Ranking on manifolds for learning with limited supervision

Giorgos Tolias

Czech Technical University in Prague

Visual Recognition Group

CTU

CZECH TECHNICAL
UNIVERSITY
IN PRAGUE

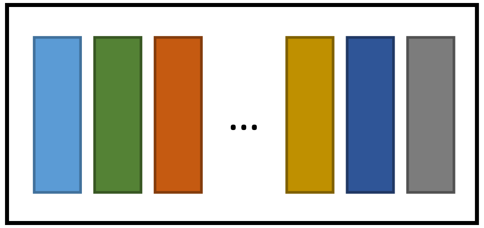
Supervision for deep learning

photo collection

human annotations

automatic annotations

training



deep network

tasks:

- **classification**
- **metric-learning**
- detection
- segmentation
- .....

[Iscen, **Tolias**, Avrithis, Chum. CVPR’18. Mining on Manifolds: Metric Learning without Labels]

[Iscen, **Tolias**, Avrithis, Chum. CVPR’19. Label Propagation for Deep Semi-supervised Learning]

[Iscen, **Tolias**, Avrithis, Chum, Schmid. ECCV’20. Graph convolutional networks for learning with few clean and many noisy labels]

Smoothness assumption

``if two points are close, then, so should be the corresponding output''.

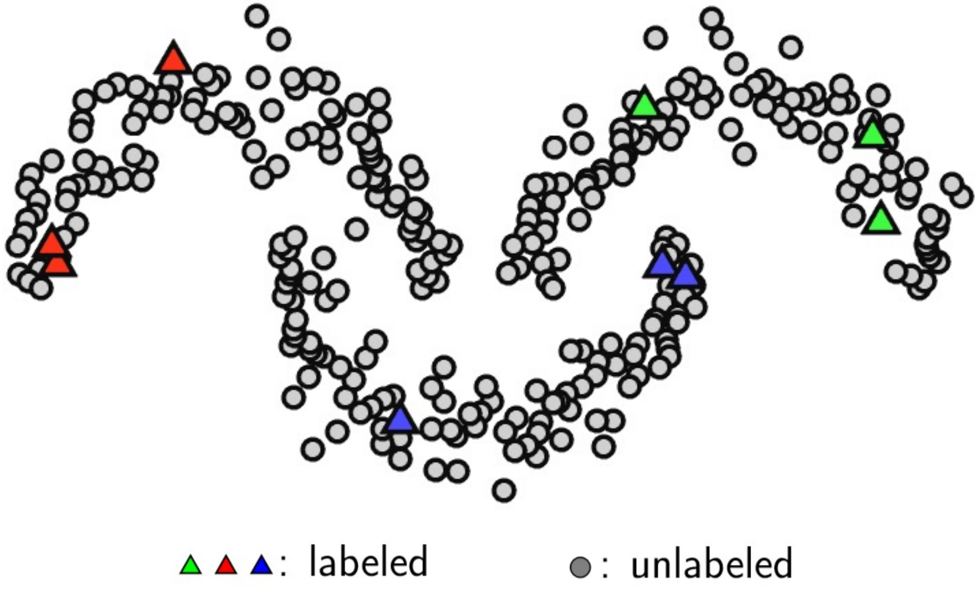
- classification → same label
  - Label propagation [Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

- similarity search → same similarity w.r.t a reference point
  - Ranking on manifolds [Zhou et al, NeurIPS 2003, Ranking on data manifolds]

Background: label propagation (LP)

- semi-supervised classification
- transductive learning

- *N* examples, *C* classes

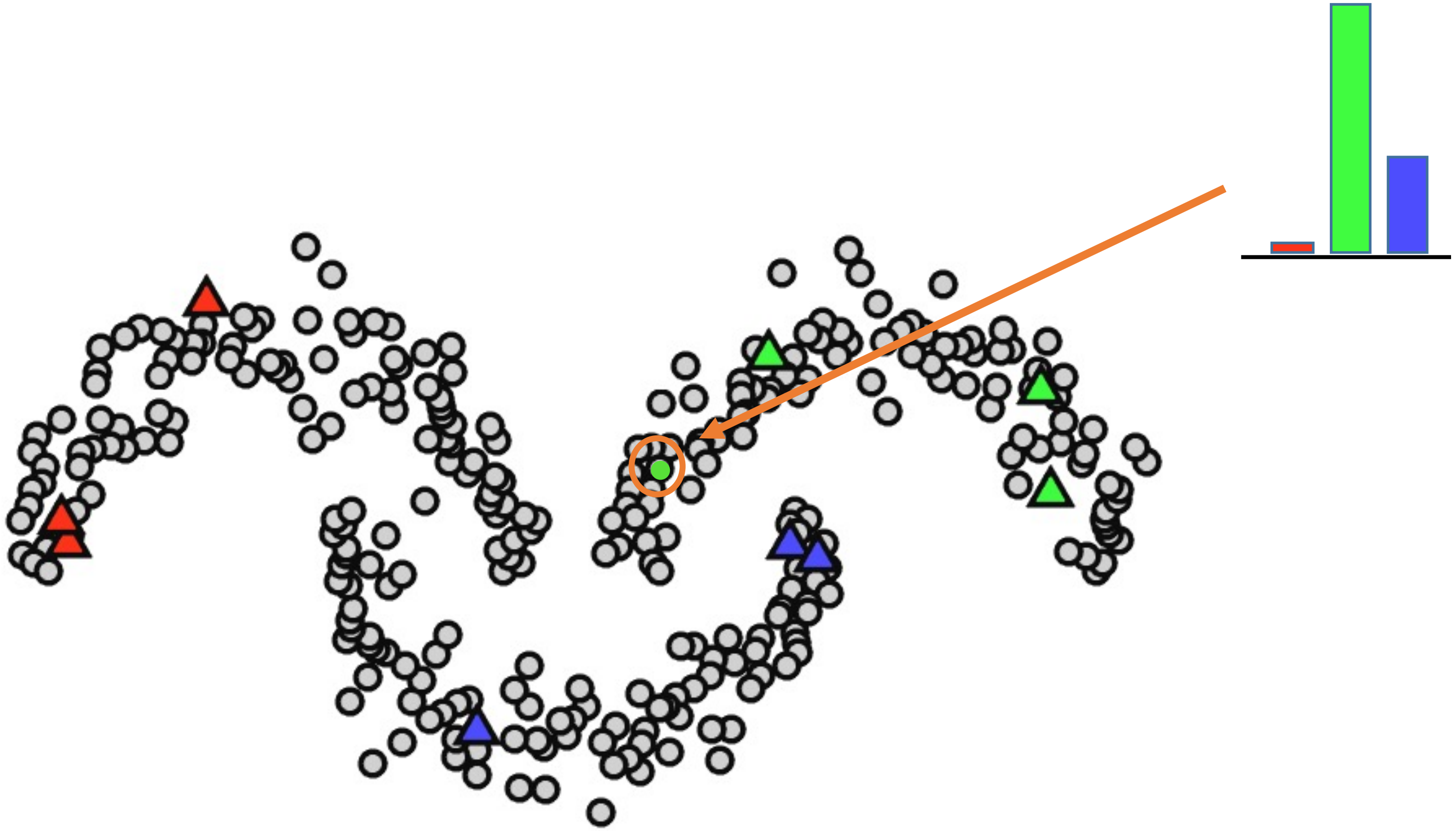


$$F^t = \alpha S F^{t-1} + (1-\alpha) Y$$

Y: *N* x *C* 1-hot-coded label matrix
S: *N* x *N* normalized affinity matrix (k-nearest neighbor graph)
α: propagation parameter in (0,1)
F: *N* x *C* class confidence matrix

