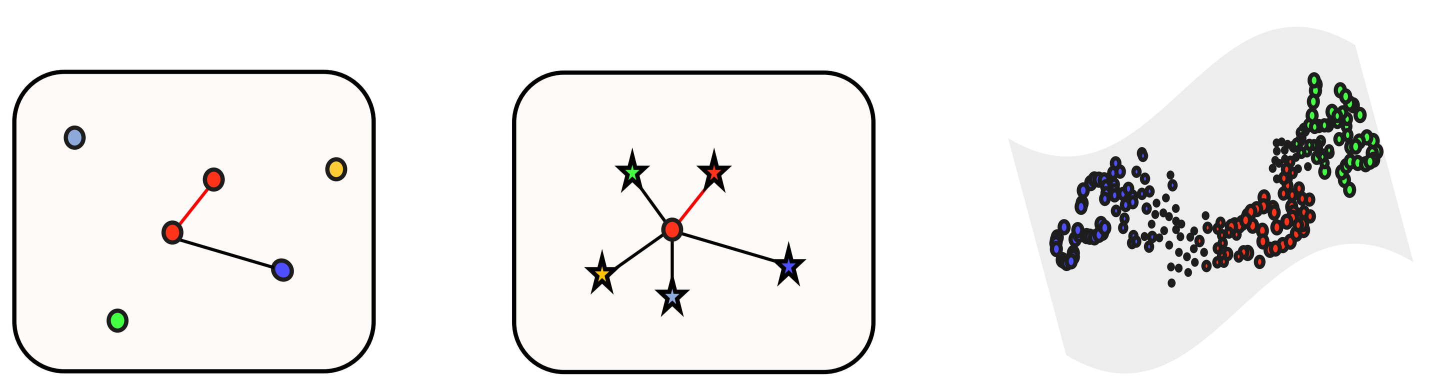
Fewer is More: A Deep Graph Metric Learning Perspective Using Fewer Proxies

Yuehua Zhu[1], Muli Yang[1], Cheng Deng[1], Wei Liu[2]
[1] Xidian University, [2] Tencent AI Lab

Paper Code

Overview

- **Goal of Deep Metric Learning** is to learn an embedding space, where the embedded vectors of similar samples are close, while those of dissimilar ones are far away from each other
- **Applied tasks:** image retrieval and clustering
- **Related work**



Sampling-based Triplet | Proxy-based Proxy-NCA | Traditional Label Propagation

- **Motivation**

◆ **Sampling-based methods** select hard samples from a subset (mini-batch) of the whole training data set, which **fail to** characterize the global geometry of the embedding space precisely

◆ **Proxy-based methods** equally treat each raw data point by calculating with either all reference points or class-specific parameters in classification layers, hence **failing to** capture the most discriminative relationships among raw data points. In addition, what follows is **expensive computational consumption** when many classes are involved

◆ **Traditional label propagation** is **good at** capturing overall neighborhood structure and possibly underlying manifold structure, iteratively determining the unknown labels of samples according to appropriate **graph structures**

