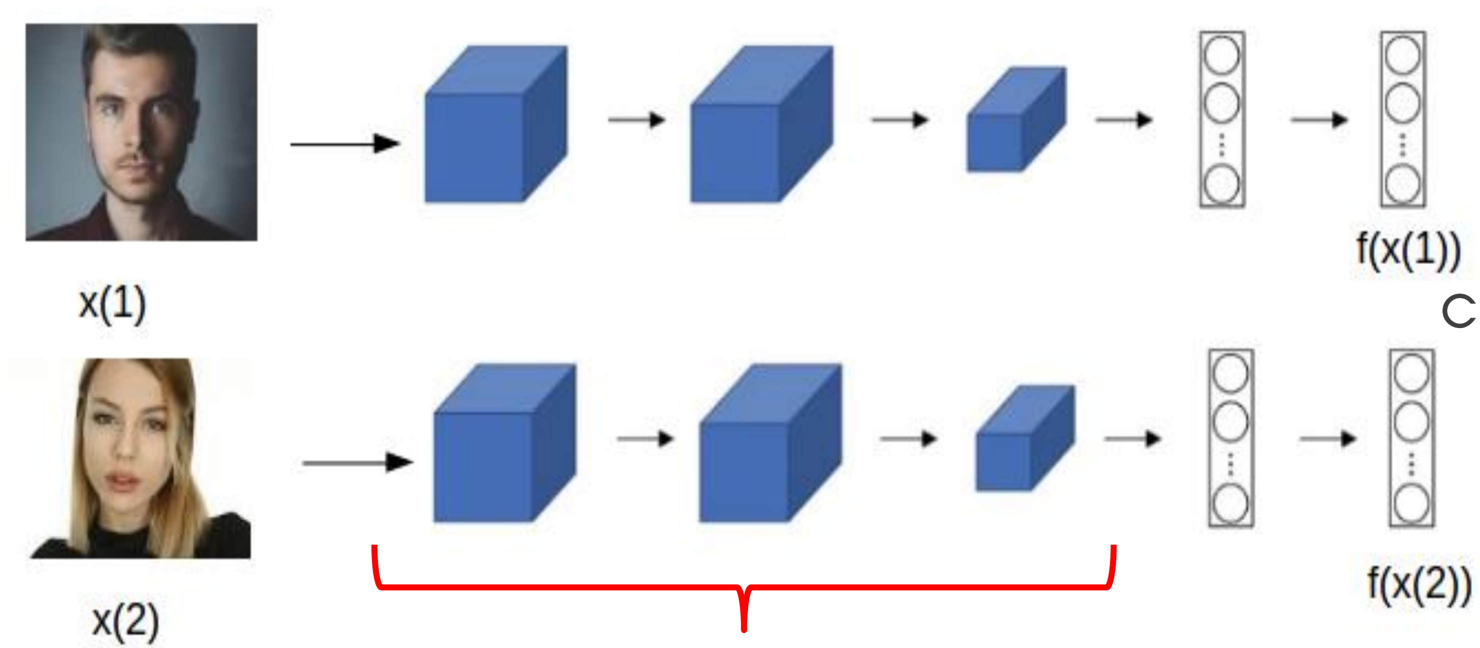
- Task-dependent metric space를 학습시키기 위해서 auxiliary task co-training을 기반으로 한 end-to-end optimization을 제안

3 Contribution

Task-dependent metric space

interaction

Similarity measure



Cost function

Feature extractor

+

Auxiliary task co-training

4 Background

Task-dependent metric space

$$\mathcal{S} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{MK}$$

$$\mathcal{Q} = \{(\mathbf{x}_i, y_i)\}_{i=1}^{q}$$

$$\mathbf{x}_i \in \mathbb{R}^{D_\mathbf{x}}$$

$$f_\phi : \mathbb{R}^{D_\mathbf{x}} \to \mathbb{R}^{D_\mathbf{z}}$$

$$\mathcal{S}_k\text{:}\ \mathbf{c}_k = \frac{1}{K} \sum_{\mathbf{x}_i \in \mathcal{S}_k} f_\phi(\mathbf{x}_i)$$

$$-\log p_\phi(y = k|\mathbf{x})$$

$$p_\phi(y = k|\mathbf{x}) = \text{softmax}(-d(f_\phi(\mathbf{x}), \mathbf{c}_k))$$

5

Model

Class-wise cross-entropy loss

$$J_k(\phi, \alpha) = \sum_{\mathbf{x}_i \in \mathcal{Q}_k} \left[\alpha d(f_\phi(\mathbf{x}_i), \mathbf{c}_k) + \log \sum_j \exp(-\alpha d(f_\phi(\mathbf{x}_i), \mathbf{c}_j)) \right]$$

$$\frac{\partial}{\partial \phi} J_k(\phi, \alpha) = \alpha \sum_{\mathbf{x}_i \in \mathcal{Q}_k} \left[\frac{\partial}{\partial \phi} d(f_\phi(\mathbf{x}_i), \mathbf{c}_k) - \frac{\sum_j \exp(-\alpha d(f_\phi(\mathbf{x}_i), \mathbf{c}_j)) \frac{\partial}{\partial \phi} d(f_\phi(\mathbf{x}_i), \mathbf{c}_j)}{\sum_j \exp(-\alpha d(f_\phi(\mathbf{x}_i), \mathbf{c}_j))} \right]$$

