Distribution Consistency based Covariance Networks for Few-shot Learning

Wenbin Li[1*], Jinglin Xu[2*], Jing Huo[1], Lei Wang[3], Yang Gao[1], Jiebo Luo[4]

[1]National Key Laboratory for Novel Software Technology, Nanjing University, China
[2]Northwestern Polytechnical University, China
[3]University of Wollongong, Australia
[4]University of Rochester, USA

Wenbin Li

January 31, 2019

Outline

- **Introduction**
  - Few-shot learning
- **Covariance Metric Network**
  - Motivation
  - Model architecture
  - Local covariance representation
  - Covariance metric function
- **Experiments**
  - Generic few-shot classification
  - Fine-grained few-shot classification
- **Conclusion**

Introduction

- **One or Few-Shot Learning**

One-shot learning is an object categorization problem in computer vision. Whereas most machine learning based object categorization algorithms require training on hundreds or thousands of images and very large datasets, one-shot learning aims to learn information about object categories from one, or only a few, training images.

(https://www.wikipedia.org/)

Introduction

Few-Shot Learning

- **Naive method**
  Directly learn a classifier only from the few training samples.

- **Generation based methods**
  Generate new samples, like data augmentation (e.g., GANs).

- **Transfer-learning based methods**
  Learn transferable knowledge from an auxiliary dataset.

Introduction

Few-Shot Learning

Three kinds of datasets:

- A **support** set (few-shot training set)
- A **query** set (testing set)
- An **auxiliary** set (additional set)

  It has it own label space that is disjoint with support/query set.

If the support set contains ***K*** labelled samples for each of ***C*** categories, the target few-shot task is called a ***C*-way *K*-shot** task.

Outline

- **Introduction**
  - Few-shot learning
- **Covariance Metric Network**
  - Motivation
  - Model architecture
  - Local covariance representation
  - Covariance metric function
- **Experiments**
  - Generic few-shot classification
  - Fine-grained few-shot classification
- **Conclusion**

Covariance Metric Network

Problem Statements

Three key aspects in few-shot Learning:

- **Transferable Knowledge**
  How to learn and store the transferable knowledge by fully utilizing the auxiliary dataset?
- **Concept Representation**
  How to represent a concept precisely in the few-shot setting?
- **Relation Measure**
  How to reasonably measure the relationship between a concept and a query sample?

Covariance Metric Network

Motivation

Conventional Methods:

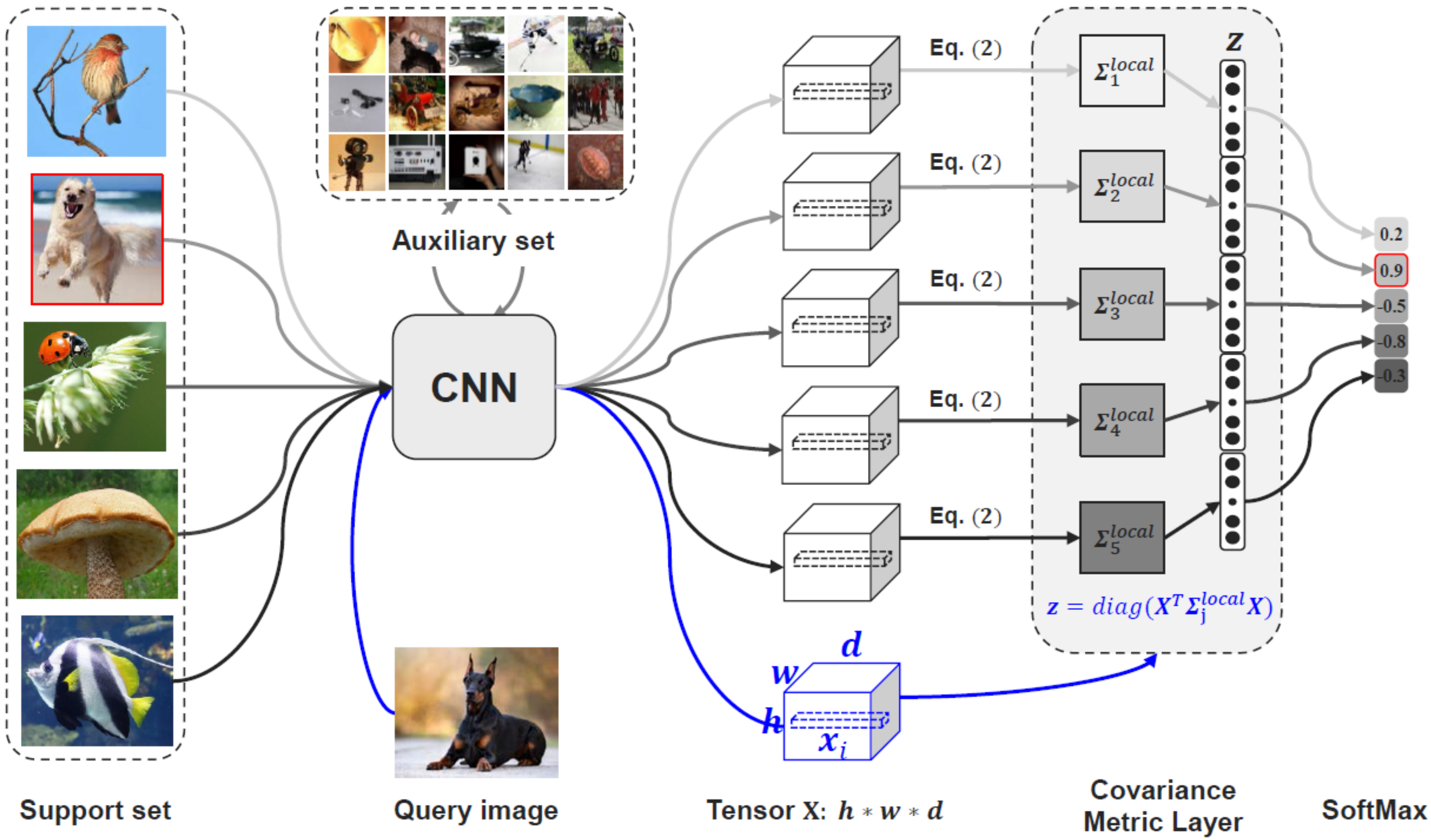
- Use **global feature** to represent an image.
- Only focus on the **first-order statistic** to represent a concept.
- Use a **fixed metric** function (e.g., Euclidean distance).

