ECCV'20 ONLINE
23-28 AUGUST 2020

Graph convolutional networks for learning with few clean and many noisy labels

Ahmet Iscen, Giorgos Tolias, Yannis Avrithis, Ondrej Chum, Cordelia Schmid

Few clean and many noisy images

Admiral

- Few-shot classification: challenging to learn with few examples

Few clean and many noisy images

Admiral

Web collection (tags, captions)

- Web-crawling with class names: weakly labeled (noisy) examples

Few clean and many noisy images

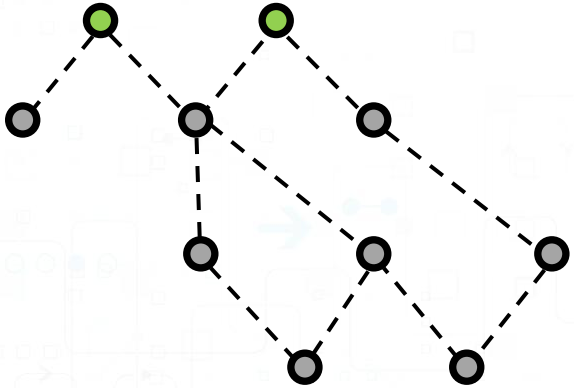
Admiral

Web collection (tags, captions)

- Our approach:
  - per-class cleaning process with GCN
  - appropriate use of noisy examples in classifier learning

GCN for noisy data cleaning

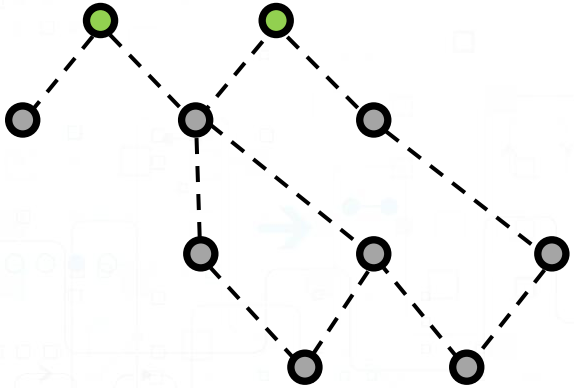
- Graph created with pre-trained embeddings
  - Based on reciprocal nearest neighbors



clean label

noisy (weak) label

GCN for noisy data cleaning



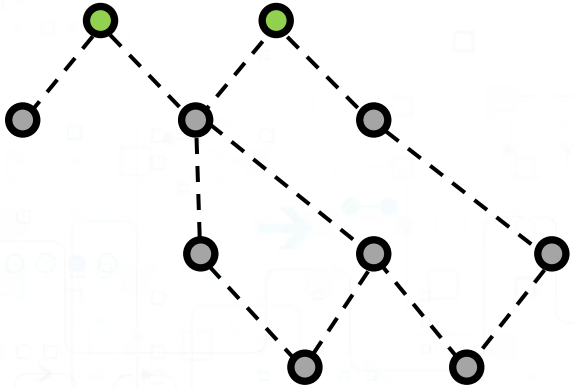
clean label

noisy (weak) label

- GCN binary classifier
  - 2 layers, 2 non-linearities: ReLU and sigmoid

$$F : \mathbb{R}^{N \times N} \times \mathbb{R}^{d \times N} \rightarrow \mathbb{R}^{N}$$

