Iterative label cleaning for transductive and semi-supervised few-shot learning

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

[1]Imperial College London, [2]Inria, Univ Rennes, CNRS, IRISA

ICCV 2021

Virtual, October 2021

Motivation

- What is few-shot learning?

?

Contributions

Previous state of the art

- Meta-learning
- Transfer learning
- Domain adaptation
- Synthetic data generation

Contributions

- Novel algorithm that consists of three modules:
  - Label propagation
  - Class balancing
  - Label cleaning

Contributions

Previous state of the art

- Meta-learning
- Transfer learning
- Domain adaptation
- Synthetic data generation

Contributions

- Novel algorithm that consists of three modules:
    - Label propagation
    - Class balancing
    - Label cleaning

Problem formulation and definitions

Pre-training

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

Inference stage

- We focus on transductive and semi-supervised few-shot learning
- 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 and in the semi-supervised setting also an unlabelled set, U

Problem formulation and definitions

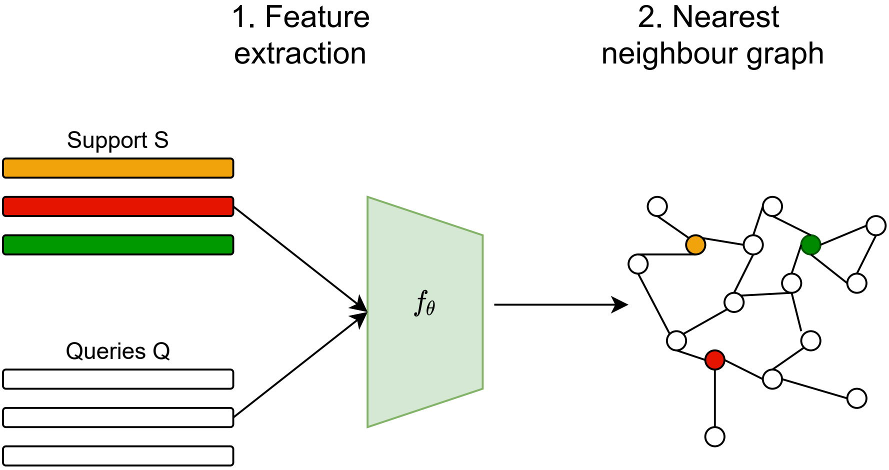
Pre-training

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

Inference stage

- We focus on transductive and semi-supervised few-shot learning
- 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 and in the semi-supervised setting also an unlabelled set, U

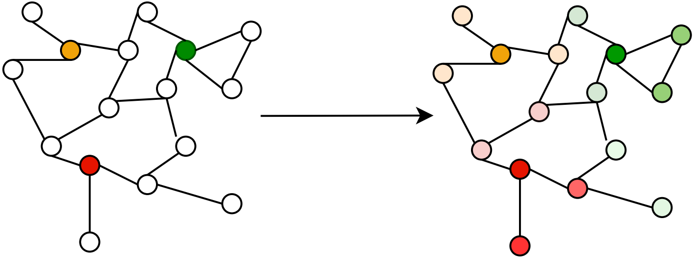
Iterative label cleaning: Nearest-neighbour graph



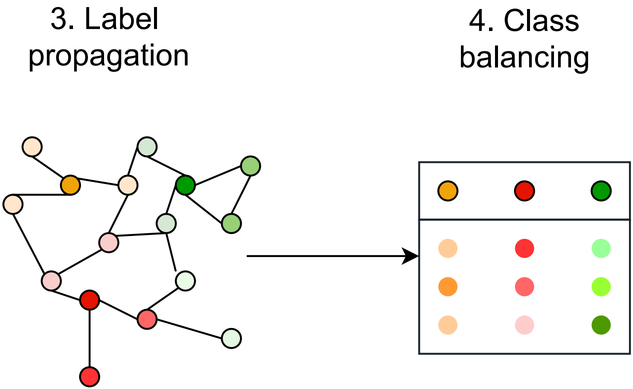
Iterative label cleaning: Label propagation

2. Nearest neighbour graph

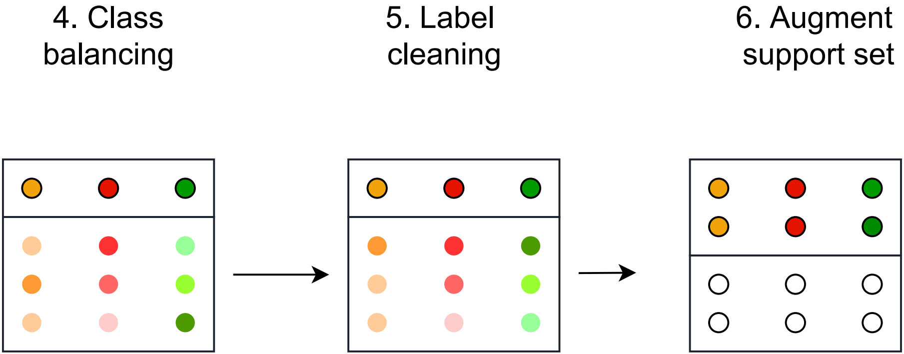
3. Label propagation



Iterative label cleaning: Class balancing



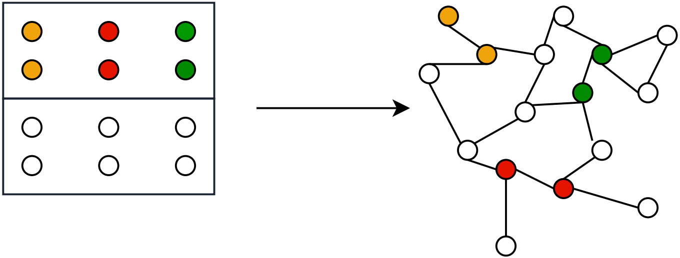
Iterative label cleaning: Label cleaning



