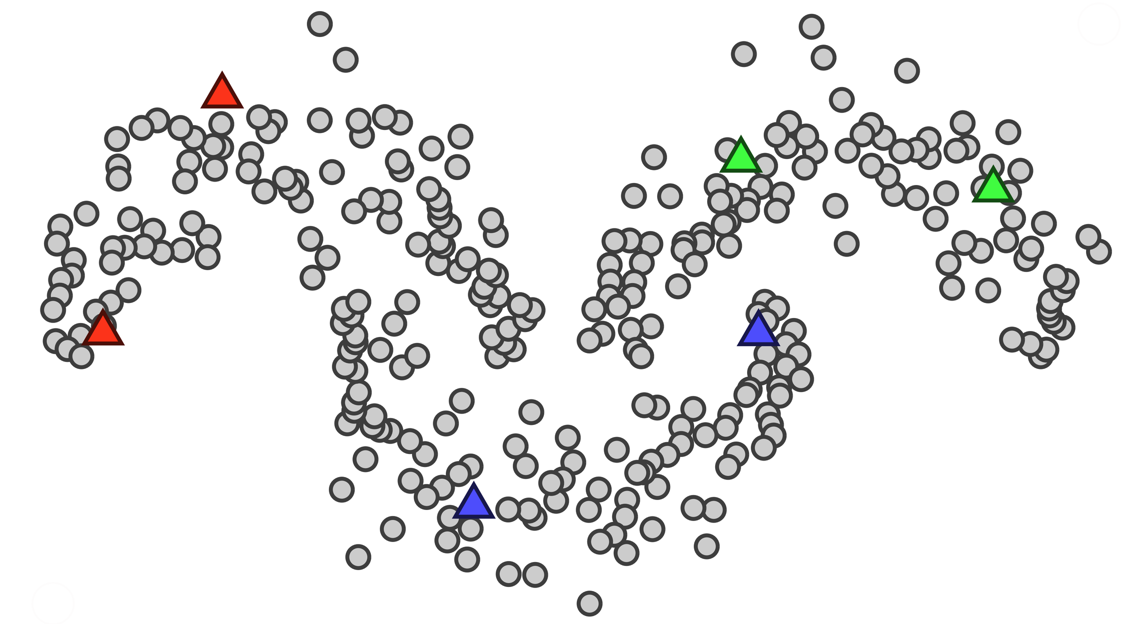
Label Propagation for Deep Semi-supervised Learning

Ahmet Iscen[1], Giorgos Tolias[1], Yannis Avrithis[2], Ondřej Chum[1]
[1]Visual Recognition Group, CTU in Prague, [2]Inria

Overview

- Labeled examples L (▲, ▲, ▲), unlabeled examples U (●)



- Use transductive learning [1] and transfer the result to deep network training with inductive setup
- Complementary to state-of-the-art approaches with consistency loss, *e.g.* Mean-Teacher [2]

Label propagation (transductive)

- Extract descriptors with a given network
- Construct normalized affinity matrix $\mathcal{W}$ (nearest neighbor graph)
- Label matrix Y with elements:

$$Y_{ij} := \begin{cases} 1, \text{ if } i \in L \wedge y_i = j \\ 0, \text{ otherwise} \end{cases}$$

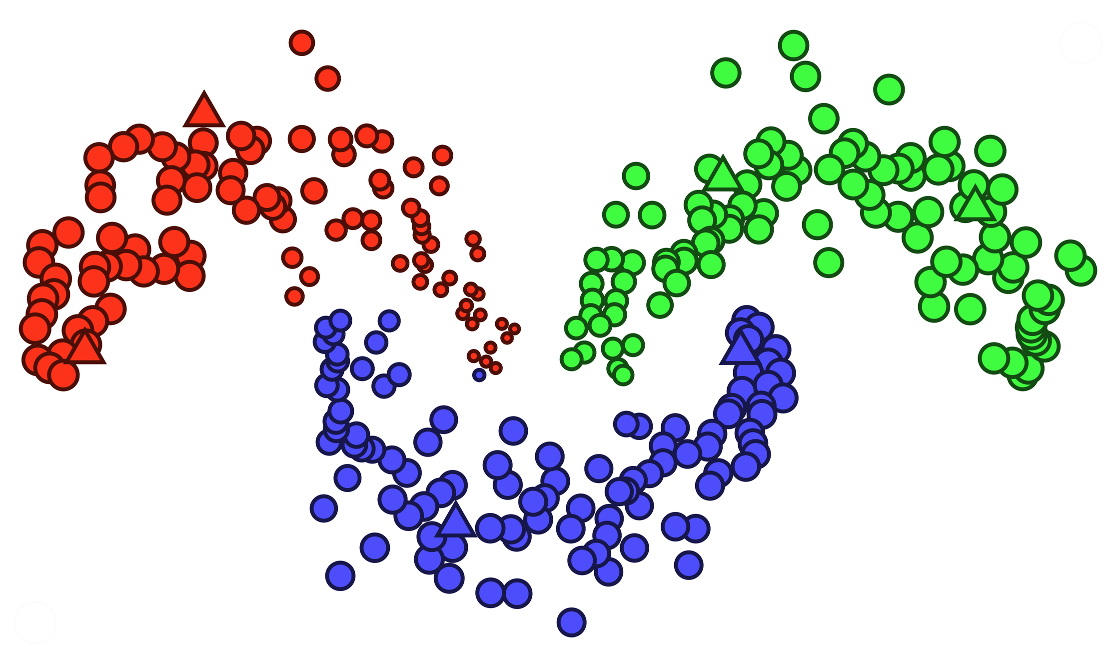
- Label propagation [1] by solving linear system (unknown Z):