The Proposed Method:

- Use richer **local descriptors** to represent an image.
- Also employ the **second-order statistic** to represent a concept.
- Use a **learnable deep metric** based on distribution consistency.

Covariance Metric Network

- **Model Architecture (CovaMNet)**



Covariance Metric Network

Solutions

- **Transferable Knowledge**
  - Employ the episodic training mechanism.
- **Concept Representation**
  - Propose a novel local covariance representation.
- **Relation Measure**
  - Define a new covariance metric function.

Covariance Metric Network

Episodic Training Mechanism

Philosophy:

Testing conditions must match the training conditions.

Episodic training:

Exploiting the auxiliary set to mimic the few-shot learning setting via episode-based training.

One episode: a support set + a query set.

Covariance Metric Network

Local Covariance Representation

Given an image set of the c-th category $\boldsymbol{D}_c = \{\boldsymbol{X}_1, \ldots, \boldsymbol{X}_K\}, \boldsymbol{X}_i|_{i=1}^{K} \in R^{d*M}$ (d is the local descriptor dimensionality), which contains *K* images with *M* local deep local descriptors per image, the local covariance metric can be defined as follows,

$$\boldsymbol{\Sigma}_c^{local} = \frac{1}{MK-1}\sum_{i=1}^{K}(\boldsymbol{X}_i - \boldsymbol{\tau})(\boldsymbol{X}_i - \boldsymbol{\tau})^\top$$

For example:
For a 5-way 5-shot task, there are $M = 400$ deep local descriptors for each image $\boldsymbol{X}_i$. It means that we have $MK = 400 * 5$ samples for one category in total. Then we use all these 2000 samples to calculate a covariance matrix as the representation.

Covariance Metric Network

Local Covariance Representation

Advantages:

- **Using local descriptors**
  - Data augmentation (***VS.*** Few-shot)
  - Capture the local details (***VS.*** Global feature)
- **Using covariance matrix**
  - Capture the second-order information (***VS.*** First-order)
  - Describe the underlying concept distribution (***VS.*** Non-distribution)

Covariance Metric Network

Covariance Metric Function

Measure the ***distribution consistency*** between a sample and a category:

$$d(\boldsymbol{x}, \boldsymbol{\Sigma}) = \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}$$

Describes the underlying distribution of one concept

Covariance Metric Network

- **Covariance Metric Function**

Compared with other metric functions:

Mahalanobis distance: $d(\boldsymbol{x}, \boldsymbol{y}) = \sqrt{(\boldsymbol{x} - \boldsymbol{y})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{x} - \boldsymbol{y})}$

Bilinear similarity: $d(\boldsymbol{x}, \boldsymbol{y}) = \boldsymbol{x}^T \boldsymbol{\Sigma} \boldsymbol{y}$

Covariance metric: $d(\boldsymbol{x}, \boldsymbol{\Sigma}) = \boldsymbol{x}^\top \boldsymbol{\Sigma} \boldsymbol{x}$

Covariance Metric Network

- **Theoretical Analysis**

Theorem 1. *Suppose that* $\mathbf{\Sigma} \in \mathbb{R}^{d \times d}$ *is the covariance matrix of one specific category from the support set* $\boldsymbol{S}$*, satisfying* $\mathbf{\Sigma} = \boldsymbol{V}\mathbf{\Lambda}\boldsymbol{V}^\top$*, where the diagonal matrix* $\mathbf{\Lambda} \in \mathbb{R}^{d \times d}$ *consists of* d *eigenvalues in descending order and the corresponding eigenvectors are denoted as the orthogonal matrix* $\boldsymbol{V} = [\boldsymbol{v}_1, \cdots, \boldsymbol{v}_d] \in \mathbb{R}^{d \times d}$*. For any nonzero sample* $\boldsymbol{x} \in \mathbb{R}^d$*,* $\boldsymbol{x}^\top \mathbf{\Sigma} \boldsymbol{x}$ *will achieve a maximum based on the first* k *eigenvalues if* $\boldsymbol{x}$ *is in the direction of the first* k *eigenvectors of* $\mathbf{\Sigma}$*.*