Iterative label cleaning: Iterative inference

6. Augment Support Set

7. Iteration, nearest neighbour graph



Ablation Study

Label propagation

Inference	ResNet-12A		WRN-28-10	
	1-shot	5-shot	1-shot	5-shot
Inductive classifier	56.30 $\pm$0.62	**75.59** $\pm$0.47	68.17 $\pm$0.60	**84.33** $\pm$0.43
Label Propagation	**61.09** $\pm$0.70	75.32 $\pm$0.50	**74.24** $\pm$0.68	84.09 $\pm$0.42

- Exploit the manifold structure of the data

Ablation Study

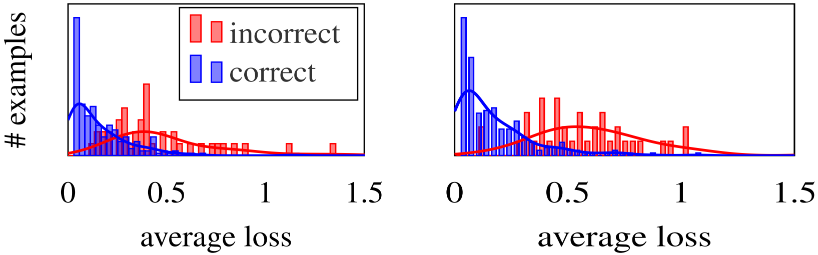
Class balancing

Balancing	Network	*mini*ImageNet		*tiered*ImageNet	
		1-shot	5-shot	1-shot	5-shot
None	WRN-28-10	78.06 ± 0.82	87.80 ± 0.42	86.04 ± 0.73	90.74 ± 0.46
True	WRN-28-10	**82.68** ± 0.82	**89.07** ± 0.41	**89.17** ± 0.70	**92.67** ± 0.44

- Incorporate prior information and search for a transport plan

Ablation Study

Label cleaning



Ablation Study

Iterative procedure

INFERENCE	RESNET-12A		WRN-28-10	
	1-shot	5-shot	1-shot	5-shot
Non-iterative	65.04$\pm$0.75	76.82$\pm$0.50	79.42$\pm$0.69	85.34$\pm$0.43
iterative (iLPC)	**69.79**$\pm$0.99	**79.82**$\pm$0.55	**83.05**$\pm$0.79	**88.82**$\pm$0.42

- Iterative selection of the most likely correctly classified queries

Experimental results

Transductive experiments

Method	Network	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	*tiered*ImageNet 1-shot	*tiered*ImageNet 5-shot
LR+ICI [63]	ResNet-12A	$66.85_{\pm 0.92}$	$78.89_{\pm 0.55}$	$82.40_{\pm 0.84}$	$88.80_{\pm 0.50}$
iLPC (ours)	ResNet-12A	$\mathbf{69.79}_{\pm 0.99}$	$\mathbf{79.82}_{\pm 0.55}$	$\mathbf{83.49}_{\pm 0.88}$	$\mathbf{89.48}_{\pm 0.47}$
PT+MAP [19]	WRN-28-10	$82.88_{\pm 0.73}$	$88.78_{\pm 0.40}$	$88.15_{\pm 0.71}$	$92.32_{\pm 0.40}$
LR+ICI [63]	WRN-28-10	$80.61_{\pm 0.80}$	$87.93_{\pm 0.44}$	$86.79_{\pm 0.76}$	$91.73_{\pm 0.40}$
iLPC (ours)	WRN-28-10	$\mathbf{83.05}_{\pm 0.79}$	$\mathbf{88.82}_{\pm 0.42}$	$\mathbf{88.50}_{\pm 0.75}$	$\mathbf{92.46}_{\pm 0.42}$

- State of the art results

Experimental results

Transductive experiments with more unlabeled queries

Method	Network	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	*tiered*ImageNet 1-shot	*tiered*ImageNet 5-shot
LR+ICI [63]	WRN-28-10	82.38$\pm$0.86	88.78$\pm$0.39	88.59$\pm$0.74	92.11$\pm$0.39
PT+MAP [19]	WRN-28-10	83.79$\pm$0.71	88.94$\pm$0.33	88.87$\pm$0.64	92.01$\pm$0.36
iLPC (ours)	WRN-28-10	**85.98**$\pm$0.74	**90.54**$\pm$0.31	**90.02**$\pm$0.70	**92.94**$\pm$0.37

- The performance gap from the other methods increases significantly because our method exploits the manifold structure of the data

Experimental results

Semi-supervised experiments

Method	Network	Split	*mini*ImageNet 1-shot	*mini*ImageNet 5-shot	*tiered*ImageNet 1-shot	*tiered*ImageNet 5-shot
LR+ICI [63]	ResNet-12A	30/50	67.57±0.97	79.07±0.56	83.32±0.87	89.06±0.51
iLPC (ours)	ResNet-12A	30/50	**70.99**±0.91	**81.06**±0.49	**85.04**±0.79	**89.63**±0.47
LR+ICI [63]	WRN-28-10	30/50	81.31±0.84	88.53±0.43	88.48±0.67	92.03±0.43
PT+MAP [19]	WRN-28-10	30/50	83.14±0.72	88.95±0.38	89.16±0.61	92.30±0.39
iLPC (ours)	WRN-28-10	30/50	**83.58**±0.79	**89.68**±0.37	**89.35**±0.68	**92.61**±0.39

- State of the art results

Thank you!

https://github.com/MichalisLazarou
http://www.commsp.ee.ic.ac.uk/~tania/
https://avrithis.net