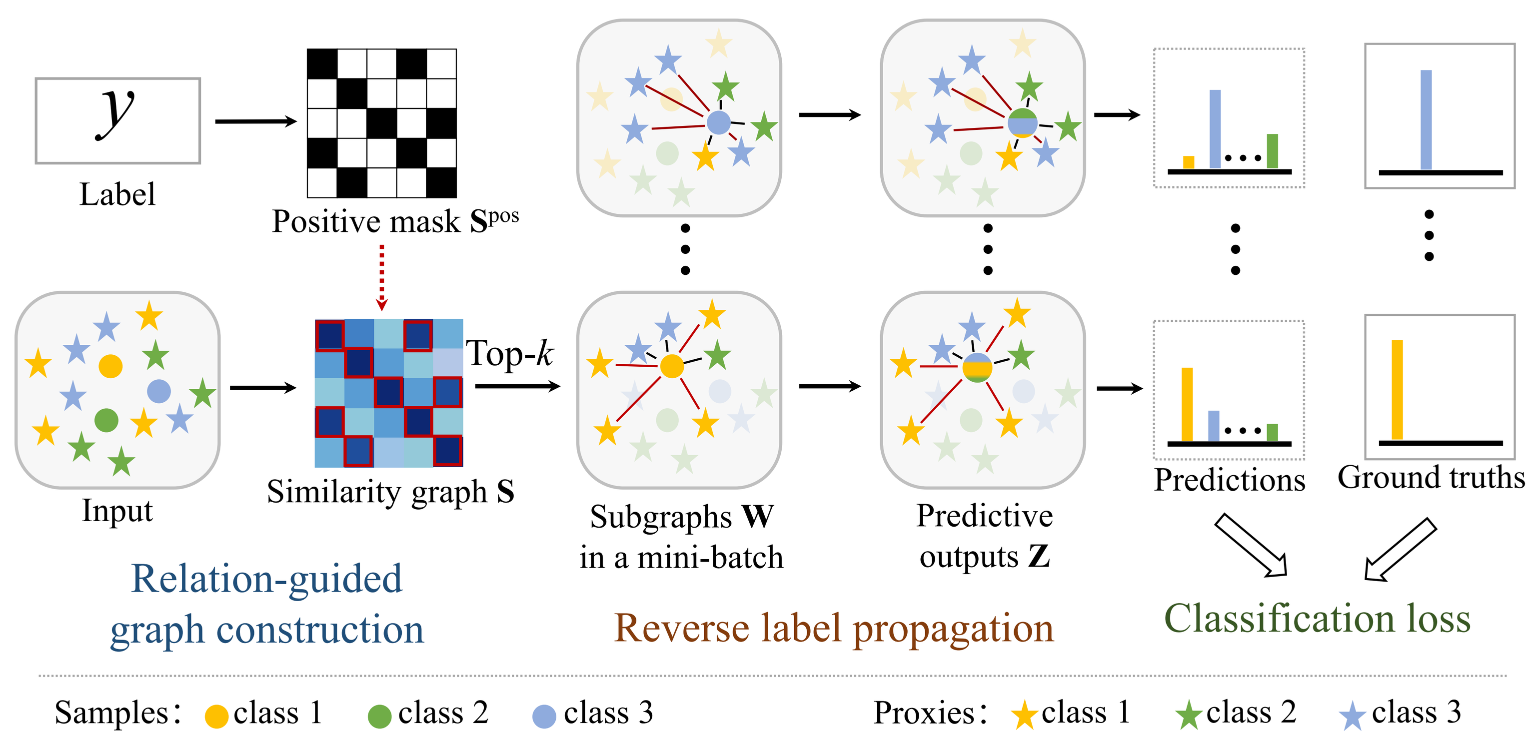
1) Multiple global proxies are leveraged to **collectively approximate** the original data points for each class to better capture intra-class variations

2) To efficiently capture local neighbor relationships, **a small number of** such proxies are **adaptively** selected to construct similarity subgraphs between these proxies and each data point

3) We design a novel **reverse label propagation algorithm**, by which the neighbor relationships are adjusted according to ground-truth labels, so that a discriminative metric space can be learned during the process of **subgraph classification**

Method

- **The pipeline of our approach**



- **Relation-Guided Graph Construction**

◆ Constructing similarity graphs $\mathbf{S}$ between proxies and samples $\mathbf{S}_{ij}=(\mathbf{x}_i^s)^\top \mathbf{x}_j^p$

◆ **Positive mask** can be regarded as a **"soft" constraint** on proxies, which makes similar proxies mutually close by encouraging proxies to be close to their relevant samples

$$\mathbf{S}_{ij}^{\mathrm{pos}}=\begin{cases}1, & \text{if} \quad y_i^s=y_j^p\\ 0, & \text{else}\end{cases}$$

◆ Under the guidance of positive mask, we calculate and store the indexes of *k*-max values in each row of ($\mathbf{S}$ + $\mathbf{S}^{\mathrm{pos}}$) into a *k*-element set $\mathcal{I}=\{(i,j),\cdots\}$

◆ ***k*-NN subgraphs** are constructed and represented by a sparse neighbor matrix $\mathbf{W}$

$$\mathbf{W}_{ij}=\begin{cases}\mathbf{S}_{ij}, & \text{if} \quad (i,j)\in\mathcal{I}\\ 0, & \text{else}\end{cases}$$

- **Reverse Label Propagation**

◆ With proposed **reverse label propagation**, all subgraphs $\mathbf{W}$ are encoded into predictive outputs $\mathbf{Z}$ $\mathbf{Z}=\mathbf{W}\mathbf{Y}^p$