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

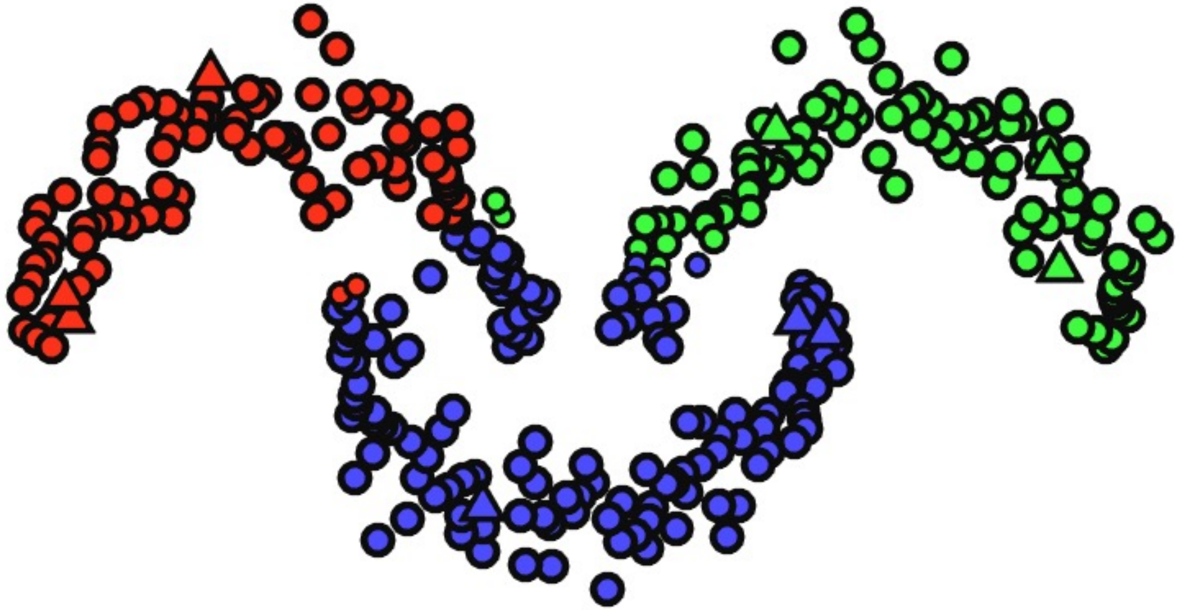
Background: LP iterations



iteration 0

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

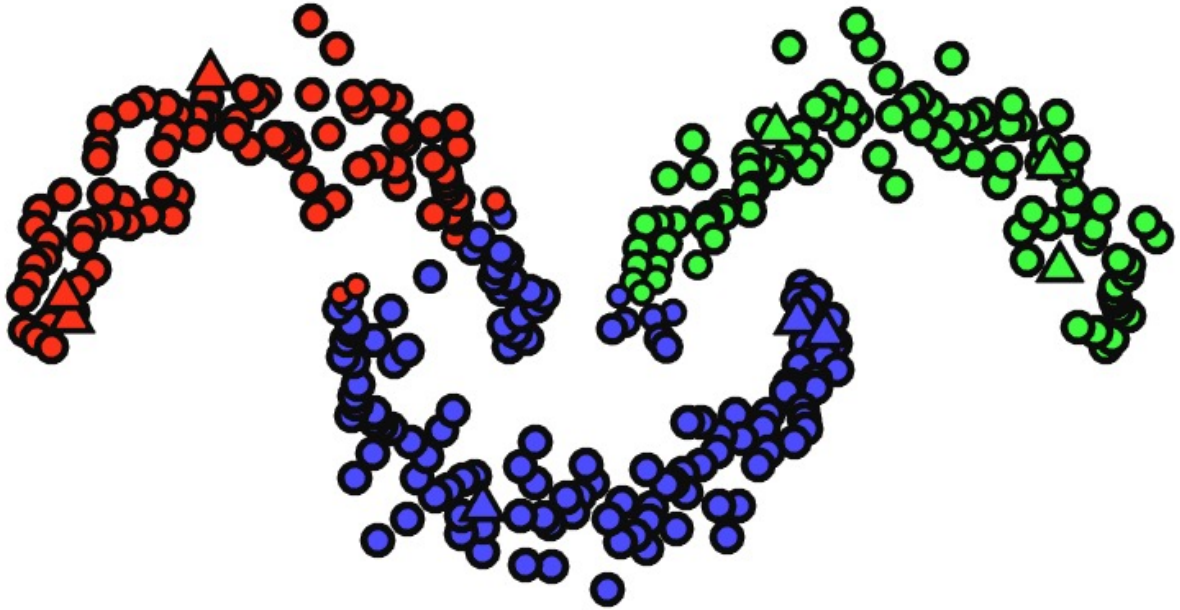
Background: LP iterations



iteration 1

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

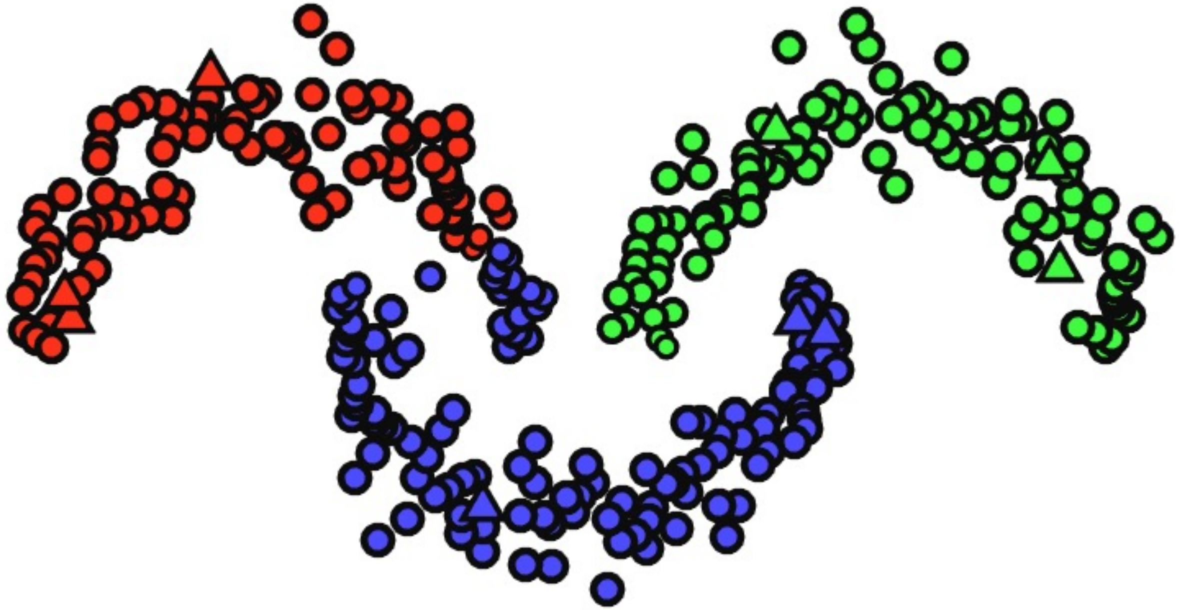
Background: LP iterations



iteration 2

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

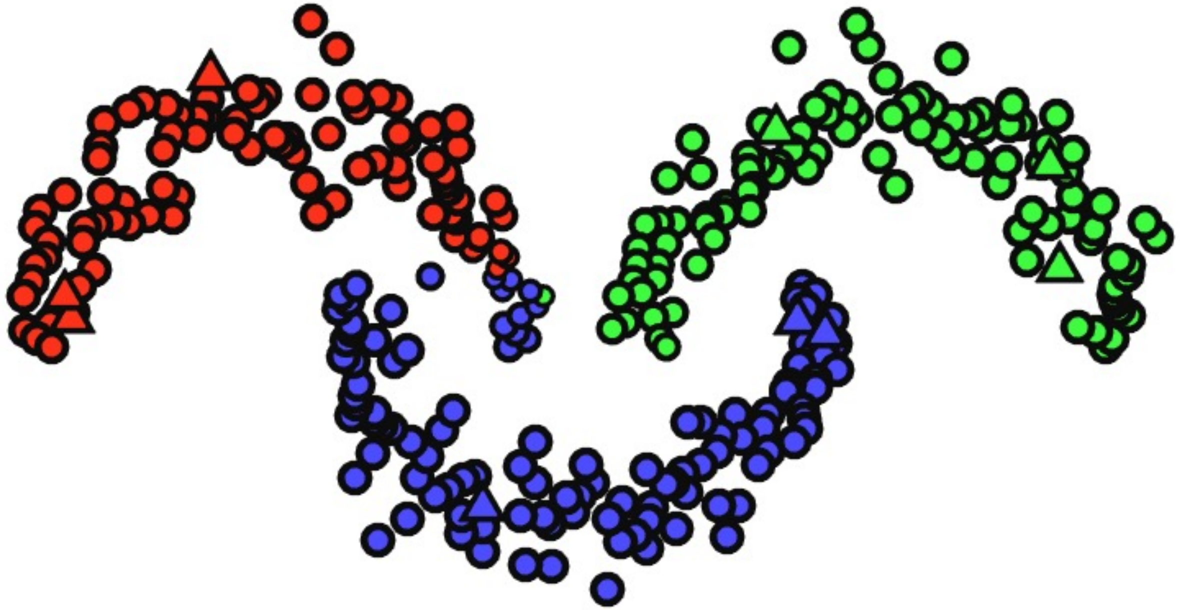
Background: LP iterations



iteration 5

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

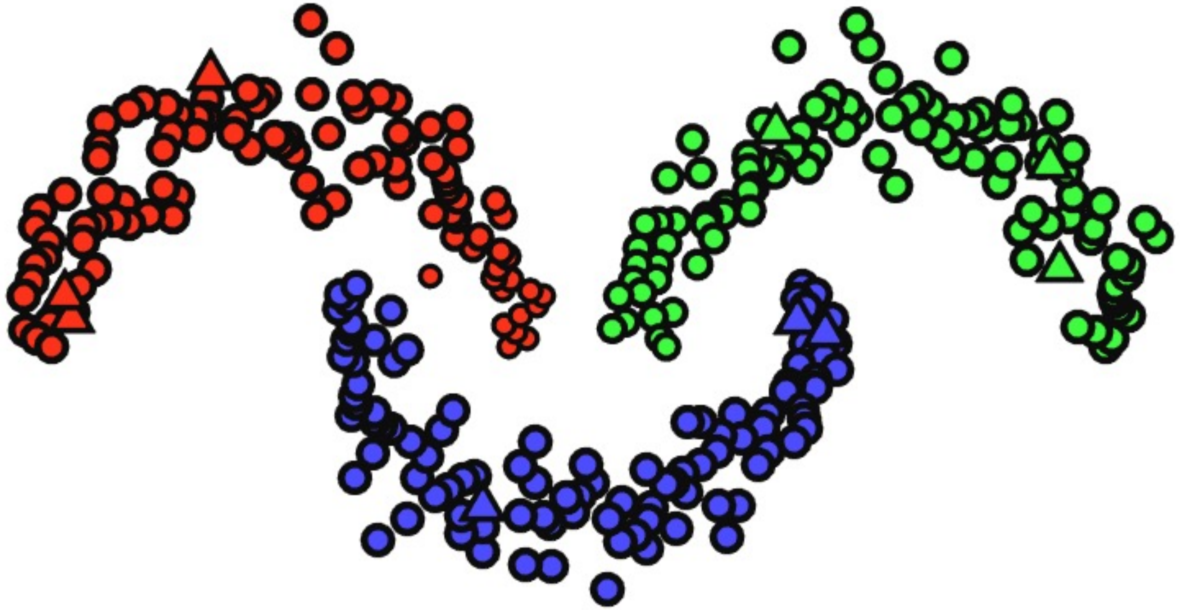
Background: LP iterations



iteration 10

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

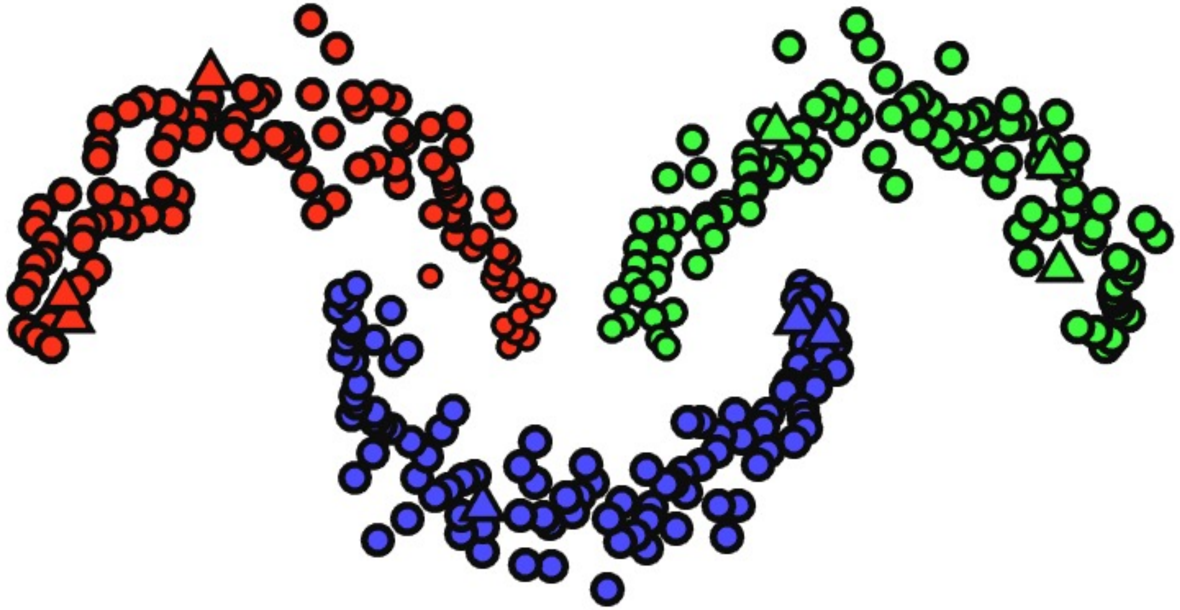
Background: LP iterations



iteration 50

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

Background: LP iterations



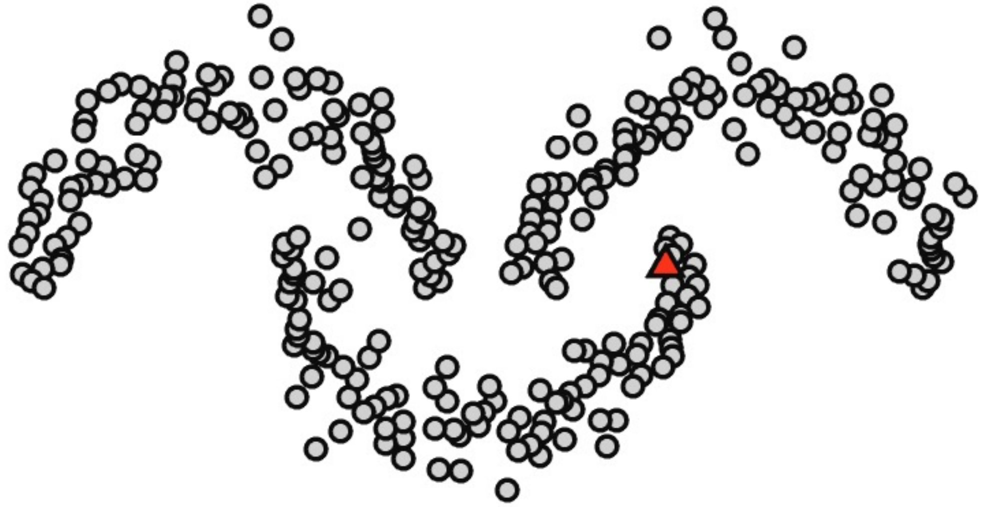
iteration 100

[Zhou et al, NeurIPS 2003, Learning with Local and Global Consistency]

Background: ranking on manifolds

- similarity search (unsupervised)

- *N* examples, 1 query

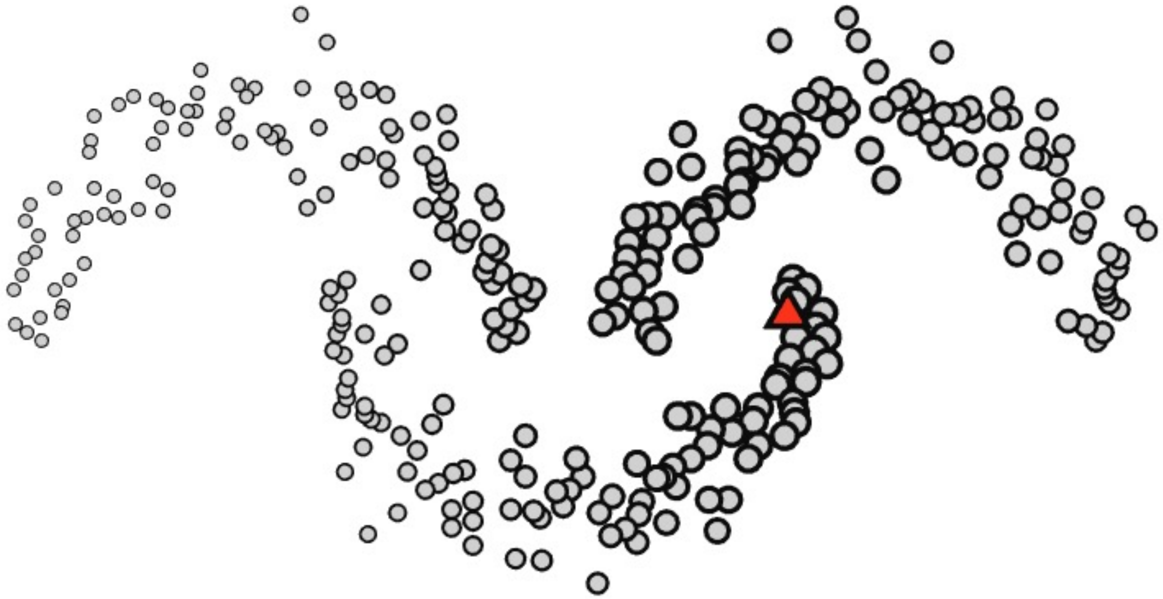


$$\mathbf{f}^t = \alpha S \mathbf{f}^{t-1} + (1-\alpha)\mathbf{y}$$

y: *N* x *1* 1-hot-coded vector that defines the query
S: *N* x *N* normalized affinity matrix (k-nearest neighbor graph)
α: propagation parameter in (0,1)
f: *N* x *1* similarity vector *w.r.t* the query