GCN for noisy data cleaning



clean label

noisy (weak) label

- GCN binary classifier
  - 2 layers, 2 non-linearities: ReLU and sigmoid

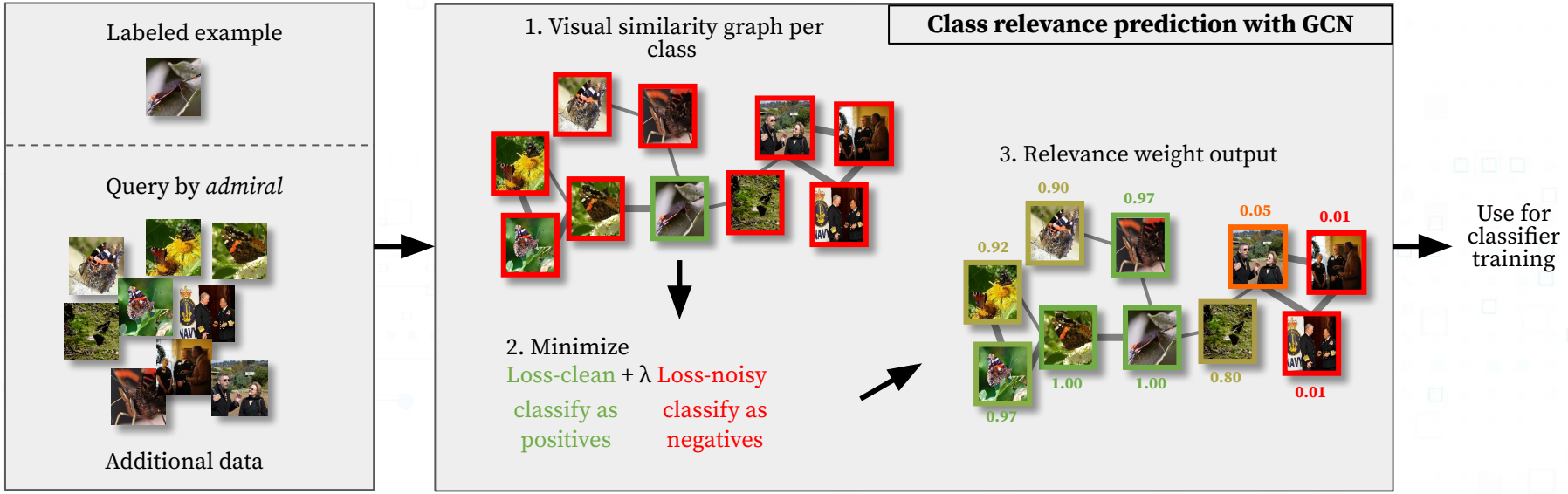
$$F : \mathbb{R}^{N \times N} \times \mathbb{R}^{d \times N} \rightarrow \mathbb{R}^{N}$$

- Binary cross-entropy loss
  - Targets output 1 for clean and 0 for noisy
  - λ - importance weight

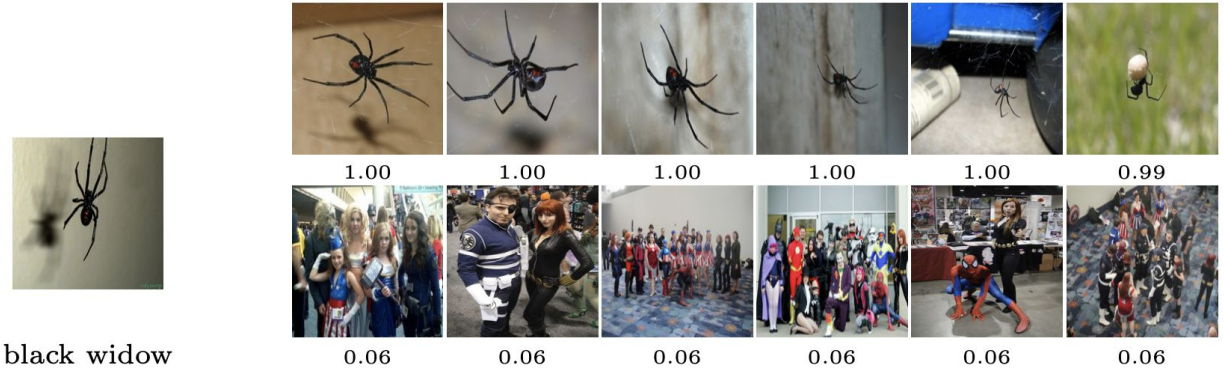
$$-\frac{1}{k}\sum_{i=1}^{k}\log\left(F(A,Z)_i\right) - \frac{\lambda}{N-k}\sum_{i=k+1}^{N}\log\left(1-F(A,Z)_i\right)$$