- **Classification Loss**

◆ We propose an improved **mask softmax** function to precisely encode the subgraph predictions

$$P(\tilde{y}_i^s=j\mid\mathbf{x}_i^s)=\frac{\mathbf{M}_{ij}\exp(\mathbf{Z}_{ij})}{\sum_{j'=1}^{C}\mathbf{M}_{ij'}\exp(\mathbf{Z}_{ij'})}$$

◆ Main classification loss on raw samples

$$\mathcal{L}^s=-\frac{1}{M}\sum_{i=1}^{M}\sum_{j=1}^{C}\mathbb{I}(y_i^s=j)\log\left(P(\tilde{y}_i^s=j\mid\mathbf{x}_i^s)\right)$$

◆ **Regularizer** can be regarded as a **"hard" constraint** on the proxies to ensure that similar proxies are close to each other while dissimilar ones are far apart from each other

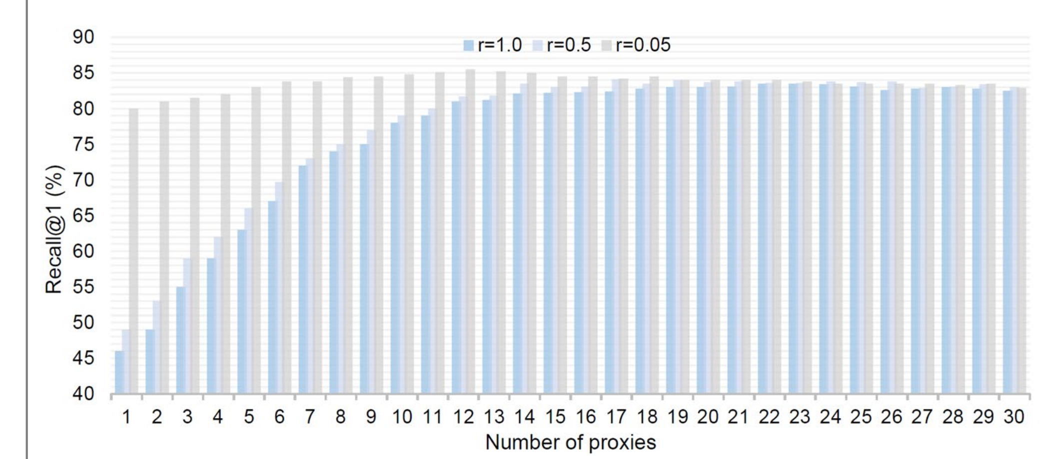
$$\mathcal{L}^p=-\frac{1}{C\times N}\sum_{i=1}^{C\times N}\sum_{j=1}^{C}\mathbb{I}(y_i^p=j)\log\left(P(\tilde{y}_i^p=j\mid\mathbf{x}_i^p)\right)$$

◆ An end-to-end training by minimizing our **ultimate objective** yields discriminative embeddings and the most informative proxies $\mathcal{L}(\Theta,\mathcal{P}):=\mathcal{L}^s+\lambda\mathcal{L}^p$

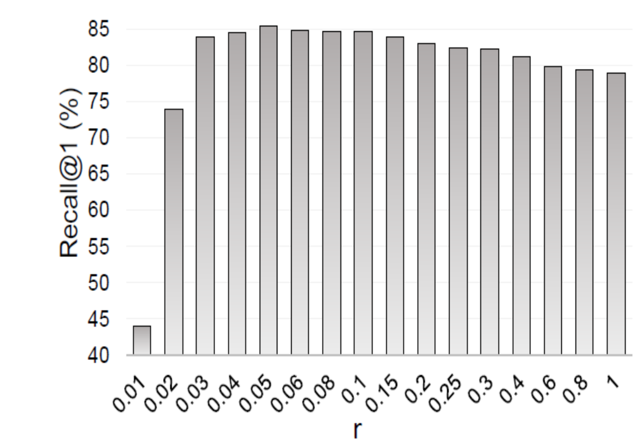
Experiments

- **Ablation Study and Training Curve**

Parameter study of *N* under three different *r* (*N* means the number of proxies assigned to each class)