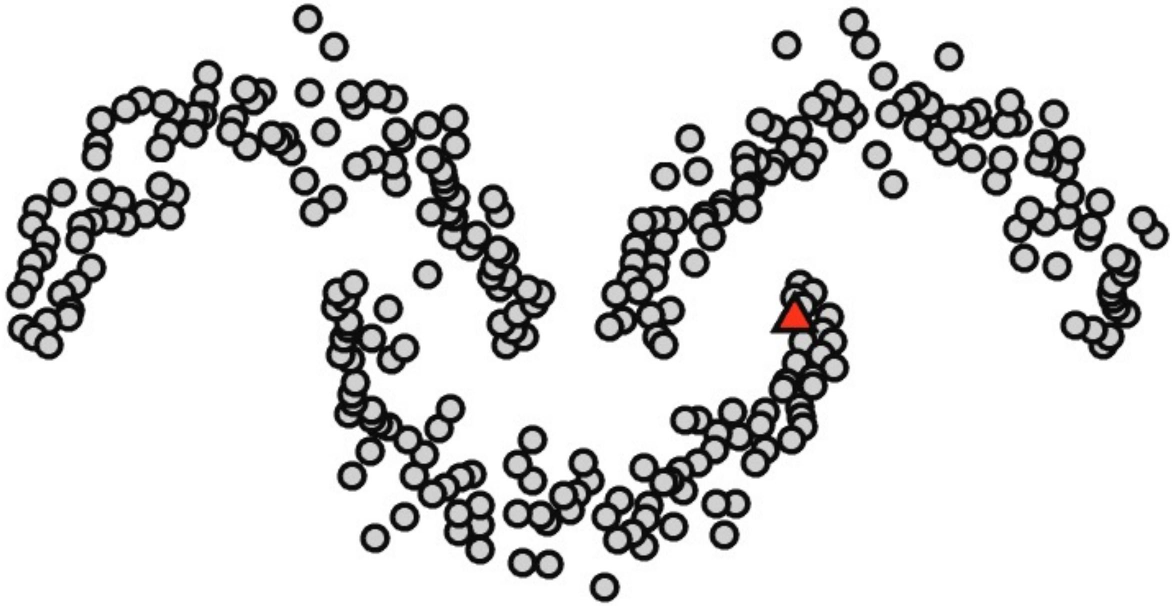
[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

Background: Euclidean similarity



[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

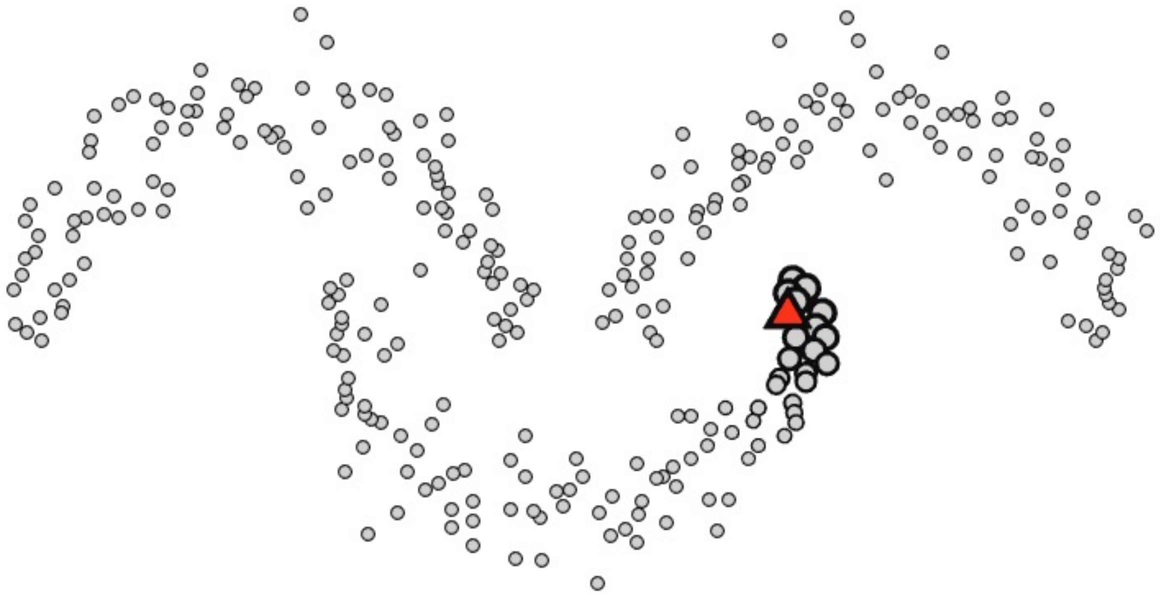
Background: similarity propagation iterations



iteration 0

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

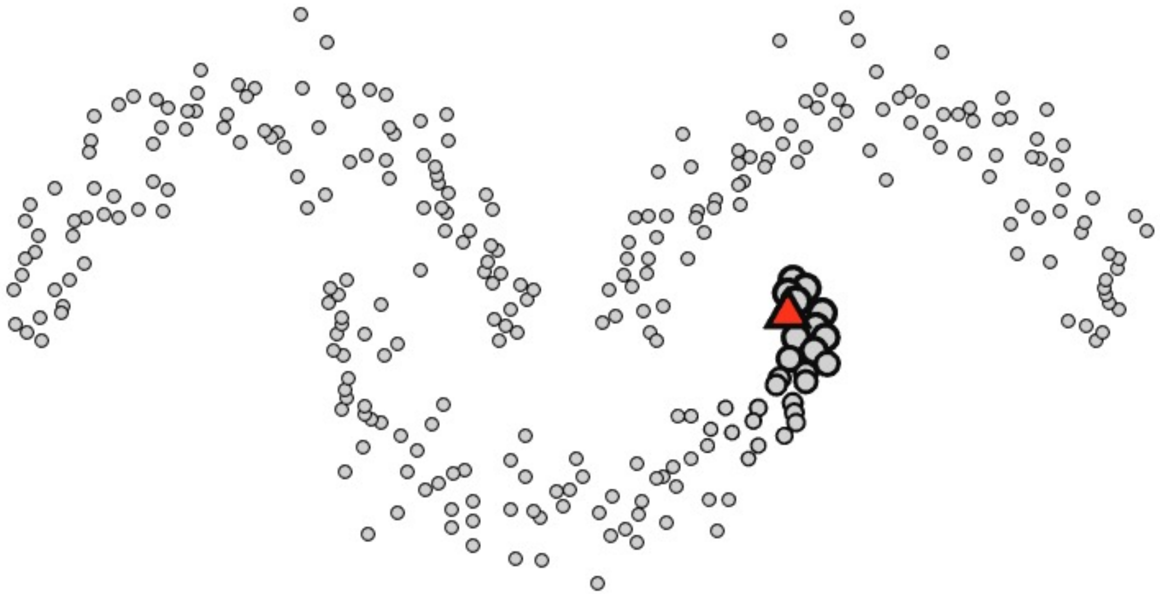
Background: similarity propagation iterations



iteration 1

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

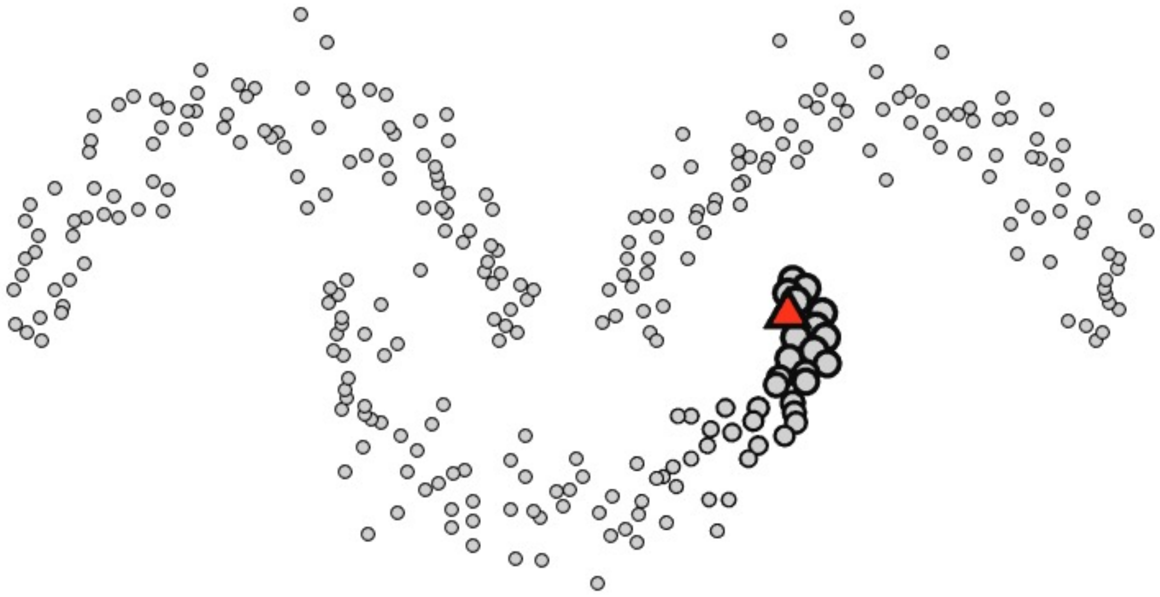
Background: similarity propagation iterations



iteration 2

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

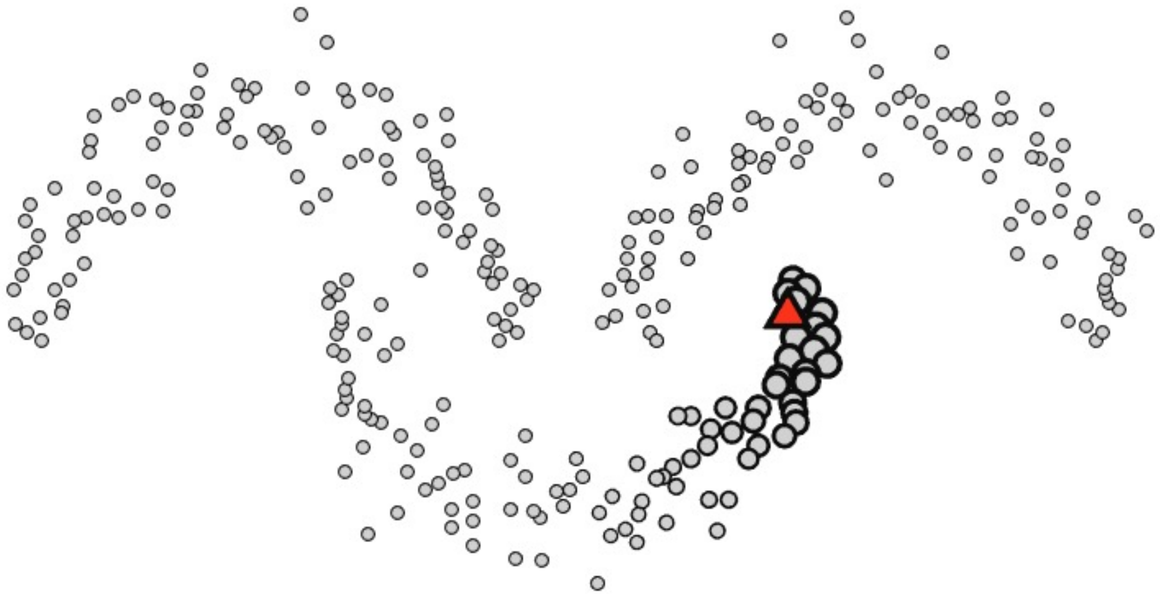
Background: similarity propagation iterations



iteration 5

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

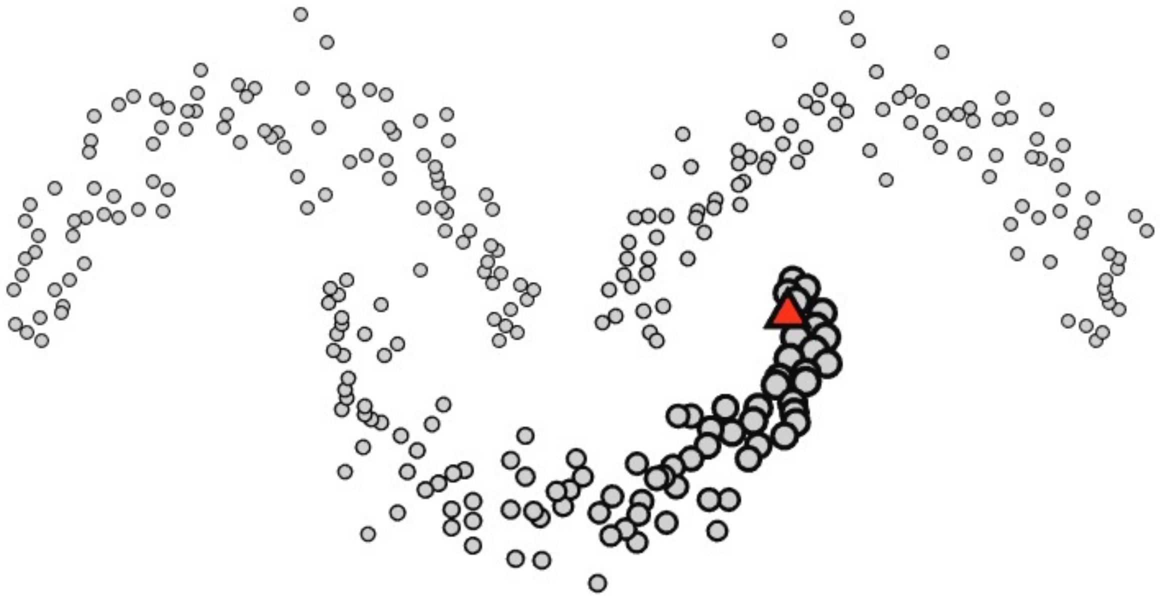
Background: similarity propagation iterations