Covariance Metric Network

Covariance Metric Function

$$d(\boldsymbol{x}, \boldsymbol{\Sigma}) = \boldsymbol{x}^{\top} \boldsymbol{\Sigma} \boldsymbol{x}$$

Advantages:

- Measure distribution consistency (***VS.*** distance between samples)
- Avoid calculating the inverse matrix (***VS.*** Mahalanobis distance)

Outline

- **Introduction**
  - Few-shot learning
- **Covariance Metric Network**
  - Motivation
  - Model architecture
  - Local covariance representation
  - Covariance metric function
- **Experiments**
  - Generic few-shot classification
  - Fine-grained few-shot classification
- **Conclusion**

Experiments

Experimental Setups

Datasets:

- **miniImageNet**
- **StanfordDog**
- **StanfordCar**
- **Cub-200**

Baselines:

- **Meta-learner** *(ICLR’17)*
- **MAML** *(ICML’17)*
- **SNAIL** *(ICLR’18)*
- **Matching Net** *(NIPS’16)*
- **GNN** *(ICLR’18)*
- **Prototypical Net** *(NIPS’17)*
- **Relation Net** *(CVPR’18)*

Embedding module:

- **Four convolutional blocks**

Convolutional block
3*3 conv, 64 filters
batch norm
Leaky ReLU

Tasks (5-way 1-shot & 5-way 5-shot):

- **Generic few-shot classification**
- **Fine-grained few-shot classification**

Experiments

- **Generic Few-shot Classification**

Model	Embed.	Type	Fine Tune	5-Way Accuracy (%)	
				1-shot	5-shot
Baseline k-NN	64F	Metric	N	27.23±1.41	49.29±1.56
Meta-Learner* (Ravi and Larochelle 2017)	32F	Meta	N	43.44±0.77	60.60±0.71
MAML* (Finn, Abbeel, and Levine 2017)	32F	Meta	Y	48.70±1.84	63.11±0.92
SNAIL* (Mishra et al. 2018)	32F	Meta	N	45.10±0.00	55.20±0.00
Matching Nets FCE* (Vinyals et al. 2016)	64F	Metric & Meta	N	43.56±0.84	55.31±0.73
GNN (Garcia and Bruna 2018)	64F	Metric	N	49.02±0.98	63.50±0.84
Prototypical Nets* (Snell, Swersky, and Zemel 2017)	64F	Metric	N	‡49.42±0.78	‡68.20±0.66
Relation Net* (Yang et al. 2018)	64F	Metric	N	50.44±0.82	65.32±0.70
Our CovaMNet	64F	Metric	N	**51.19**±0.76	**67.65**±0.63

Experiments

Fine-grained Few-shot Classification

Model	Embed.	5-Way Accuracy (%)					
		Stanford Dogs		*Stanford Cars*		*CUB Birds*	
		1-shot	5-shot	1-shot	5-shot	1-shot	5-shot
Baseline k-NN	64F	26.14±0.91	43.14±1.02	23.50±0.88	34.45±0.98	25.81±0.90	45.34±1.03
Matching Nets FCE	64F	35.80±0.99	47.50±1.03	34.80±0.98	44.70±1.03	45.30±1.03	59.50±1.01
Prototypical Nets	64F	37.59±1.00	48.19±1.03	40.90±1.01	52.93±1.03	37.36±1.00	45.28±1.03
GNN	64F	46.98±0.98	62.27±0.95	55.85±0.97	71.25±0.89	51.83±0.98	63.69±0.94
Our CovaMNet	64F	**49.10±0.76**	**63.04±0.65**	**56.65±0.86**	**71.33±0.62**	**52.42±0.76**	**63.76±0.64**

Outline

- **Introduction**
  - Few-shot learning
- **Covariance Metric Network**
  - Motivation
  - Model architecture
  - Local covariance representation
  - Covariance metric function
- **Experiments**
  - Generic few-shot classification
  - Fine-grained few-shot classification
- **Conclusion**

Conclusion

Problems:

- How to learn transferable knowledge form the auxiliary data?
- How to represent a concept precisely in the few-shot setting?
- How to measure the relationship between a concept and a query sample?

Model: An end-to-end Covariance Metric Network (**CovaMNet**)

- Employ the episodic training mechanism.
- Design a novel local covariance representation.
- Construct a new covariance metric function.

Conclusion

- **Code**

https://github.com/WenbinLee/CovaMNet

Acknowledgement

Jiebo Luo	VIStA Group
Lei Wang	VILA Group
Yang Gao	R&L Group
Jing Huo	R&L Group
Jinglin Xu	NWPU

Thank you

Q & A