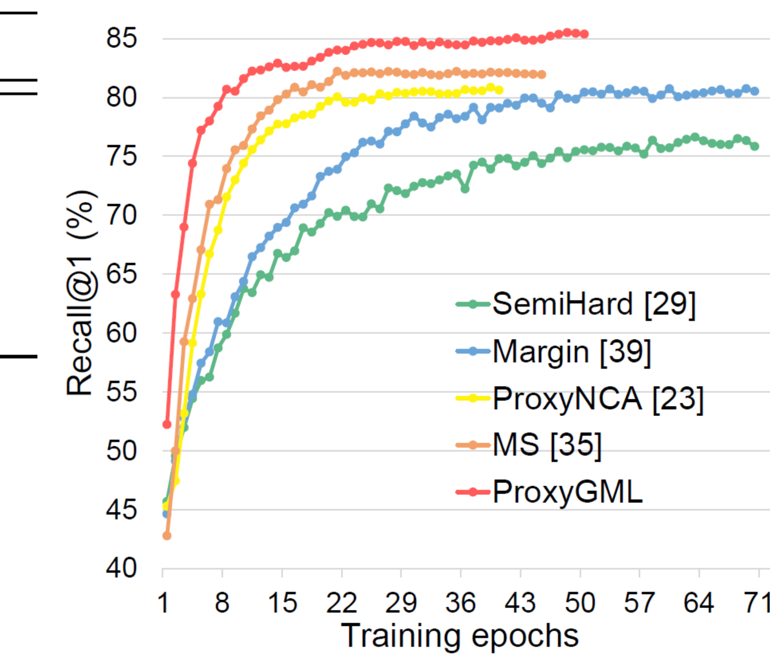
Parameter study of *r* ($r\in(0,1]$ implies the scale of the graph)



Ablation study of three proposed modules

#	$\mathbf{S}^{\mathrm{pos}}$	$\mathbf{M}$	$\mathcal{L}^p$	NMI	R@1
1	×	×	×	52.1	47.3
2	✓	×	×	69.6	83.3
3	×	✓	×	54.9	66.1
4	×	×	✓	67.1	81.7
5	×	✓	✓	68.8	82.6
6	✓	×	✓	71.6	84.5
7	✓	✓	×	70.7	84.0
8	✓	✓	✓	72.4	85.5

Training convergence curve of image retrieval task on the Cars196 dataset



- **Comparisons**

Comparison with SOTAs

Methods		CUB-200-2011 NMI	CUB-200-2011 R@1	CUB-200-2011 R@2	CUB-200-2011 R@4	Cars196 NMI	Cars196 R@1	Cars196 R@2	Cars196 R@4	Stanford Online Products NMI	Stanford Online Products R@1	Stanford Online Products R@10	Stanford Online Products R@100
SemiHard^{64} [24]	BN	55.4	42.6	55.0	66.4	53.4	51.5	63.8	73.5	89.5	66.7	82.4	91.9
Clustering^{64} [19]	BN	59.2	48.2	61.4	71.8	59.0	58.1	70.6	80.3	89.5	67.0	83.7	93.2
LiftedStruct^{64} [20]	G	56.6	43.6	56.6	68.6	56.9	53.0	65.7	76.0	88.7	62.5	80.8	91.9
ProxyNCA^{64} [18]	BN	59.5	49.2	61.9	67.9	64.9	73.2	82.4	86.4	90.6	73.7	–	–
HDC^{384} [39]	G	–	53.6	65.7	77.0	–	73.7	83.2	89.5	–	69.5	84.4	92.8
HTL^{512} [6]	BN	–	57.1	68.8	78.7	–	81.4	88.0	92.7	–	74.8	88.3	94.8
DAMLRRM^{512} [34]	G	61.7	55.1	66.5	76.8	64.2	73.5	82.6	89.1	88.2	69.7	85.2	93.2
HDML^{512} [40]	G	62.6	53.7	65.7	76.7	69.7	79.1	87.1	92.1	89.3	68.7	83.2	92.4
SoftTriple^{512} [21]	BN	69.3	65.4	76.4	84.5	70.1	84.5	90.7	94.5	**92.0**	**78.3**	90.3	95.9
MS^{512} [29]	BN	–	65.7	77.0	86.3	–	84.1	90.4	94.0	–	78.2	90.5	96.0
ProxyGML^{64}	BN	65.1	59.4	70.1	80.4	67.9	78.9	87.5	91.9	89.8	76.2	89.4	95.4
ProxyGML^{384}	BN	68.4	65.2	76.4	84.3	70.9	84.5	90.4	94.5	90.1	77.9	90.0	96.0
ProxyGML^{512}	BN	**69.8**	**66.6**	**77.6**	**86.4**	**72.4**	**85.5**	**91.8**	**95.3**	90.2	78.0	**90.6**	**96.2**