iteration 10

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

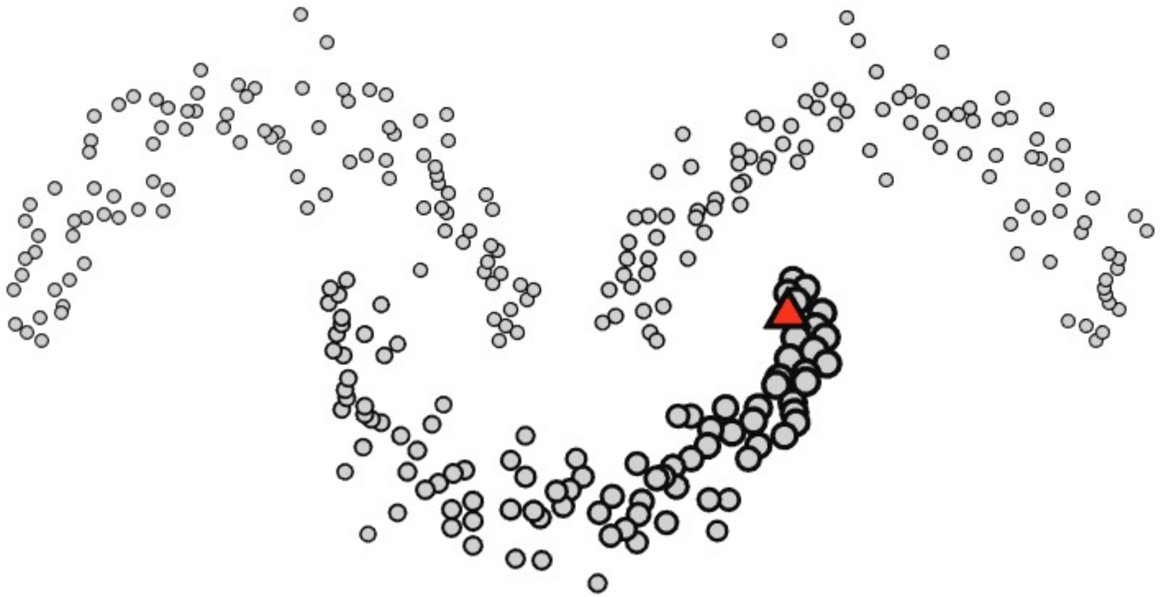
Background: similarity propagation iterations



iteration 50

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

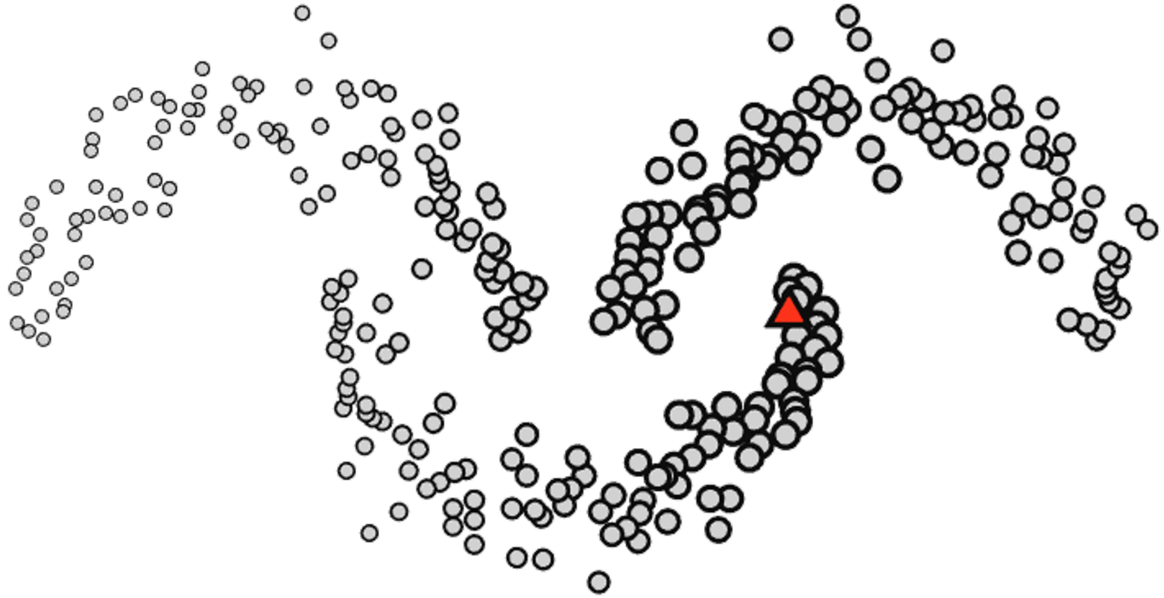
Background: similarity propagation iterations



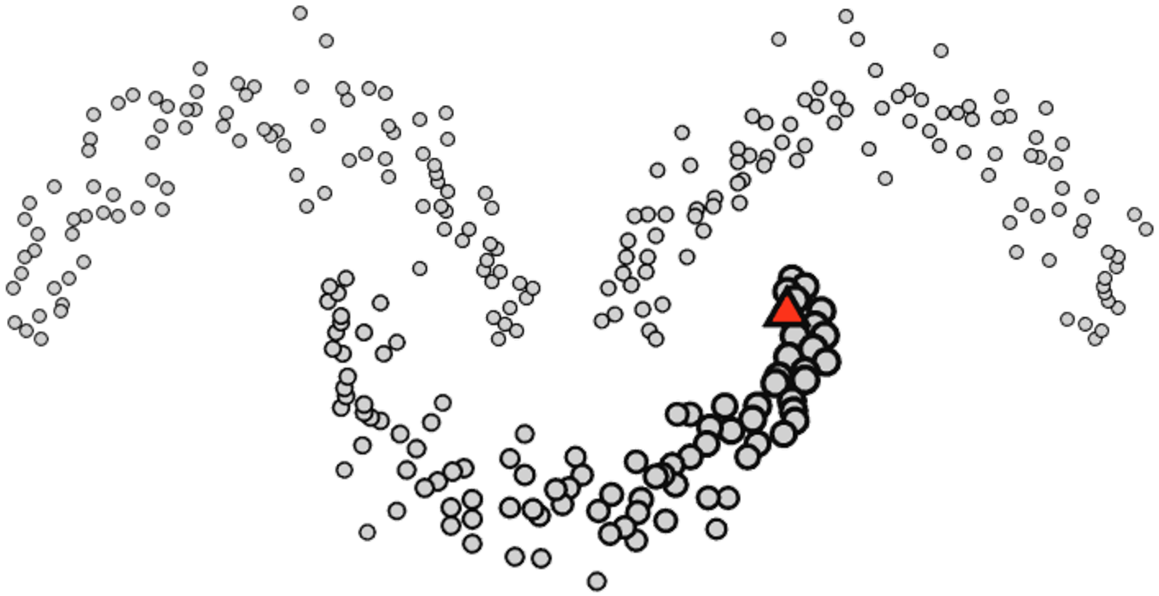
iteration 100

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

Background: Euclidean vs manifold similarity



Euclidean



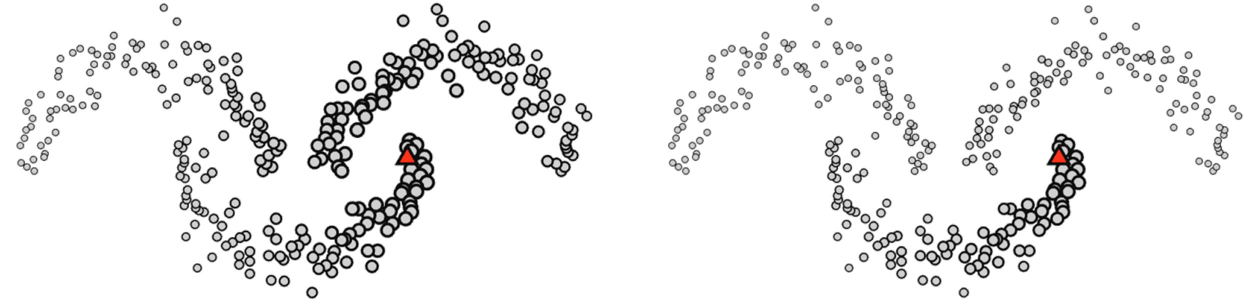
Manifold

[Zhou et al, NeurIPS 2003, Ranking on data manifolds]

PCA example

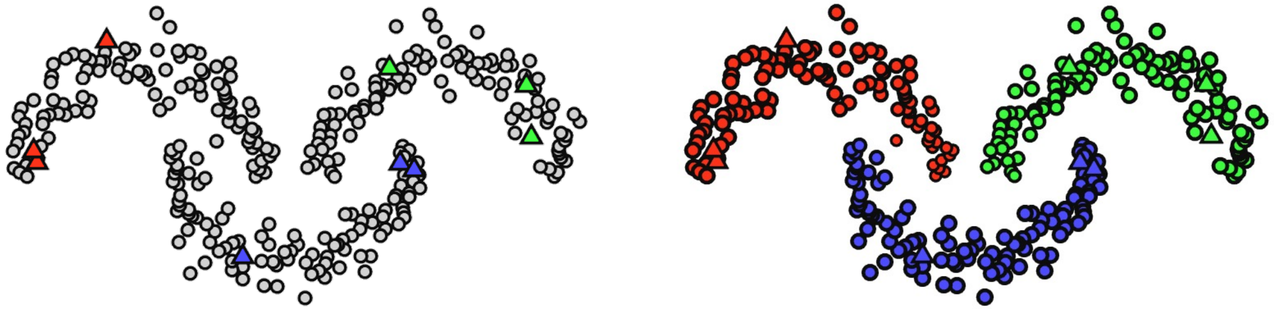
PCA example

Propagation for deep learning



self-supervised deep metric learning

[Iscen, **Tolias**, Avrithis, Chum. CVPR’18. Mining on Manifolds: Metric Learning without Labels]



semi-supervised deep learning for classification

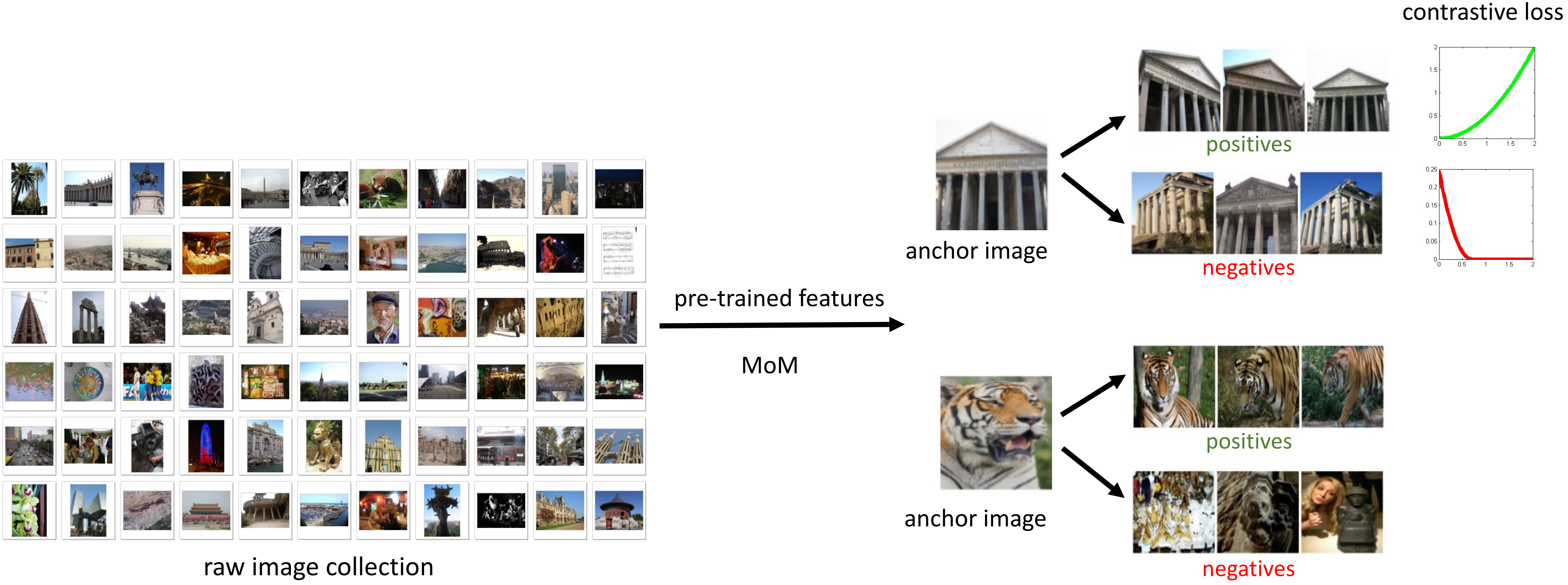
[Iscen, **Tolias**, Avrithis, Chum. CVPR’19. Label Propagation for Deep Semi-supervised Learning]

Self-supervised deep metric learning

self-supervised deep metric learning

[Iscen, **Tolias**, Avrithis, Chum. CVPR’18. Mining on Manifolds: Metric Learning without Labels]

Mining on manifolds (MoM) - goal



- automatically mine pairs of matching and non-matching images
  - use with any standard pairwise loss

Anchor images

- nearest neighbor graph *A*
- stationary probability distribution π of random walk on *A*

