南京航空航天大學
Nanjing University of Aeronautics and Astronautics

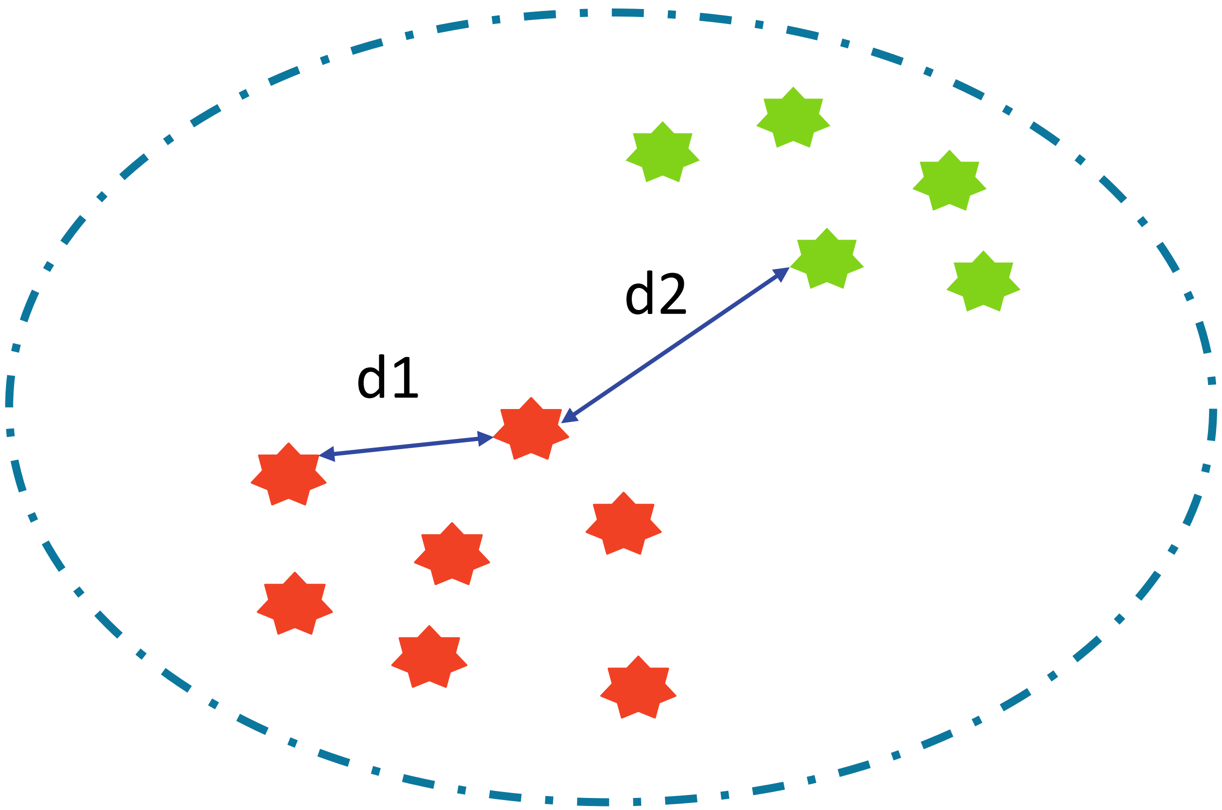
Noise-resistant Deep Metric Learning with Ranking-based Instance Selection

Chang Liu[1], **Han Yu**[1], **Boyang Li**[1*], **Zhiqi Shen**[1*], **Zhanning Gao**[2*],
Peiran Ren[2], **Xuansong Xie**[2], **Lizhen Cui**[3,4], **Chunyan Miao**[1*]
[1]School of Computer Science and Engineering, Nanyang Technological University (NTU), Singapore
[2]Alibaba Group, Hangzhou, China [3]School of Software, Shandong University (SDU), Jinan, China
[4]Joint SDU-NTU Centre for Artificial Intelligence Research (C-FAIR), SDU, Jinan, China
*{boyang.li, zqshen}@ntu.edu.sg, zhanning.gzn@alibaba-inc.com, ascymiao@ntu.edu.sg

CVPR 2021

Deep Metric Learning

Cosine similarity: $$S(f(x_i), f(x_j)) = \frac{f(x_i)^T f(x_j)}{\|f(x_i)\| \|f(x_j)\|}. \quad (1)$$



Memory Bank

$$\mathcal{M} = \{(v_0, y_0), (v_1, y_1), \ldots, (v_M, y_M)\}$$

First in first out
Memory bank:

Feature1, label: 3
Feature2, label: 1
Feature3, label: 2
Feature4, label: 3
Feature5, label: 2
Feature6, label: 2
Feature7, label: 4