Overview

Cleaning is performed separately for each class



Examples of relevance weights



Experimental setup

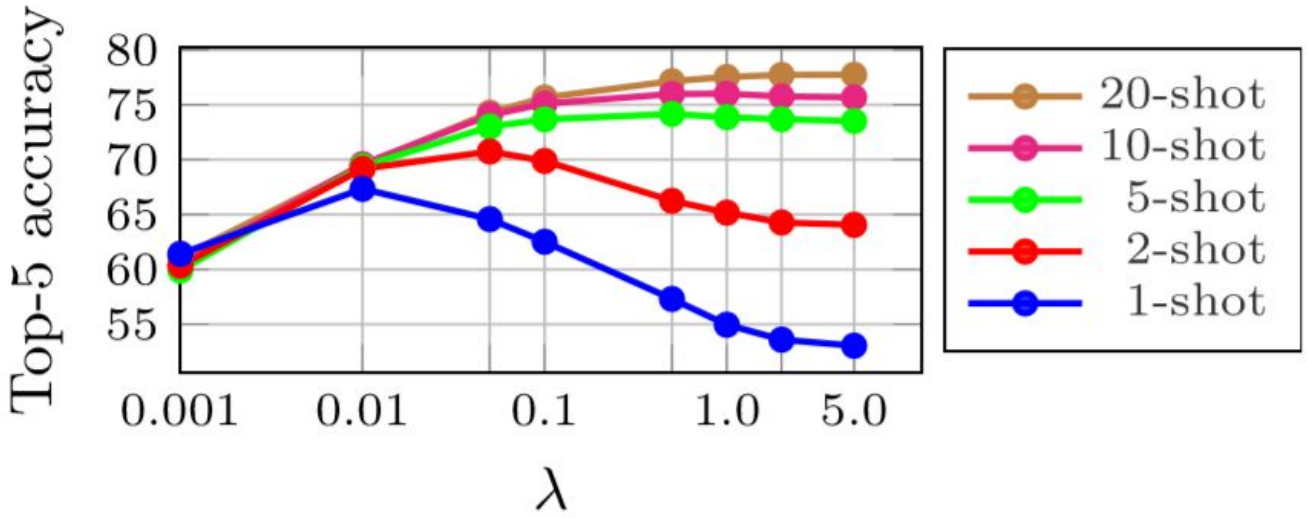
- Extend the Low-Shot ImageNet Benchmark (Hariharan and Girschick)
- Clean data: subset of ImageNet or Places365
  - k random examples
- Noisy data from YFCC100M
  - free-form user-tags and captions
- Feature extractor trained on non-overlapping subset of ImageNet or Places365
  - ResNet10 or ResNet50

Classifiers

- Prototypical classifier
  - Average embedding
  - Ours: weighted average according to relevance
- Linear classifier
  - FC layer trained with SGD and cross-entropy loss
  - Ours: weighted training examples according to relevance
- CNN classifier
  - Assumes access to images
  - Feature extractor and FC layer trained with SGD and cross-entropy loss
  - Ours: weighted training examples according to relevance

Importance of negative examples

Less clean examples -> more dependance on noisy data (smaller λ)



Results on Extended Low-Shot ImageNet

Method	k=1	5
	FEW CLEAN EXAMPLES	
Class proto. [9]	45.3±0.65	69.3±0.32
	FEW CLEAN & MANY NOISY EXAMPLES	
Similarity	49.8±0.29	64.2±0.32
β-weighting, $\beta = 1$	56.1±0.06	57.1±0.05
β-weighting, β^*	55.6±0.24	63.4±0.25
Linear	59.8±0.00	58.4±0.00
Label Propagation	62.6±0.35	74.6±0.30
MLP	63.6±0.41	73.7±0.25
Ours	67.8±0.10	73.9±0.17

without noisy

[9]: Gidaris, Komodakis, CVPR'18, Dynamic few-shot visual learning without forgetting

Results on Extended Low-Shot ImageNet

Method	k=1	5
	FEW CLEAN EXAMPLES	
Class proto. [9]	45.3±0.65	69.3±0.32
	FEW CLEAN & MANY NOISY EXAMPLES	
Similarity	49.8±0.29	64.2±0.32
β-weighting, $\beta = 1$	56.1±0.06	57.1±0.05
β-weighting, β^*	55.6±0.24	63.4±0.25
Linear	59.8±0.00	58.4±0.00
Label Propagation	62.6±0.35	74.6±0.30
MLP	63.6±0.41	73.7±0.25
Ours	67.8±0.10	73.9±0.17

without noisy

noisy without cleaning

[9]: Gidaris, Komodakis, CVPR'18, Dynamic few-shot visual learning without forgetting

Results on Extended Low-Shot ImageNet

Method	k=1	5
	FEW CLEAN EXAMPLES	
Class proto. [9]	45.3±0.65	69.3±0.32
	FEW CLEAN & MANY NOISY EXAMPLES	
Similarity	49.8±0.29	64.2±0.32
β-weighting, $\beta = 1$	56.1±0.06	57.1±0.05
β-weighting, β^*	55.6±0.24	63.4±0.25
Linear	59.8±0.00	58.4±0.00
Label Propagation	62.6±0.35	74.6±0.30
MLP	63.6±0.41	73.7±0.25
Ours	67.8±0.10	73.9±0.17

without noisy

noisy without cleaning

cleaning with MLP

[9]: Gidaris, Komodakis, CVPR'18, Dynamic few-shot visual learning without forgetting