  $\pi = \pi P$,

  $P = D^{-1}A$

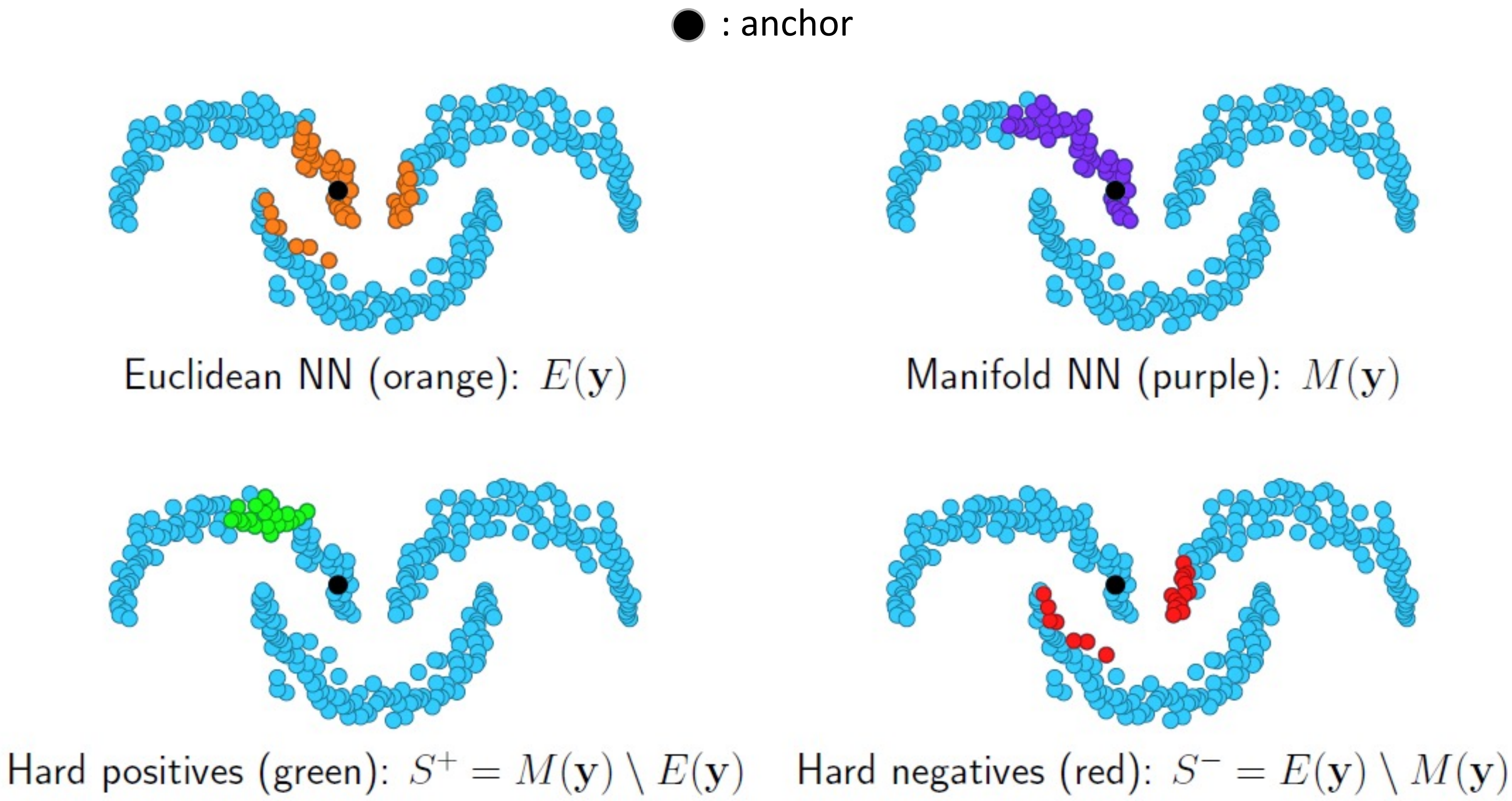
  D = degree matrix of A
- anchors: graph local-max based on π

top-100 anchors: random sample

top-1000 anchors: random sample

example from 10^6 input images

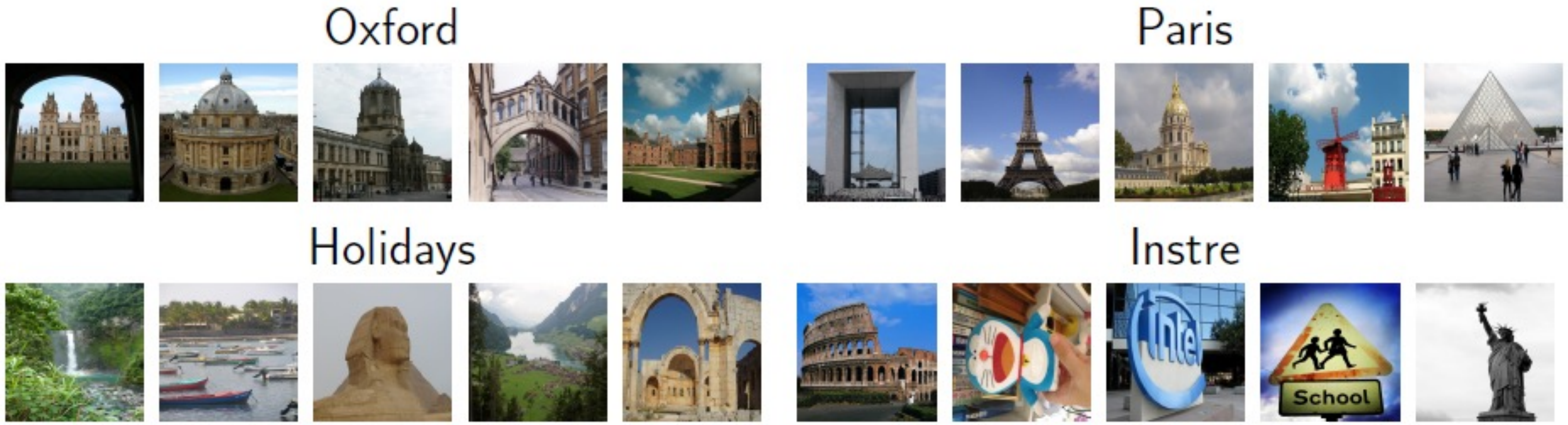
Mining on manifolds

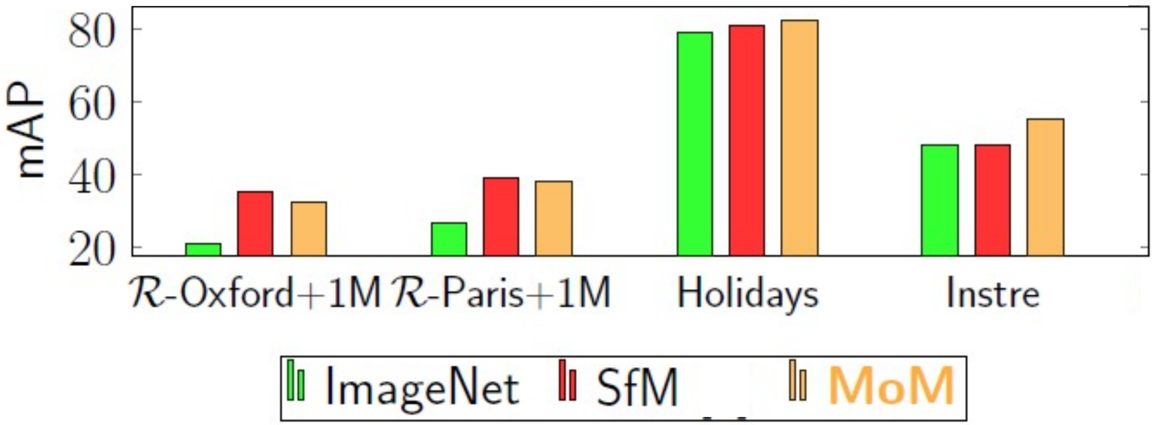


Mining of training samples from 10^6 images

Use VGG16 pre-trained on ImageNet as feature extractor

Experiments on instance search



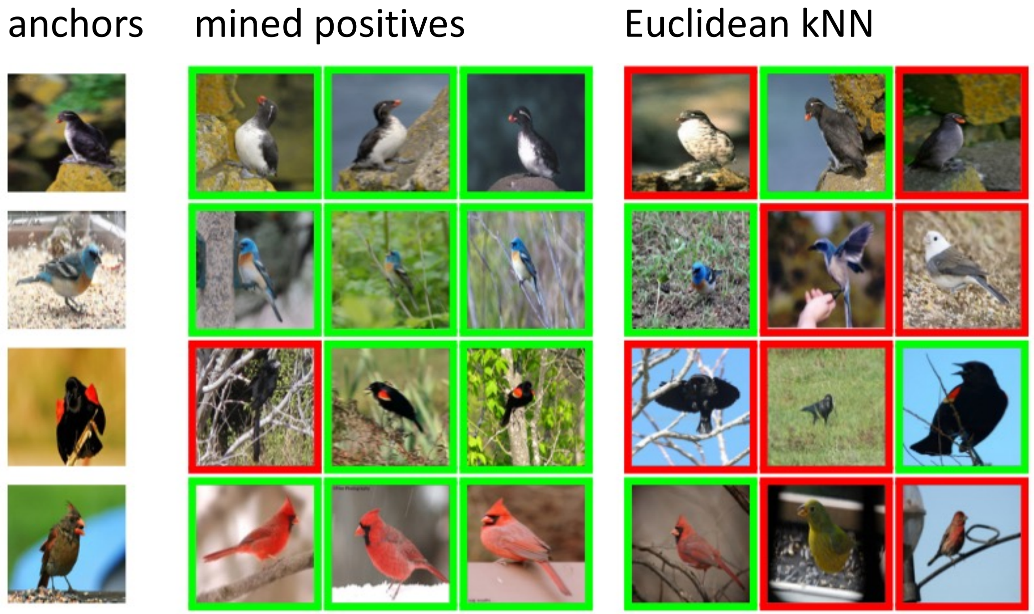


initialize: pre-training on ImageNet -- fine-tune: MoM on 10^6 images

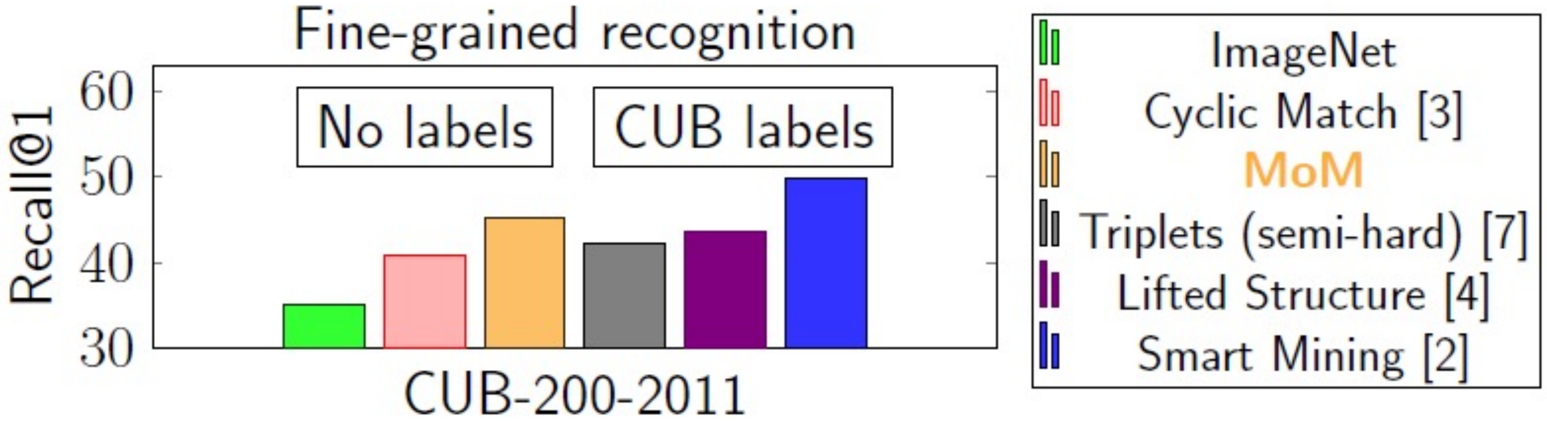
SfM: [Radenović et al. ECCV'16. CNN Image Retrieval Learns from BoW: Unsupervised Fine-Tuning with Hard Examples]

Mining of training samples

CUB 200



Experiments on fine-grained recognition



Cyclic Match: [Li et al. ECCV'16. Unsupervised visual representation learning by graph-based consistent constraints]
Triplets (semi-hard): [Schroff et al. CVPR'15. Facenet: A unified embedding for face recognition and clustering]
Lifted Structure: [Song et al. CVPR'16. Deep metric learning via lifted structured feature embedding]
Smart Mining: [Harwood et al. ICCV'17. Smart mining for deep metric learning. In ICCV, 2017]

Ablation experiment

Positive	Negative	CUB	Oxford5k	
Anchors		All	Random	$\mathcal{A}$
Initial		35.0	52.6	
NN_5^e	$X \setminus \mathrm{NN}_5^e$	38.5	37.4	41.9
P^+	$X \setminus \mathrm{NN}_5^e$	43.0	48.2	38.1
NN_5^e	P^-	42.1	57.8	71.3
P^+	P^-	43.5	64.4	73.7
$P^+ + W$	$P^- + W$	**45.3**	67.0	**76.7**