Capacity: M

Identifying Noisy Labels

$$P_{\text{clean}}(i) = \frac{\exp\left(T(x_i, y_i)\right)}{\sum_{k \in C} \exp\left(T(x_i, k)\right)} \tag{2}$$

$$T(x_i, k) = \frac{1}{M_k} \sum_{(v_j, y_j) \in \mathcal{M}, y_j = k} S(f(x_i), v_j) \tag{3}$$

M_k : The number of samples in class k in the memory bank

$T(x_i, k)$: The average similarity between x_i and all the stored features v_j in class k

Identifying Noisy Labels

$P_{\text{clean}}(i)$ > m: clean sample

Feature i, label: y(i)
⋮
⋮
⋮
⋮
⋮
⋮

$$m = \frac{1}{\tau} \sum_{j=t-\tau}^{t} Q_j$$

Accelerating PRISM

$$\sum_{\substack{(v_j,y_j)\in\mathcal{M}\\ y_j=k}} \frac{S(f(x_i),v_j)}{M_k} = \left(\frac{1}{M_k}\sum_{\substack{(v_j,y_j)\in\mathcal{M}\\ y_j=k}} \frac{v_j}{\|v_j\|}\right)\frac{f(x_i)}{\|f(x_i)\|}$$
$$= w_k \frac{f(x_i)}{\|f(x_i)\|}, \tag{5}$$

$$w_k = \frac{1}{M_k}\sum_{(v_j,y_j)\in\mathcal{M},y_j=k} \frac{v_j}{\|v_j\|}. \tag{6}$$

Plugging Eq. (5) into Eq. (2), $P_{\text{clean}}(i)$ can be expressed as:

$$\boxed{P_{\text{clean}}(i) = \exp\left(w_{y_i}\frac{f(x_i)}{\|f(x_i)\|}\right) / \sum_{k\in C} \exp\left(w_k \frac{f(x_i)}{\|f(x_i)\|}\right). \quad (7)}$$

PRISM

Algorithm 1: A training iteration of PRISM

Input : $\mathcal{B} = \{(x_0, y_0), (x_1, y_1), ..., (x_B, y_B)\}$: a given minibatch of data with size B;
$f(\cdot)$: a given deep metric model;

Parameter: $\{w_k | k \in C\}$: the set of mean feature vectors of all classes, all initialized to zero before training commences;

for each $(x_i, y_i) \in \mathcal{B}$ **do**
  if $w_{y_i} = \vec{0}$ **then**
    $P_{\text{clean}}(i) = 1$;
  else
    Calculate $P_{\text{clean}}(i)$ according to Eq. (7);
  end
end
Calculate the threshold m using TRM or sTRM;
Initialize $\mathcal{B}_{clean}$ as an empty set;
for each $(x_i, y_i) \in \mathcal{B}$ **do**
  if $P_{clean}(i) > m$ **then**
    Add $(f(x_i), y_i)$ to $\mathcal{B}_{clean}$;
  end
end
Enqueue $\mathcal{B}_{clean}$ into the Memory Bank
for each $(v_i, y_i) \in \mathcal{B}_{clean}$ **do**
  Update w_{y_i} according to Eq. (6);
end
Calculate loss $L(\mathcal{B}_{clean})$ and update the parameters of $f(\cdot)$

Loss Function

More formally, the loss for mini-batch $\mathcal{B}$ is

$$\begin{aligned} L_{\text{batch}}(\mathcal{B}) = &\sum_{\substack{(x_i,y_i)\in B,(x_j,y_j)\in B \\ y_i\neq y_j}} \max(S(f(x_i), f(x_j)) - \lambda, 0) \\ &- \sum_{\substack{(x_i,y_i)\in B,(x_j,y_j)\in B \\ y_i=y_j}} S(f(x_i), f(x_j)) \end{aligned} \tag{8}$$

Total Loss = (8)+(9)

$$\begin{aligned} L_{\text{bank}}(\mathcal{M},\mathcal{B}) = &\sum_{\substack{(x_i,y_i)\in \mathcal{B},\,(v_j,y_j)\in \mathcal{M} \\ y_i\neq y_j}} \max(S(f(x_i), v_j) - \lambda, 0) \\ &- \sum_{\substack{(x_i,y_i)\in \mathcal{B},\,(v_j,y_j)\in \mathcal{M} \\ y_i=y_j}} S(f(x_i), v_j). \end{aligned} \tag{9}$$

Experiments

Table 1: Precision@1 (%) on CARS, SOP, and CUB dataset with symmetric label noise.