$$(I - \alpha\mathcal{W})Z = Y$$

$Z^{(t)} = \alpha\mathcal{W}Z^{(t-1)} + (1-\alpha)Y$ converges to solution Z

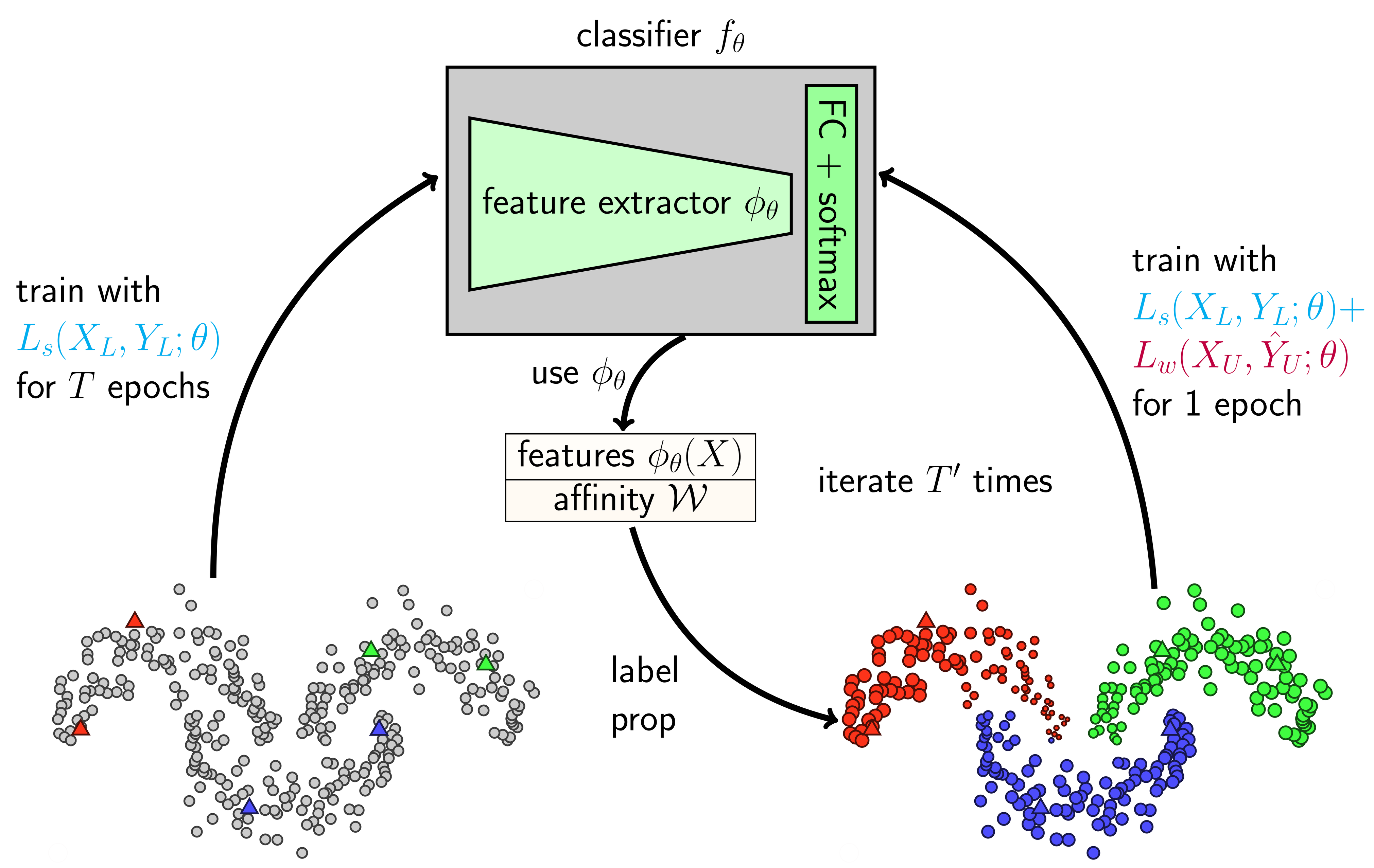
- Prediction for unlabeled examples:

$$\hat{y}_i := \arg\max_j z_{ij},$$



- Pseudo-labels (●, ●, ●), bigger circle → higher certainty

Label propagation (inductive)



▲ ▲ ▲: labeled ●: unlabeled ● ● ●: pseudo-labeled ● ● ●: certainty

Loss function

- Weighted cross-entropy loss ℓ_{CE} with labeled and unlabeled examples:

$$L = L_s(X_L, Y_L; \theta) + L_w(X_U, \hat{Y}_U; \theta)$$

$$= \sum_{i \in L} \zeta_{y_i} \ell_{\text{CE}}\left(f_\theta(x_i), y_i\right) + \sum_{i \in U} \omega_i \zeta_{\hat{y}_i} \ell_{\text{CE}}\left(f_\theta(x_i), \hat{y}_i\right)$$

- Weight ω_i is the entropy-based certainty of the pseudo-label prediction for example x_i :

$$\omega_i := 1 - \frac{H(\hat{\mathbf{z}}_i)}{\log(c)}$$

- ζ_j is the class balancing weight for class j:

$$\zeta_j := (|L_j| + |U_j|)^{-1}$$

$$L_j = \{i \in L \wedge y_i = j\} \text{ and } U_j = \{i \in U \wedge \hat{y}_i = j\}$$

References

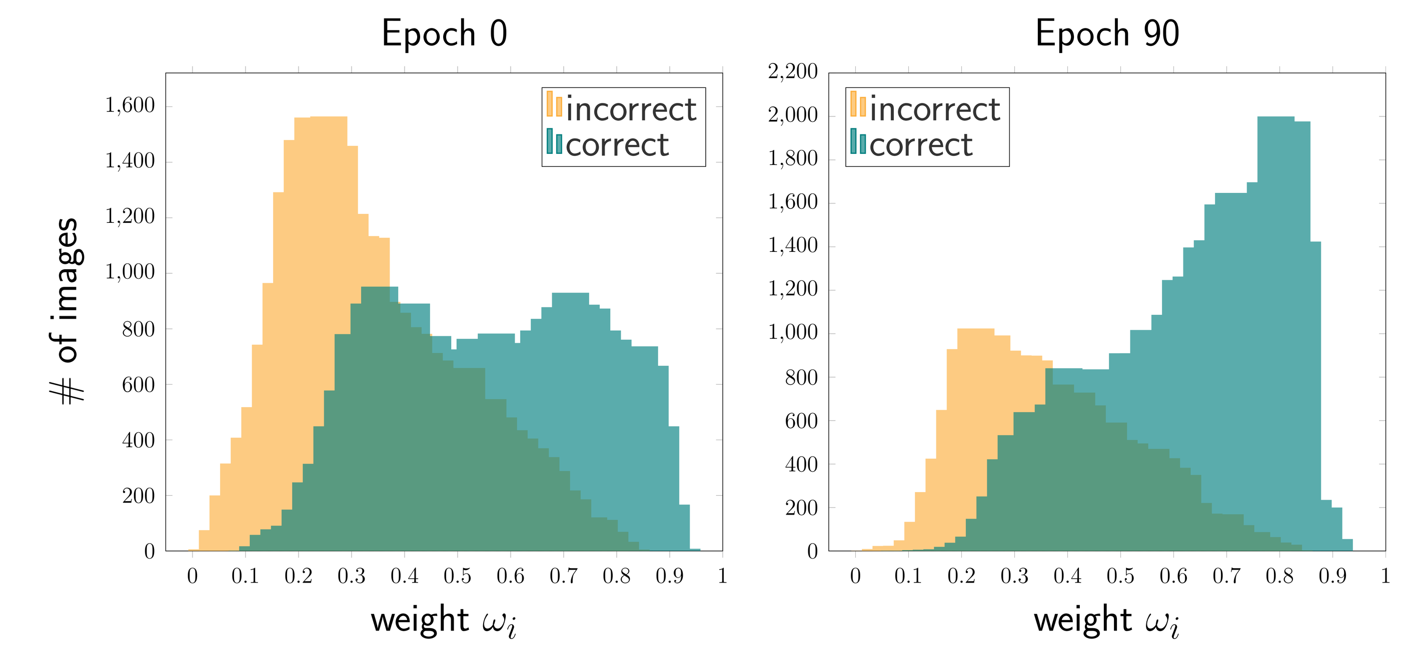
[1] Dengyong Zhou, Olivier Bousquet, Thomas Navin Lal, Jason Weston, and Bernhard Schölkopf. Learning with local and global consistency. In *NIPS*, 2003.