Ablation experiment: anchors

Positive	Negative	CUB	Oxford5k	
Anchors		All	Random	$\mathcal{A}$
Initial		35.0	52.6	
NN_5^e	$X \backslash \text{NN}_5^e$	38.5	37.4	41.9
P^+	$X \backslash \text{NN}_5^e$	43.0	48.2	38.1
NN_5^e	P^-	42.1	57.8	71.3
P^+	P^-	43.5	$\boxed{64.4}$	$\boxed{73.7}$
$P^+ + W$	$P^- + W$	**45.3**	67.0	**76.7**

random anchors vs proposed anchors

Ablation experiment: weights

Positive	Negative	CUB	Oxford5k	
Anchors		All	Random	$\mathcal{A}$
Initial		35.0	52.6	
NN_5^e	$X \setminus \text{NN}_5^e$	38.5	37.4	41.9
P^+	$X \setminus \text{NN}_5^e$	43.0	48.2	38.1
NN_5^e	P^-	42.1	57.8	71.3
P^+	P^-	43.5	64.4	$\boxed{73.7}$
$P^+ + W$	$P^- + W$	**45.3**	67.0	$\boxed{\mathbf{76.7}}$

manifold similarity as pair weight

Ablation experiment: hard positives

Positive	Negative	CUB	Oxford5k	
Anchors		All	Random	$\mathcal{A}$
Initial		35.0	52.6	
NN_5^e	$X \backslash \mathrm{NN}_5^e$	38.5	37.4	41.9
P^+	$X \backslash \mathrm{NN}_5^e$	43.0	48.2	38.1
NN_5^e	P^-	42.1	57.8	$\boxed{71.3}$
P^+	P^-	43.5	64.4	$\boxed{73.7}$
$P^+ + W$	$P^- + W$	**45.3**	67.0	**76.7**

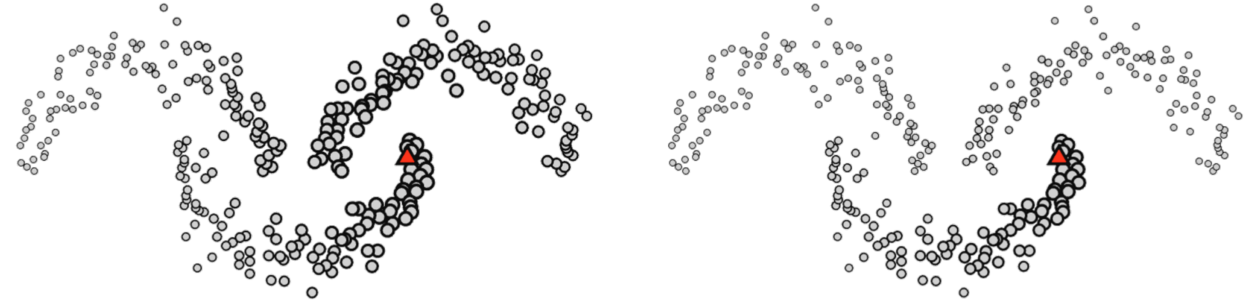
hard positives (P^+) vs easy positives (NN^e)

Ablation experiment: hard negatives

Positive	Negative	CUB	Oxford5k	
Anchors		All	Random	$\mathcal{A}$
Initial		35.0	52.6	
NN_5^e	$X \setminus \mathrm{NN}_5^e$	38.5	37.4	41.9
P^+	$X \setminus \mathrm{NN}_5^e$	43.0	48.2	$\boxed{38.1}$
NN_5^e	P^-	42.1	57.8	71.3
P^+	P^-	43.5	64.4	$\boxed{73.7}$
$P^+ + W$	$P^- + W$	**45.3**	67.0	**76.7**

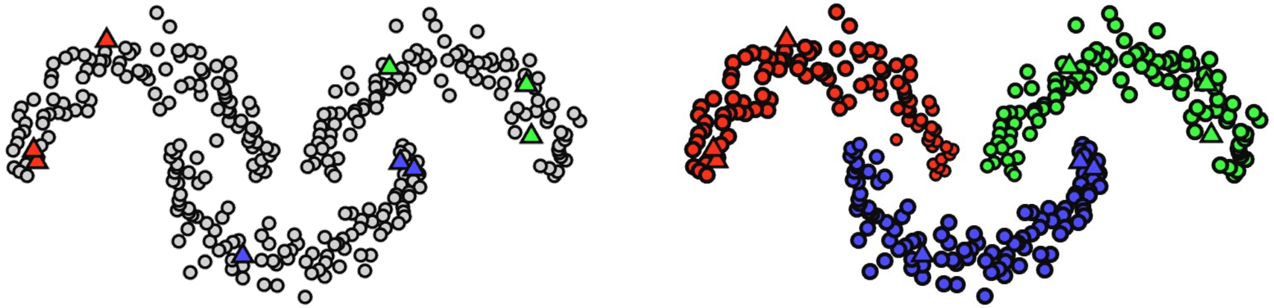
hard negatives (P^-) vs easy negatives ($X \setminus \mathrm{NN}^e$)

Propagation for deep learning



self-supervised deep metric learning

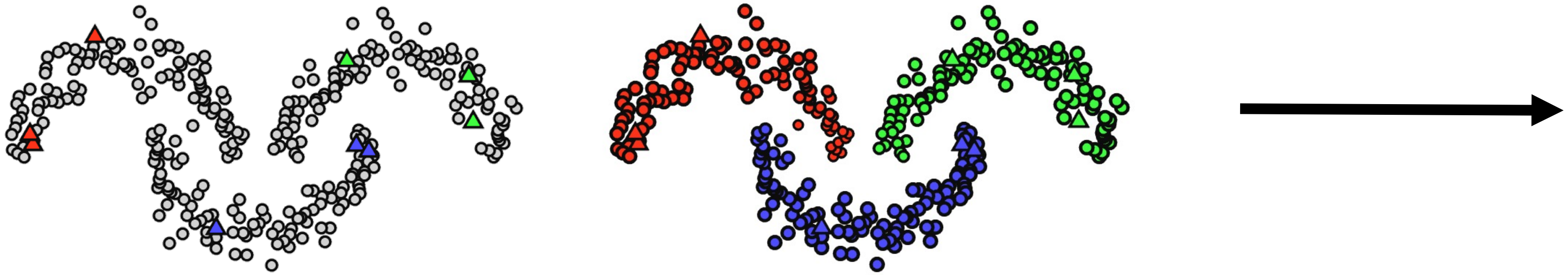
[Iscen, **Tolias**, Avrithis, Chum. CVPR'18. Mining on Manifolds: Metric Learning without Labels]



semi-supervised deep learning for classification

[Iscen, **Tolias**, Avrithis, Chum. CVPR'19. Label Propagation for Deep Semi-supervised Learning]

Semi-supervised deep learning for classification

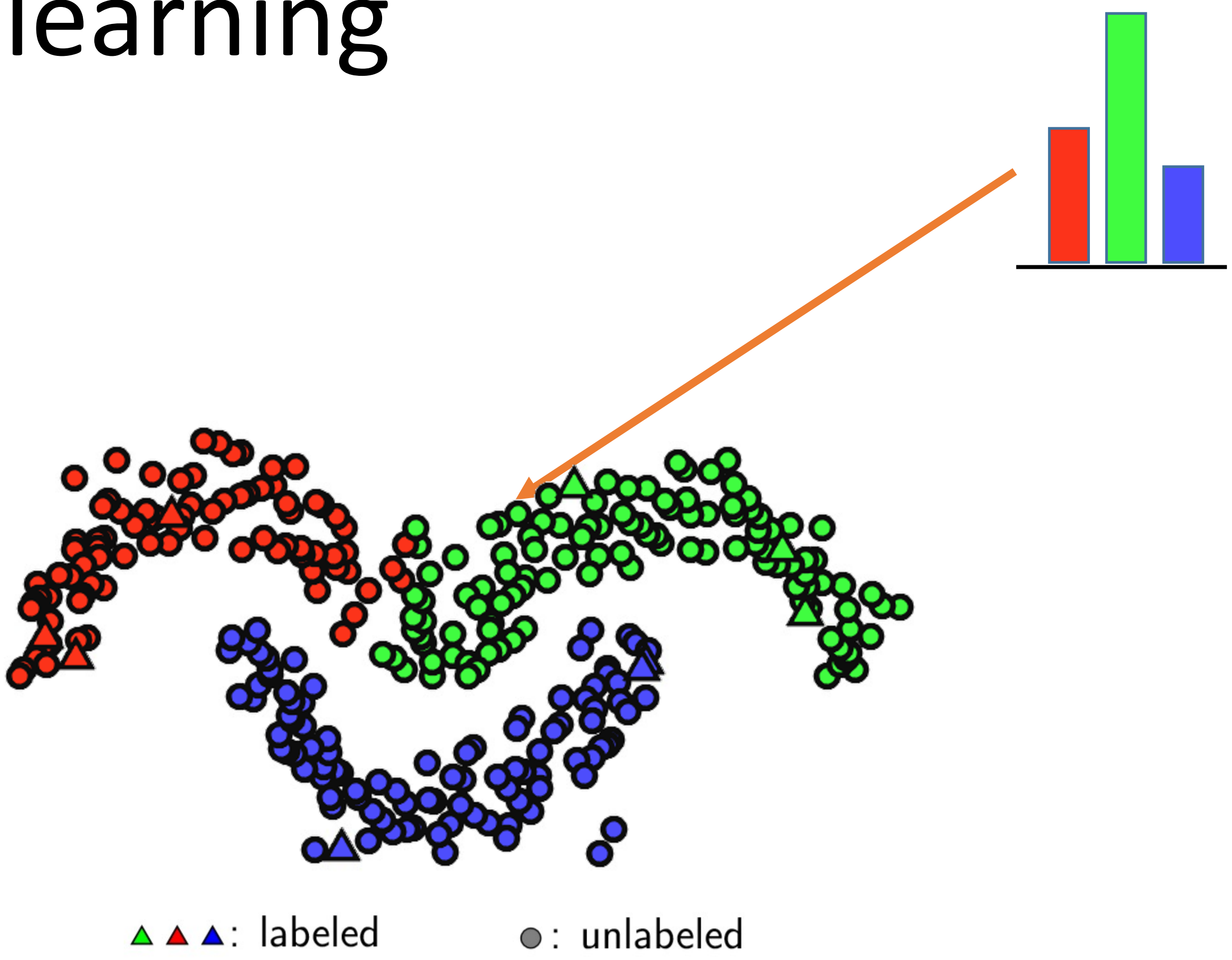


semi-supervised deep learning for classification

[Iscen, **Tolias**, Avrithis, Chum. CVPR’19. Label Propagation for Deep Semi-supervised Learning]

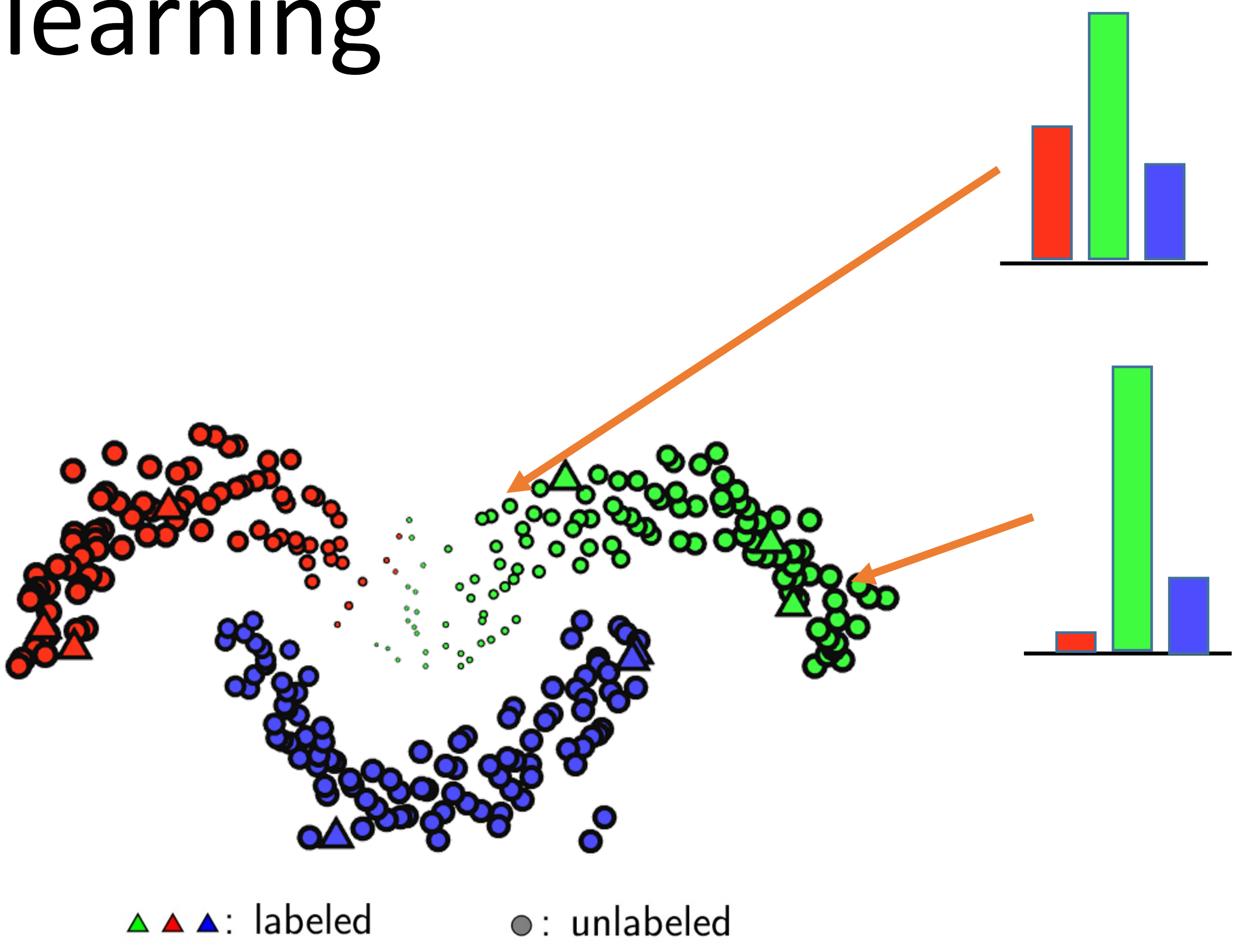
Semi-supervised deep learning

1. fully supervised training
    - on labeled examples
    - cross-entropy loss
2. use trained network
    - extract features for all examples
    - nearest neighbor graph
3. perform label propagation
4. obtain vectors of class confidence
5. assign pseudo-label (arg-max)



