IMT Atlantique
Bretagne-Pays de la Loire
École Mines-Télécom

La recherche CARNOT pour l'innovation des entreprises

Improving Classification Accuracy with Graph Filtering

Mounia Hamidouche, Carlos Lassance, Yuqing Hu, Lucas Drumetz, Bastien Pasdeloup, Vincent Gripon

ICIP Conference, Anchorage, Alaska, USA

mounia.hamidouche@gmail.com

September 19-22, 2021

Outline

1. Deep Learning and Few-shot Learning
2. Methodology
3. Experiments
4. Conclusion and Perspectives

Outline

1. Deep Learning and Few-shot Learning
2. Methodology
3. Experiments
4. Conclusion and Perspectives

Introduction

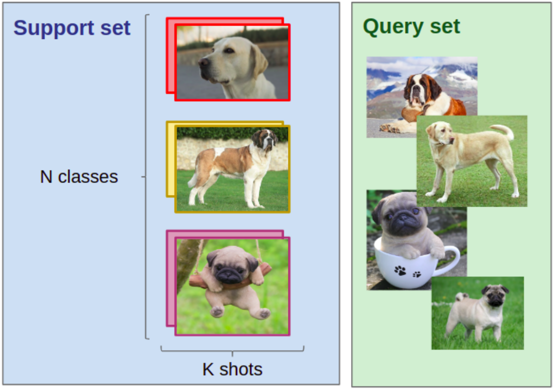
- Deep Learning is the state of the art for many machine learning problems

- Reaching peak performance typically requires large annotated datasets

- Performing predictions on batches of unlabeled inputs can boost performance: this setting is called *transductive*.

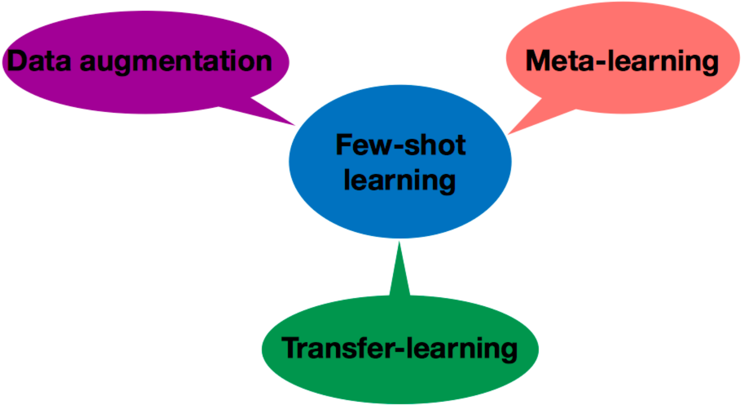
Transductive predictions are even more impactful in the case of few-shot learning

- N-way K-shot image classification



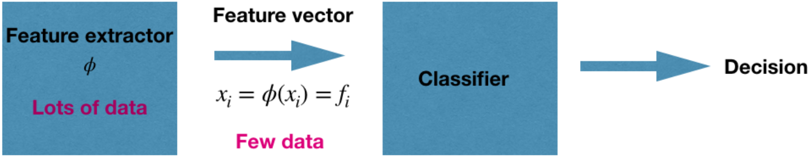
- For $K \leq 5$, classify the query images among the N classes given the $N \times K$ images in the support set

Few-shot Learning Methods



Main idea of the paper:

- Idea: Insert a denoising method after feature extraction and before classification after a model has been trained



- Goal: filter out the high frequencies within the feature representation of data to improve classification performance

- Tool: graph signal processing framework

Outline

1. Deep Learning and Few-shot Learning
2. Methodology
3. Experiments
4. Conclusion and Perspectives

Graph Signal Processing (GSP)

- GSP allows to manipulate signals defined on graph structures

  - Assume $\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$ is connected. If not, just process each component separately

    $\mathcal{V} = \{ 1, 2, ... n \}$: finite set of n nodes or vertices
    $\mathcal{E}$: set of edges
    $\mathbf{W} : \mathcal{E} \to \mathbb{R}$: map from the set of edges to scalar values

  - Assume undirected graph with fixed graph structure

  - Vertex signal $\mathbf{f}_i : \mathcal{V} \to \mathbb{R}$

  - Graph signal $\mathbf{f} = [\mathbf{f}_1, \mathbf{f}_2, ..., \mathbf{f}_n]^\top$

- Graph filters are systems that take a graph signal as an input and produce another signal indexed by the same graph as the output

Class Graphs

- Define a similarity matrix $\mathbf{W}$ between labeled samples

$$\mathbf{W}_{i,j} = w\left(\mathbf{f}_i, \mathbf{f}_j\right) \text{ if } i \neq j \text{ and } 0 \text{ otherwise,}$$

where s is a similarity measure on tensors.

- This matrix defines a graph $\mathcal{G}$

$$\mathcal{G} = (\mathcal{V}, \mathcal{E}, \mathbf{W})$$

with $\mathcal{V} = \{1, ..., |\mathbf{f}|\}$.

- Generate the adjacency matrix $\mathbf{W}$ of $\mathcal{G}$ for each class using k-nearest neighbors

$$\mathbf{W}_{i,j} = \begin{cases} \mathbf{W}_{i,j} & \text{if } \mathbf{W}_{i,j} \text{ is among the } k \text{ largest entries of row,} \\ 0 & \text{otherwise.} \end{cases}$$

- Normalized Laplacian matrix of $\mathcal{G}$

$$\mathbf{L} = \mathbf{I} - \mathbf{D}^{-\frac{1}{2}}\mathbf{W}\mathbf{D}^{-\frac{1}{2}},$$

where $\mathbf{D}$ is the diagonal degree matrix.

