Adaptive Anchor Label Propagation for Transductive Few-Shot Learning

Michalis Lazarou[1], Yannis Avrithis[2],
Guangyu Ren[1] and Tania Stathaki[1]

[1]Imperial College London, [2]IARAI

IEEE International Conference of Image Processing

Kuala Lumpur, Malaysia, October 2023

What is few-shot learning

- Why is it important?

?

Current approaches in few-shot learning

Relevant literature

- Meta-learning:
  - Optimization based.
  - Metric based.
  - Model based.
- Transfer learning.
- Feature adaptation.
- Data Augmentation.

Utilizing unlabeled data

- Data manifold exploitation:
  - Label propagation.
  - Embedding propagation.
- Iterative pseudo-label selection.
- Class centroid refinement:
  - Soft k-means.
  - Minimizing loss functions.

Motivation

Traditional Label Propagation algorithm

- We investigate the widely popular label propagation algorithm.
- We identify a limitation of label propagation, that the labeled data are fixed and may be in sub optimal positions.

Contributions

- We propose a novel variant of label propagation algorithm named Adaptive Anchor Label Propagation ($\mathbf{A^2LP}$).
-
- $\mathbf{A^2LP}$ outperforms significantly the traditional label propagation in transductive few-shot learning.

Motivation

Traditional Label Propagation algorithm

- We investigate the widely popular label propagation algorithm.
- We identify a limitation of label propagation, that the labeled data are fixed and may be in sub optimal positions.

Contributions

- We propose a novel variant of label propagation algorithm named Adaptive Anchor Label Propagation ($\mathbf{A^2LP}$).
-
- $\mathbf{A^2LP}$ outperforms significantly the traditional label propagation in transductive few-shot learning.

Problem formulation and definitions

Pre-training stage

- We use publicly available pre-trained networks from published works.
- Base class dataset: $D_{\text{base}} := \{(x_i, y_i)\}_{i=1}^{I}$ where $y_i \in C_{\text{base}}$.
- Network $f_\theta : \mathcal{X} \to \mathbb{R}^d$ is trained on D_{base}.

Inference stage

- Novel class dataset D_{novel} with C_{novel} disjoint from C_{base}.
- Assume access to f_θ, a support set, S, a query set, Q.
- We focus on transductive few-shot learning, where all S and Q are available at inference at the same time.

Problem formulation and definitions

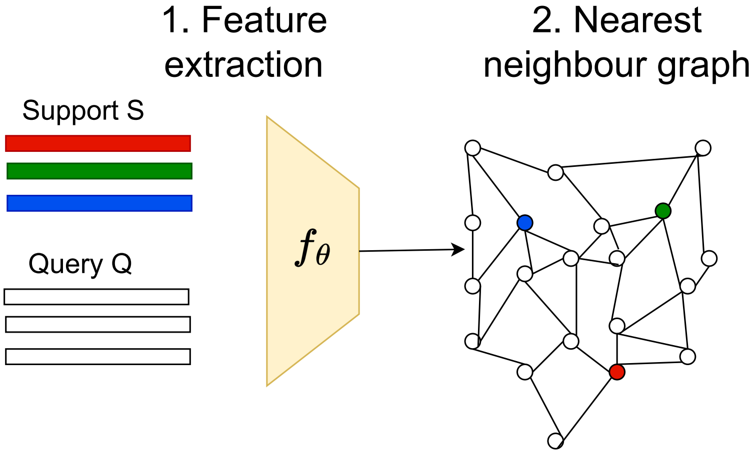
Pre-training stage

- We use publicly available pre-trained networks from published works.
- Base class dataset: $D_{\text{base}} := \{(x_i, y_i)\}_{i=1}^{I}$ where $y_i \in C_{\text{base}}$.
- Network $f_\theta : \mathcal{X} \to \mathbb{R}^d$ is trained on D_{base}.

Inference stage

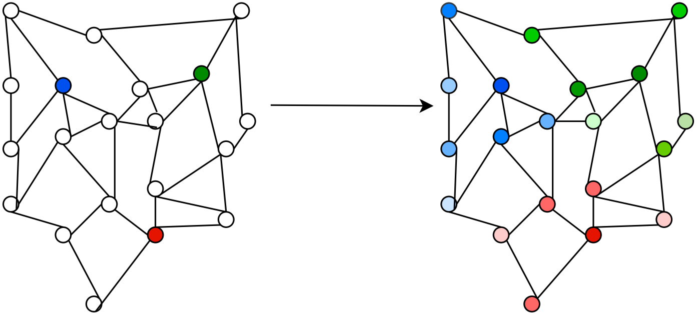
- Novel class dataset D_{novel} with C_{novel} disjoint from C_{base}.
- Assume access to f_θ, a support set, S, a query set, Q.
- We focus on transductive few-shot learning, where all S and Q are available at inference at the same time.