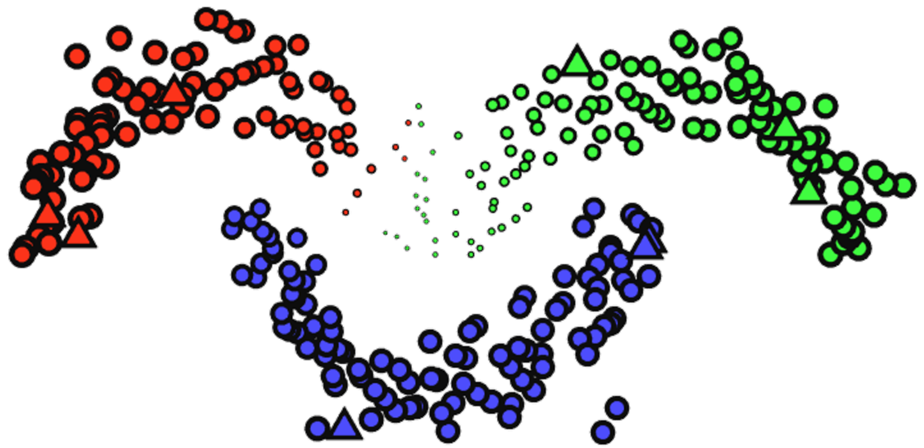
Semi-supervised deep learning

1. fully supervised training
    - on labeled examples
    - cross-entropy loss
2. use trained network
    - extract features for all examples
    - nearest neighbor graph
3. perform label propagation
4. obtain vectors of class confidence
5. assign pseudo-label (arg-max)
6. estimate certainty of the prediction
7. train with labels + pseudo-labels
    - weighted cross-entropy loss
    - certainty → example weight

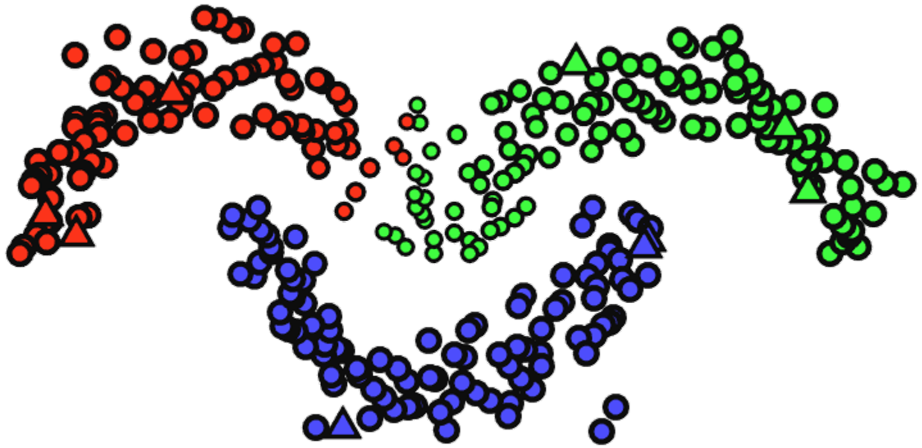


8. go to step-2

Example weights

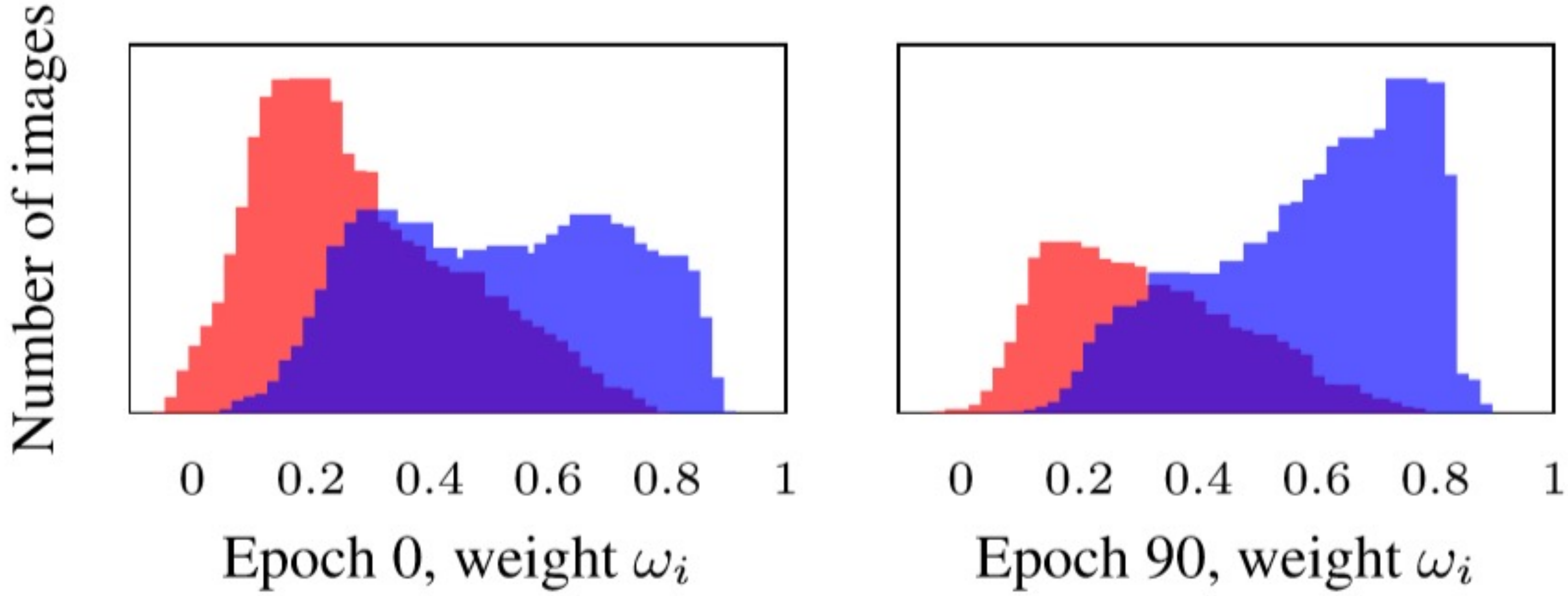


certainty = 1 – normalized entropy



maximum confidence value

Prediction certainty as weight on CIFAR10



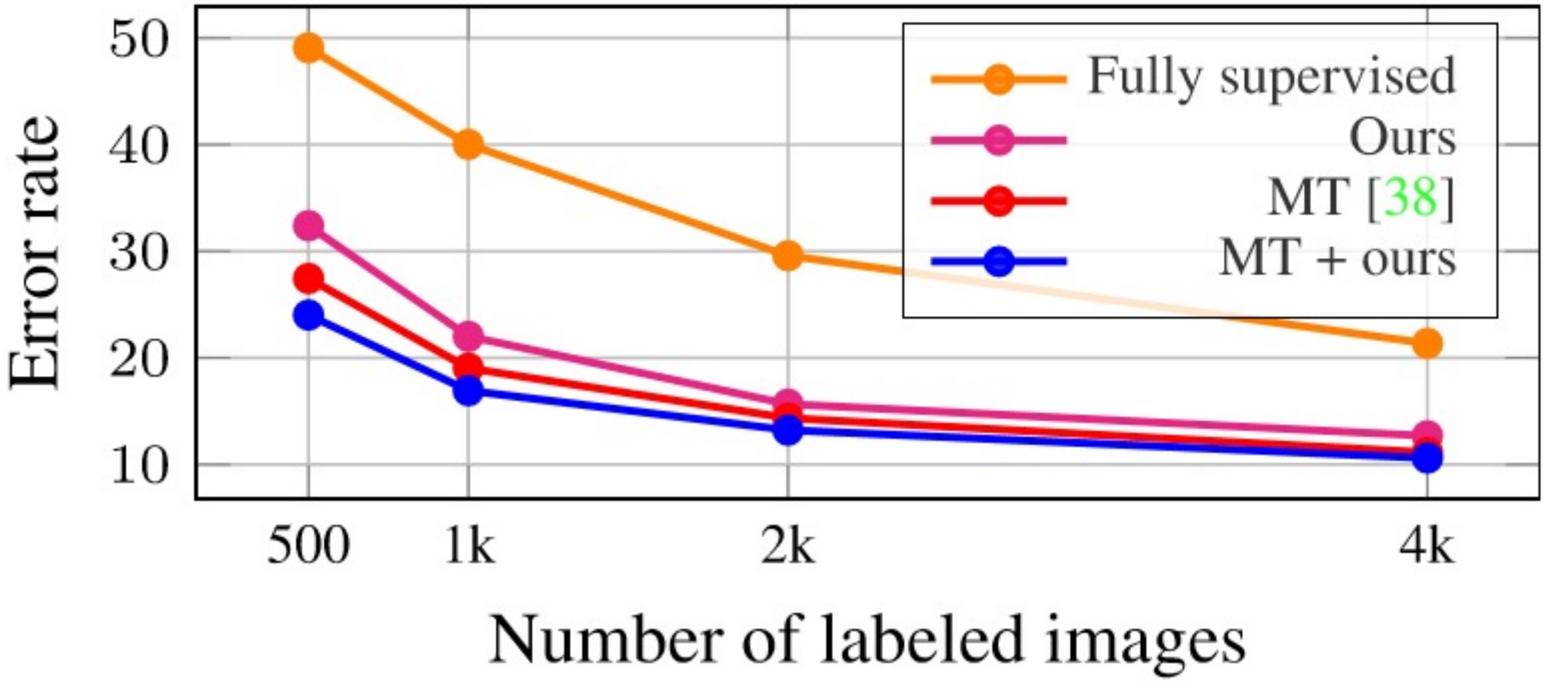
correctly and incorrectly pseudo-labeled

Ablation experiment on CIFAR10

Pseudo-labeling	example weight ω_i	class balancing ζ_j	CIFAR-10
LP-based			36.53 ± 1.42
		✓	36.17 ± 1.98
	✓		33.32 ± 1.53
	✓	✓	$\mathbf{32.40 \pm 1.80}$
Network-based	✓	✓	35.17 ± 2.46

Network-based pseudo-labeling in [Shi et al. ECCV'18. Transductive semi-supervised deep learning using min-max features.]

Number of labeled images



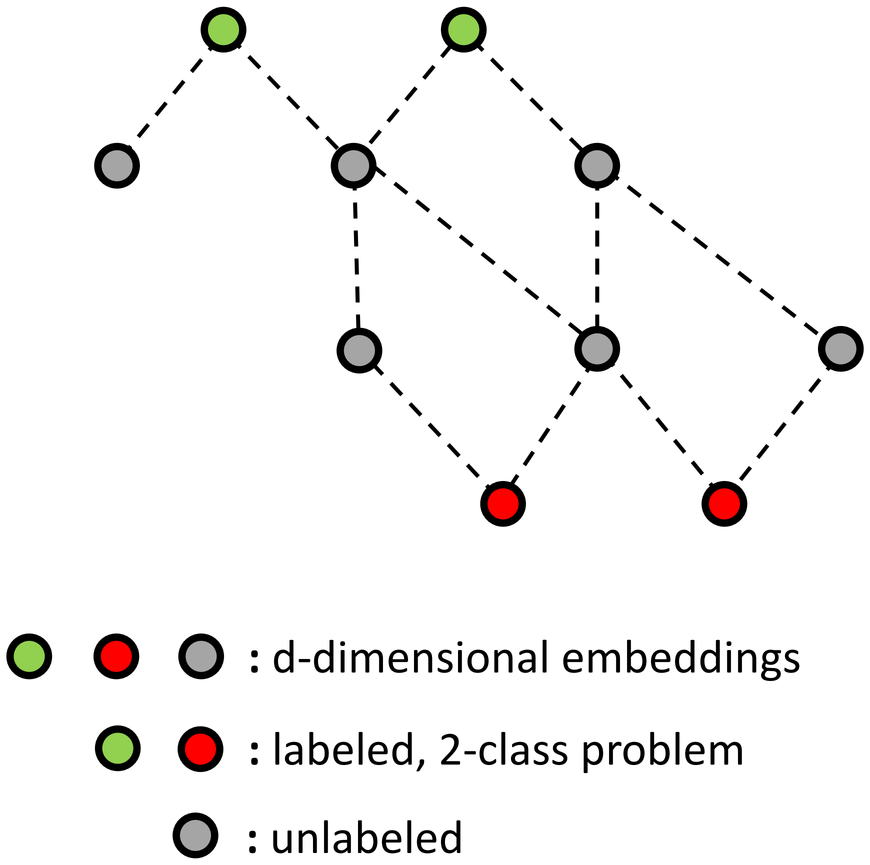
Mean Teacher (MT): [Tarvainen et al. NeurIPS'17. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results]

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

ECCV'20

Iscen, **Tolias**, Avrithis, Chum, Schmid

From MLP to GCN



Inductive approach

- train Multi-Layer Perceptron (MLP) on labeled
- pseudo-labeling by prediction on unlabeled