5 Model

Class-wise cross-entropy loss

Lemma 1 (Metric scaling). *If the following assumptions hold:*

$$\mathcal{A}_1 : d(f_\phi(\mathbf{x}), \mathbf{c}_k) \neq d(f_\phi(\mathbf{x}'), \mathbf{c}_k), \forall k, \mathbf{x} \neq \mathbf{x}' \in \mathcal{Q}_k; \quad \mathcal{A}_2 : \left| \tfrac{\partial}{\partial \phi} d(f_\phi(\mathbf{x}), \mathbf{c}) \right| < \infty, \forall \mathbf{x}, \mathbf{c}, \phi,$$

then it is true that:

$$\lim_{\alpha \to 0} \frac{1}{\alpha} \frac{\partial}{\partial \phi} J_k(\phi, \alpha) = \sum_{\mathbf{x}_i \in \mathcal{Q}_k} \left[\frac{K-1}{K} \frac{\partial}{\partial \phi} d(f_\phi(\mathbf{x}_i), \mathbf{c}_k) - \frac{1}{K} \sum_{j \neq k} \frac{\partial}{\partial \phi} d(f_\phi(\mathbf{x}_i), \mathbf{c}_j) \right],$$

$$\lim_{\alpha \to \infty} \frac{1}{\alpha} \frac{\partial}{\partial \phi} J_k(\phi, \alpha) = \sum_{\mathbf{x}_i \in \mathcal{Q}_k} \left[\frac{\partial}{\partial \phi} d(f_\phi(\mathbf{x}_i), \mathbf{c}_k) - \frac{\partial}{\partial \phi} d(f_\phi(\mathbf{x}_i), \mathbf{c}_{j_i^*}) \right];$$

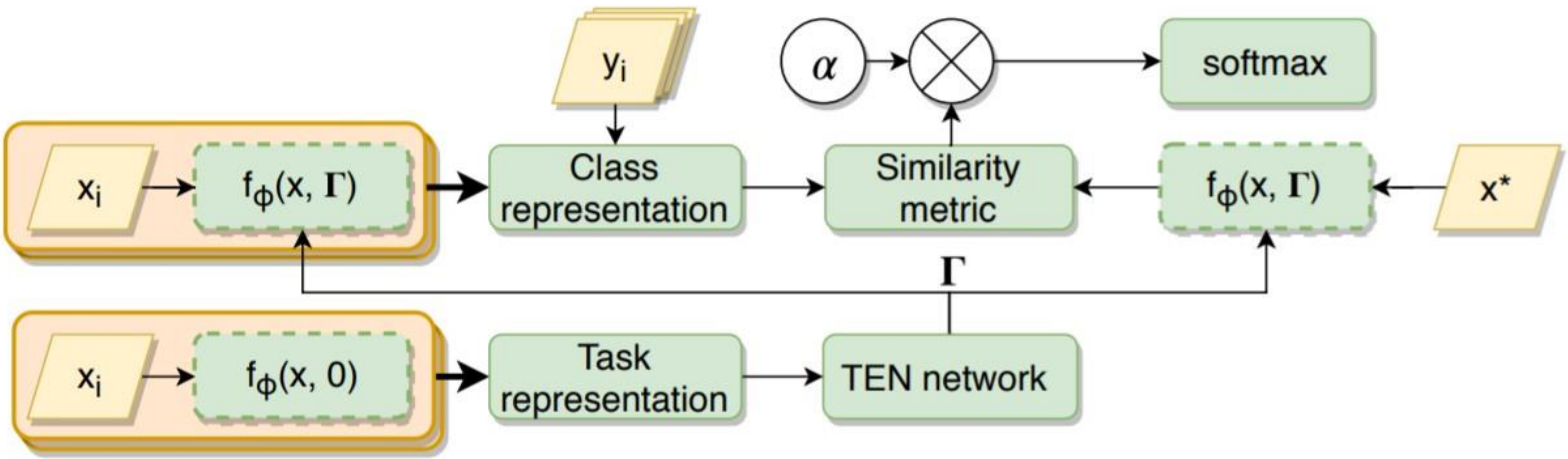
where $j_i^* = \arg\min_j d(f_\phi(\mathbf{x}_i), \mathbf{c}_j)$.

Proof. Please refer to Appendix A.

DAVIAN
Data and Visual Analytics Lab
KOREA UNIVERSITY

Model

Task representation



DAVIAN Data and Visual Analytics Lab

KOREA UNIVERSITY

5 Model

Task conditioning

Table 1: mini-Imagenet (Vinyals et al. [33]), 5-way classification results. †Our re-implementation.

	1-shot	5-shot	10-shot
Meta Nets [22]	43.4	60.6	-
Matching Networks [33]	46.6	60.0	-
MAML [6]	48.7	63.1	-
Proto Nets [28]	49.4	68.2	74.3†
Relation Net [29]	50.4	65.3	-
SNAIL [16]	55.7	68.9	-
Discriminative k-shot [1]	56.3	73.9	78.5
adaResNet [17]	56.9	71.9	-
Ours	**58.5**	**76.7**	**80.8**