A^2LP: Nearest Graph Construction



2. Nearest neighbour graph

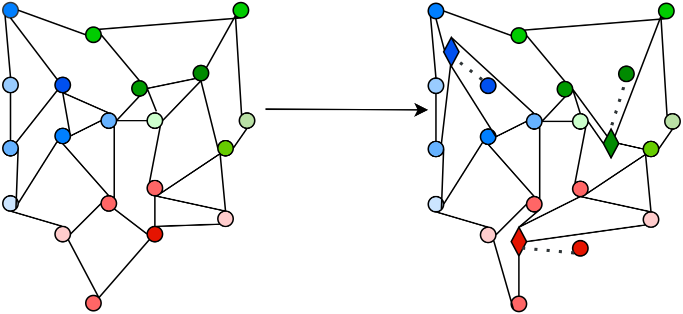
3. Label propagation



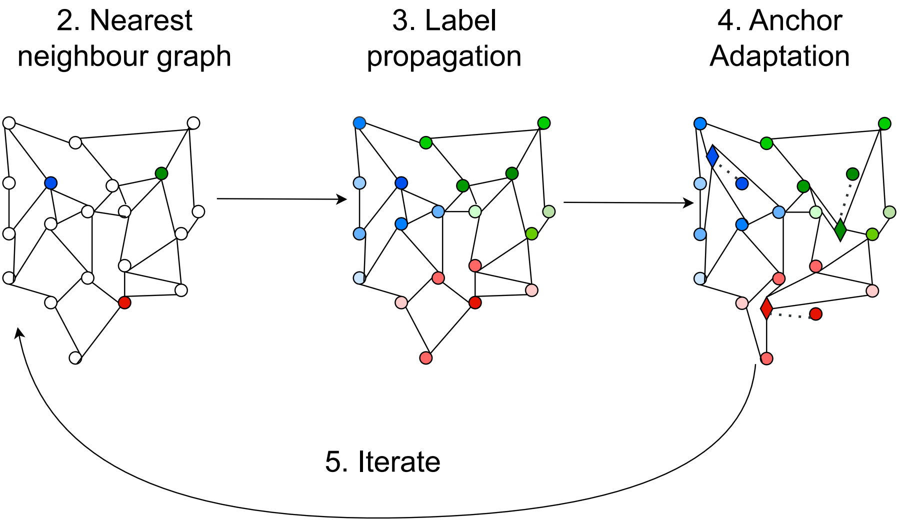
A^{2}LP: Anchor Adaptation

3. Label propagation

4. Anchor Adaptation



A²LP: Iteration



Experimental results: Implemented baselines comparisons

ALGORITHM	*mini*IMAGENET		*tiered*IMAGENET	
	1-shot	5-shot	1-shot	5-shot
RESNET-12				
Prototypical classifier	54.23$\pm$0.61	74.98$\pm$0.48	67.21$\pm$0.72	85.19$\pm$0.48
Imprinting+L_{ce}	54.53$\pm$0.61	74.31$\pm$0.49	67.53$\pm$0.72	84.86$\pm$0.48
LP	59.15$\pm$0.66	73.50$\pm$0.52	72.67$\pm$0.74	84.61$\pm$0.53
A^{2}LP	**64.35**$\pm$0.77	**75.00**$\pm$0.53	**80.49**$\pm$0.77	**85.80**$\pm$0.51
WIDERESNET-28-10				
Prototypical classifier	65.35$\pm$0.63	83.37$\pm$0.43	73.47$\pm$0.70	88.22$\pm$0.45
Imprinting+L_{ce}	66.07$\pm$0.62	83.34$\pm$0.42	74.07$\pm$0.69	88.55$\pm$0.43
LP	69.50$\pm$0.64	81.28$\pm$0.47	78.14$\pm$0.72	87.63$\pm$0.50
A^{2}LP	**76.48**$\pm$0.75	**83.57**$\pm$0.45	**82.82**$\pm$0.73	**88.80**$\pm$0.46

- **A^{2}LP** is significantly outperforming all implemented baselines.

Experimental results: PLC pre-processing

ALGORITHM	*mini*IMAGENET		*tiered*IMAGENET	
	1-shot	5-shot	1-shot	5-shot
	WIDERESNET-28-10			
Prototypical classifier	69.64 $\pm$0.60	84.61 $\pm$0.42	77.26 $\pm$0.65	89.22 $\pm$0.42
Imprint+L_{ce}	68.77 $\pm$0.60	84.24 $\pm$0.42	76.13 $\pm$0.66	88.95 $\pm$0.42
LP	74.24 $\pm$0.66	84.59 $\pm$0.44	82.48 $\pm$0.70	90.07 $\pm$0.45
$\mathbf{A^2LP}$	**75.94** $\pm$0.72	**85.67** $\pm$0.42	**83.68** $\pm$0.72	**90.53** $\pm$0.43

- PLC: power transform, ℓ_2-normalization, centering.
- Even when PLC pre-processing is used $\mathbf{A^2LP}$ is outperforming all baselines.

Experimental results: State of the art

Algorithm	*mini*ImageNet		*tiered*ImageNet	
	1-shot	5-shot	1-shot	5-shot
ResNet-12				
LR+ICI	66.80	79.26	80.79	87.92
CAN+Top-*k*	67.19±0.55	80.64±0.35	73.21±0.58	84.93 ±0.38
DPGN	67.77±0.32	84.60±0.43	72.45±0.51	87.24±0.39
WideResNet-28-10				
EP	70.74±0.85	84.34±0.53	78.50±0.91	88.36±0.57
SIB	70.00±0.60	79.20±0.40	72.90	82.80
SIB+E^{3}BM	71.40±0.50	81.20±0.40	75.60±0.60	84.30±0.40
LaplacianShot	74.86±0.19	84.13±0.14	80.18±0.21	87.56±0.15
A^{2}LP	**76.48**±0.75	83.57±0.45	82.82±0.73	88.80±0.46
A^{2}LP+PLC	75.94±0.72	**85.67**±0.42	**83.68**±0.72	**90.53**±0.43

- Outperforming several state of the art methods.

Conclusion

- We propose a novel variant of label propagation.
- Our algorithm iteratively adapts the labeled anchors.
- Significantly outperforms the traditional algorithm.
- Outperforms several state of the art methods.
- Future directions:
    - Different loss functions.
    - Beyond few-shot learning.

Imperial College London IARAI

Thank you!

https://github.com/MichalisLazarou

https://avrithis.net

http://www.commsp.ee.ic.ac.uk/~tania/