	CARS			SOP			CUB		
Noisy Label Rate	10%	20%	50%	10%	20%	50%	10%	20%	50%
Algorithms for image classification under label noise									
Co-teaching [11]	73.47	70.39	59.55	62.60	60.26	52.18	53.74	51.12	45.01
Co-teaching+ [52]	71.49	69.62	62.35	63.44	67.93	58.29	53.31	51.04	45.16
Co-teaching [11] w/ Temperature [53]	77.51	76.30	66.87	73.71	71.97	64.07	55.25	54.18	50.65
F-correction [31]	71.00	69.47	59.54	51.18	46.34	48.92	53.41	52.60	48.84
DML with Proxy-based Losses									
FastAP [6]	66.74	66.39	58.87	69.20	67.94	65.83	54.10	53.70	51.18
nSoftmax [53]	72.72	70.10	54.80	70.10	68.90	57.32	51.99	49.66	42.81
ProxyNCA [26]	69.79	70.31	61.75	71.10	69.50	61.49	47.13	46.64	41.63
Soft Triple [32]	76.18	71.82	52.53	68.60	55.21	38.45	51.94	49.14	41.46
DML with Pair-based Losses									
MS [43]	66.31	67.14	38.24	69.90	67.60	59.58	57.44	54.52	40.70
Circle [37]	71.00	56.24	15.24	72.80	70.50	41.17	47.48	45.32	12.98
Contrastive Loss [7]	72.34	70.93	22.91	68.70	68.80	61.16	51.77	51.50	38.59
Memory Contrastive Loss (MCL) [46]	74.22	69.17	46.88	79.00	76.60	67.21	56.72	50.74	31.18
MCL + PRISM (Ours)	**80.06**	**78.03**	**72.93**	**80.11**	**79.47**	**72.85**	**58.78**	**58.73**	**56.03**

Experiments

Table 2: Precision@1 (%) on CARS, SOP, and CUB with Small Cluster label noise.

	CARS		SOP		CUB	
Noisy Label Rate	25%	50%	25%	50%	25%	50%
Algorithms for image classification under label noise						
Co-teaching	70.57	62.91	61.97	58.08	51.75	48.85
Co-teaching+	70.05	61.58	62.57	59.27	51.55	47.62
Co-teaching w/ Temperature	75.26	66.19	70.19	68.50	54.59	48.32
DML with Proxy-based Losses						
FastAP	62.49	53.07	70.66	67.55	52.18	48.46
nSoftmax	71.61	62.29	70.00	61.92	49.61	41.78
ProxyNCA	69.50	58.34	67.95	62.25	42.07	36.48
Soft Triple	73.26	66.66	**73.63**	64.14	56.18	50.35
Soft Triple + PRISM (Ours)	**77.60**	**70.45**	70.99	**69.38**	**57.61**	**54.27**
DML with Pair-based Losses						
MS	63.92	43.73	67.32	62.17	53.60	41.66
Circle	53.03	19.95	70.33	40.48	44.07	22.96
Contrastive Loss	65.60	26.45	68.25	64.27	47.27	39.43
Memory Contrastive Loss (MCL)	69.46	36.43	75.61	68.71	52.25	41.58
MCL + PRISM (Ours)	**77.08**	**68.26**	**78.56**	**73.84**	**55.77**	**53.46**

Experiments

Table 3: Precision@1 (%) and Mean Average Precision@R (%) on CARS-98N and Food-101N [21].

	CARS-98N		Food-101N	
	P@1	MAP@R	P@1	MAP@R
Algorithms for image classification under label noise				
Co-teaching	58.74	9.10	59.08	14.66
Co-teaching+	56.66	8.40	57.59	14.72
Co-teaching w/ Temperature	60.72	9.61	63.18	17.38
DML with Proxy-based Losses				
ProxyNCA	53.55	8.75	48.41	9.30
Soft Triple	63.36	10.88	63.61	16.23
Soft Triple + PRISM (Ours)	**64.81**	**11.21**	**64.46**	**17.53**
DML with Pair-based Losses				
MS	49.00	5.92	52.53	9.82
Contrastive	44.91	4.76	50.04	9.42
Memory Contrastive Loss (MCL)	38.73	3.34	**52.58**	**9.88**
MCL + PRISM (Ours)	**57.95**	**8.04**	52.47	9.64

Experiments

Algorithm	Training Time (Seconds)
Memory Contrastive Loss (MCL)	1,679.22
MCL + PRISM without centers	12,294.76
MCL + PRISM with centers	1,777.38
Soft Triple	1,685.47
Soft Triple + PRISM with centers	1,767.97

THANKS