[2] Antti Tarvainen and Harri Valpola. Mean teachers are better role models: Weight-averaged consistency targets improve semi-supervised deep learning results. In *NIPS*, 2017.

[3] Weiwei Shi, Yihong Gong, Chris Ding, Zhiheng Ma, Xiaoyu Tao, and Nanning Zheng. Transductive semi-supervised deep learning using min-max features. In *ECCV*, 2018.

Certainty weight distribution

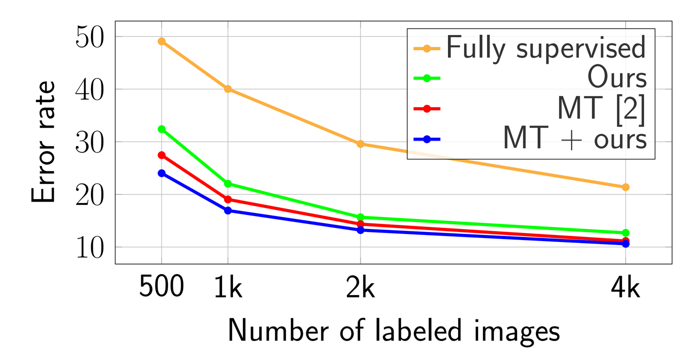
- Distribution of weights ω_i in CIFAR-10



Experiments

- Error rate is reported (lower is better)
- "13-layer" network for CIFAR-10 and CIFAR-100
- Resnet-18 for Mini-ImageNet

Pseudo-labeling	ω_i	ζ_j	CIFAR-10
Label propagation			36.53 ± 1.42
		✓	36.17 ± 1.98
	✓		33.32 ± 1.53
	✓	✓	$\mathbf{32.40 \pm 1.80}$
Network	✓	✓	35.17 ± 2.46



Comparison with state of the art:

Dataset	CIFAR-10			
Nb. labeled images	500	1000	2000	4000
Fully supervised	49.08 ± 0.83	40.03 ± 1.11	29.58 ± 0.93	21.63 ± 0.38
TDCNN [3]†	-	32.67 ± 1.93	22.99 ± 0.79	16.17 ± 0.37
Network prediction	35.17 ± 2.46	23.79 ± 1.31	16.64 ± 0.48	13.21 ± 0.61
Ours	32.40 ± 1.80	22.02 ± 0.88	15.66 ± 0.35	12.69 ± 0.29
VAT †	-	-	-	11.36
Π model †	-	-	-	12.36 ± 0.31
Temporal Ensemble †	-	-	-	12.16 ± 0.24
MT [2]†	-	27.36 ± 1.30	15.73 ± 0.31	12.31 ± 0.28
MT [2]	27.45 ± 2.64	19.04 ± 0.51	14.35 ± 0.31	11.41 ± 0.25
MT + Ours	$\mathbf{24.02 \pm 2.44}$	$\mathbf{16.93 \pm 0.70}$	$\mathbf{13.22 \pm 0.29}$	$\mathbf{10.61 \pm 0.28}$

† denotes scores reported in prior work

Dataset	CIFAR-100		Mini-ImageNet-*top5*	
Nb. labeled images	4000	10000	4000	10000
Fully supervised	55.43 ± 0.11	40.67 ± 0.49	53.07 ± 0.68	38.28 ± 0.38
Ours	46.20 ± 0.76	38.43 ± 1.88	$\mathbf{47.58 \pm 0.94}$	36.14 ± 2.19
MT [2]	45.36 ± 0.49	36.08 ± 0.51	49.35 ± 0.22	32.51 ± 1.31
MT + Ours	$\mathbf{43.73 \pm 0.20}$	$\mathbf{35.92 \pm 0.47}$	50.52 ± 0.39	$\mathbf{31.99 \pm 0.55}$

Contact: ahmet.iscen@cmp.felk.cvut.cz