Results on Extended Low-Shot ImageNet

Method	k=1	5
	FEW CLEAN EXAMPLES	
Class proto. [9]	45.3±0.65	69.3±0.32
	FEW CLEAN & MANY NOISY EXAMPLES	
Similarity	49.8±0.29	64.2±0.32
β-weighting, $\beta = 1$	56.1±0.06	57.1±0.05
β-weighting, β^*	55.6±0.24	63.4±0.25
Linear	59.8±0.00	58.4±0.00
Label Propagation	62.6±0.35	74.6±0.30
MLP	63.6±0.41	73.7±0.25
Ours	67.8±0.10	73.9±0.17

without noisy

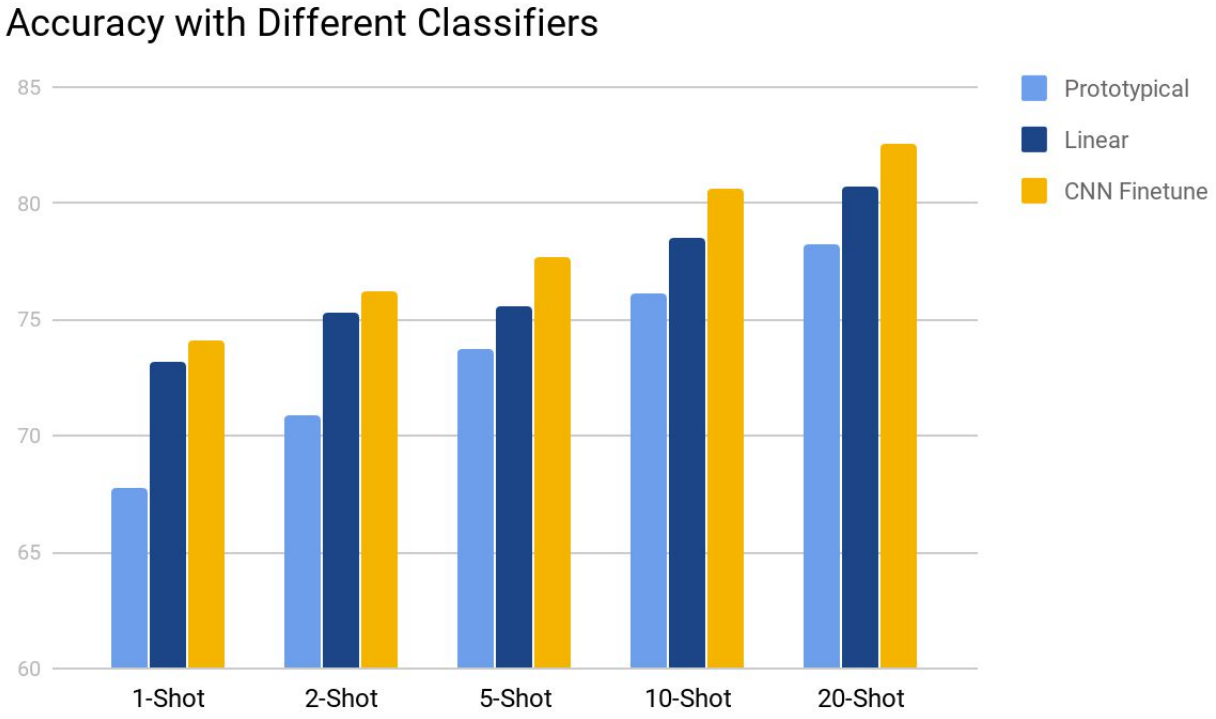
noisy without cleaning

cleaning with MLP

cleaning with GCN

[9]: Gidaris, Komodakis, CVPR'18, Dynamic few-shot visual learning without forgetting

Results on Extended Low-Shot ImageNet



Conclusions

- Supplement few-shot learning with additional noisy data
- GCN-based cleaning method to choose appropriate instances
- Significant improvement in accuracy
  - 45.3 to 74.1 in 1-shot learning

Paper is available at

arXiv:1910.00324