Graph Filters

- Eigenvalues $\boldsymbol{\lambda}$ of $\mathbf{L}$ are interpreted as frequencies
- A filter is usually defined by its spectral response

$$h : \lambda \mapsto h(\lambda)$$

where $\lambda \in \boldsymbol{\lambda}$.

- $\mathbf{H} = \mathrm{diag}(h(\boldsymbol{\lambda}))$, the filtered signal $\mathbf{f}^{\mathrm{filter}}$ is defined as

$$\mathbf{f}^{\mathrm{filter}} = \mathbf{U}\mathbf{H}\mathbf{U}^{\top}\mathbf{f} \ .$$

where $\mathbf{U}$ are the eigenvectors of $\mathbf{L}$.

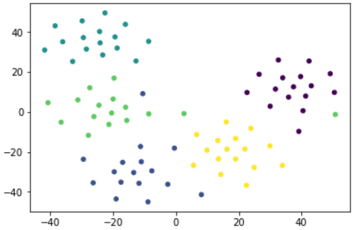
Observations

- To remove high frequencies in $\mathbf{f}$, a low-pass filter h nullifies large values of λ
- We substitute $\mathbf{f}$ with $\mathbf{f}^{\mathrm{filter}}$ as an input for the classifier

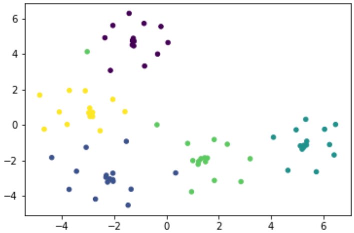
Effect of Low-pass Graph Filters on Centroids (1)

Low-pass filter: $h(\lambda) = 1$ if $\lambda \leq \lambda_m$ and $h(\lambda) = 0$ if not.

- Mini-imageNet data representation before applying any filter



- Mini-imageNet data representation after applying a graph filter



Effect of Low-pass Graph Filters on Centroids (2)

- $\mathbf{f}_i \sim \mathcal{N}(\boldsymbol{\mu}, \sigma^2\mathbf{I})$
- Centroids are computed as

$$\boldsymbol{\gamma} = \frac{1}{K}\sum_{i=1}^{K}\mathbf{f}_i \ .$$

- Filter $\mathbf{H}$ such that $h(\lambda) = 1$ if $\lambda \leq \boldsymbol{\lambda}_m$ and $h(\lambda) = 0$ otherwise ($\lambda \in \boldsymbol{\lambda}$)

Result

- We build a complete graph for each class, and let the number of labeled samples of that class $K \to \infty$, then

$$\mathbb{E}(\boldsymbol{\gamma}^{\text{filter}}) = \mathbb{E}(\boldsymbol{\gamma}) \ ,$$

$$\mathrm{Cov}(\boldsymbol{\gamma}^{\text{filter}}) = o\left(\mathrm{Cov}(\boldsymbol{\gamma})\right) \ .$$

Outline

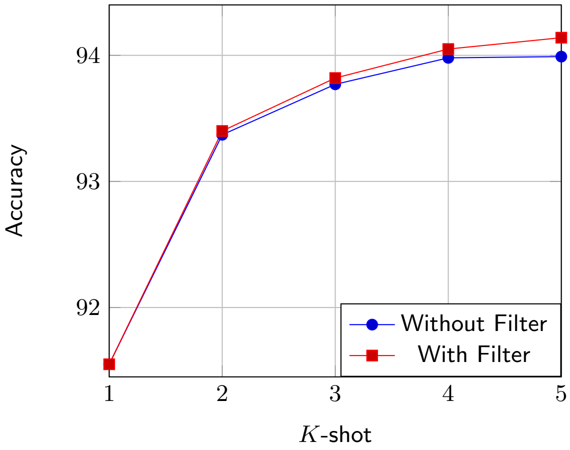
1. Deep Learning and Few-shot Learning
2. Methodology
3. Experiments
4. Conclusion and Perspectives

5-shot Accuracy of the Methodology Compared with State-of-the-art Method, for Various Extractors

- Runs $= 100.000$
- Number of classes $N = 5$
- Shots $K = 5$
- Queries $q = 15$

Classifier			
Dataset	Extractors	No Filter %	With Filter %
MINet	**WRN**	88.82 ± 0.013	$\mathbf{88.90 \pm 0.011}$
	DNet121	86.82 ± 0.014	$\mathbf{86.87 \pm 0.014}$
CUB	**WRN**	93.99 ± 0.011	$\mathbf{94.14 \pm 0.009}$
CIFAR	**WRN**	90.68 ± 0.015	$\mathbf{90.75 \pm 0.015}$
TINet	**DNet121**	90.44 ± 0.014	$\mathbf{90.62 \pm 0.013}$

Evolution of the Accuracy on CUB (extractor: WRN) as a Function of Shots K



Accuracy on the CIFAR-10 Dataset

	CIFAR-10		
Method	**WRN**	**ShakeNet**	**PyramidNet**
original paper	95.82 %	97.96 %	98.56 %
NCM	85.81 %	**97.97** %	98.52 %
1-NN	95.81 %	97.95 %	98.54 %
1-NN+Filter	**95.92** %	**97.97** %	**98.61** %

Outline

Conclusion and Perspectives

- A graph-based method is proposed to improve the accuracy of classification methods
- Use of GSP to reduce the noise in the extracted feature vectors
- The effectiveness of the method was shown theoretically
- Gains are obtained in two different settings: few-shot classification and standard classification
- Explore an automatic way of choosing the best filter parameters

Thanks for Your Attention.

mounia.hamidouche@gmail.com