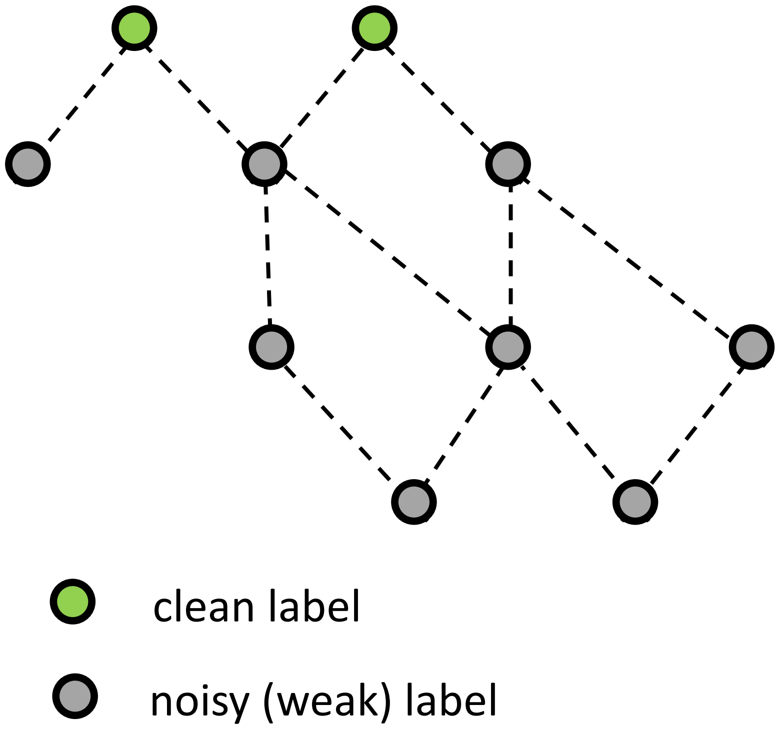
$$Z_o = f_\Theta(Z_i) = h(\Theta^\top Z_i)$$

Transductive approach

- construct affinity matrix
- train Graph Convolutional Network (GCN)
  - loss on labeled
- pseudo-labeling by prediction on unlabeled

$$Z_o = f_\Theta(Z_i) = h(\Theta^\top Z_i A)$$

Graph Convolutional Networks for noisy data cleaning



GCN binary classifier

- 2 layers: d → m → 1
- non-linearity: 1st: ReLU, 2nd: sigmoid

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

- binary cross-entropy loss
  - targets output 1 for clean and 0 for noisy
  - noise examples with importance weight λ

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

- relevance prediction for noisy examples
  - GCN scalar output after training

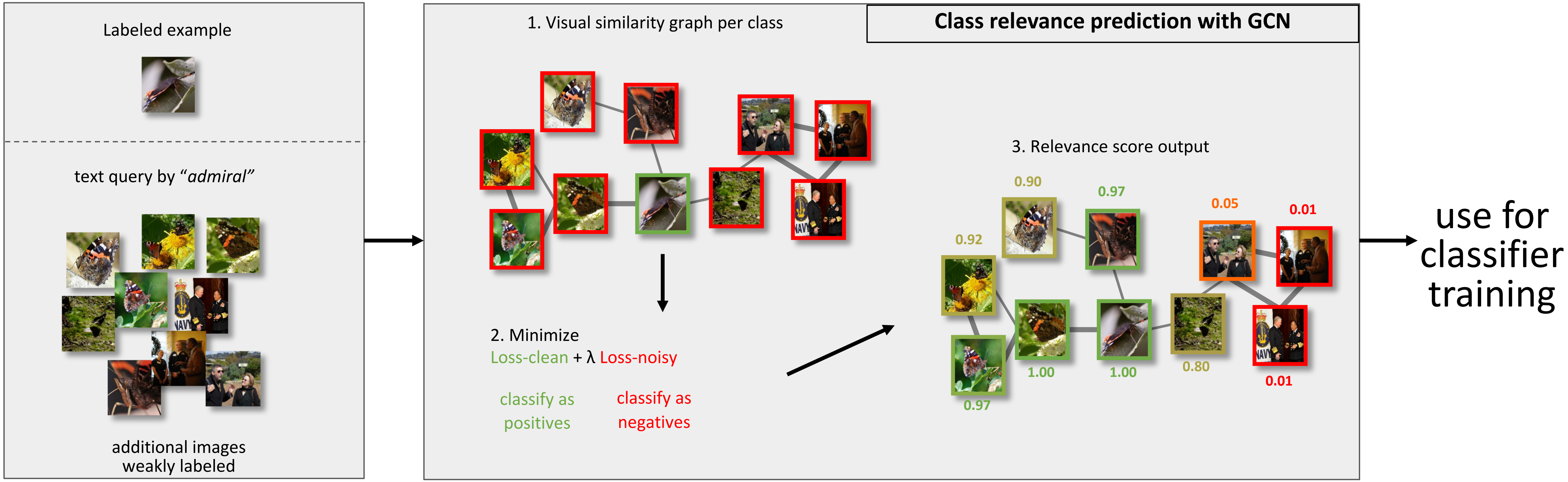
Task description

- Extension of standard few-shot learning
  - available feature extractor
  - k labeled examples per class
  - **class names are known**
  - **textual web crawling: retrieve noisy images**
  - **learn classifier with clean and noisy examples**

Our approach

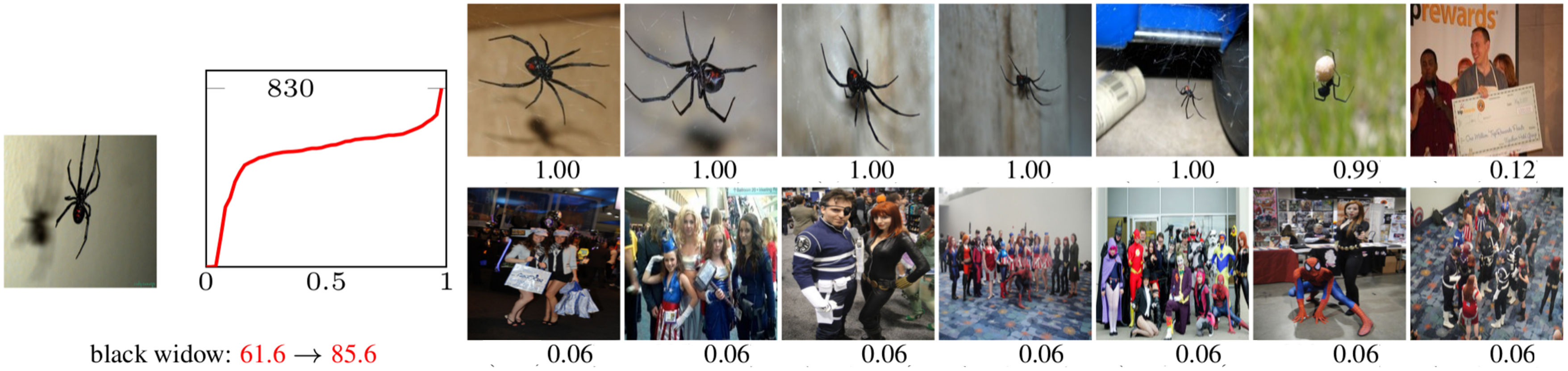
- appropriate use of noisy examples: GCN-based cleaning

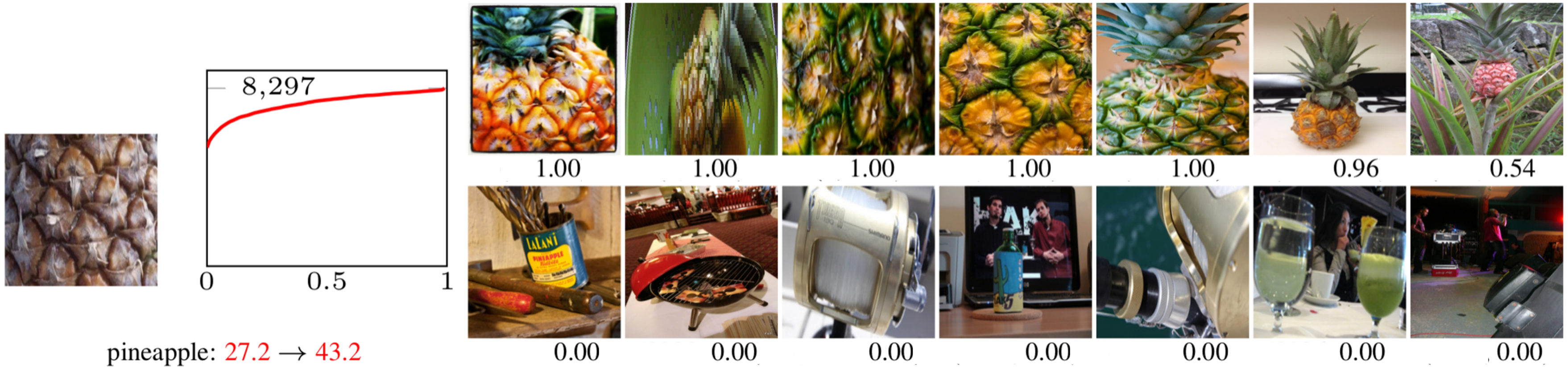
Overview



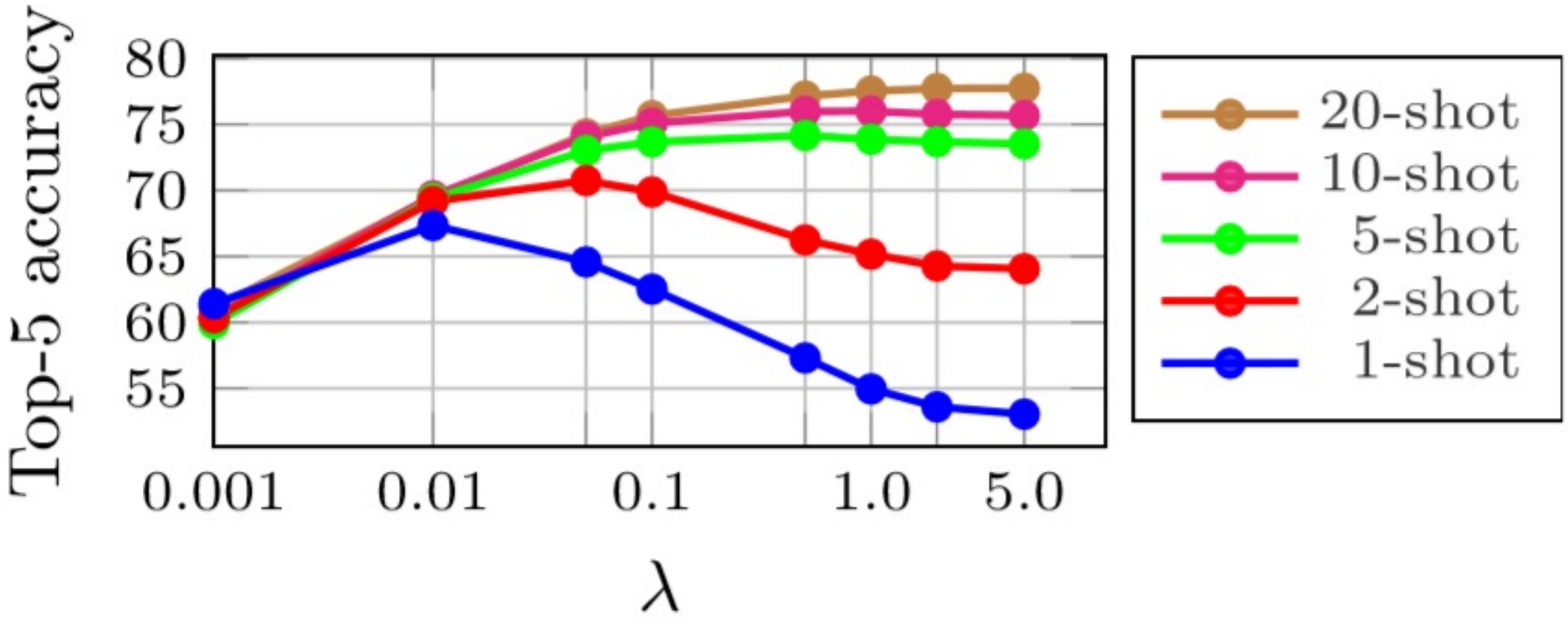
cleaning is performed separately per class

Examples of relevance scores





Importance of negative examples



- Low-shot learning benchmark based on ImageNet [Hariharan & Girshick]
- Extended with noisy images from YFCC100M (free form user tags)

Results on extended low-shot Imagenet

Method	k=1	2	5	
		Few Clean Examples		
Class proto. [9]	45.3±0.65	57.1±0.37	69.3±0.32	without noisy
		Few Clean & Many Noisy Examples		
Similarity	49.8±0.29	56.3±0.27	64.2±0.32	
β-weighting, $\beta = 1$	56.1±0.06	56.4±0.08	57.1±0.05	noisy without cleaning
β-weighting, β^*	55.6±0.24	58.3±0.14	63.4±0.25	
Linear	59.8±0.00	59.3±0.00	58.4±0.00	
Label Propagation	62.6±0.35	67.0±0.41	74.6±0.30	
MLP	63.6±0.41	68.8±0.42	73.7±0.25	cleaning with MLP
Ours	67.8±0.10	70.9±0.30	73.9±0.17	cleaning with GCN

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

Ahmet Iscen

@google

Yannis Avrithis

@inria

Ondřej Chum

@ctu

Cordelia Schmid

@google

co-authors

- Mining on manifolds: https://github.com/gtolias/mom
- LP for semi-supervised deep learning: https://github.com/ahmetius/LP-DeepSSL/
- GCN for noisy data cleaning: https://github.com/google-research/noisy-fewshot-learning