Circle Loss: A Unified Perspective of Pair Similarity Optimization

Yifan Sun, Changmao Cheng, Yuhan Zhang, Chi Zhang, Liang Zheng, Zhongdao Wang, Yichen Wei

Megvii Inc., Beihang University, Australian National University, Tsinghua University

CVPR 2020

Introduction

Feature learning:

- maximize the within-class similarity s_p
- minimize the between-class similarity s_n

Many popular losses, e.g., triplet loss:

embed s_n and s_p into similarity pairs and seek to make (s_n – s_p) < m

(symmetric optimization)

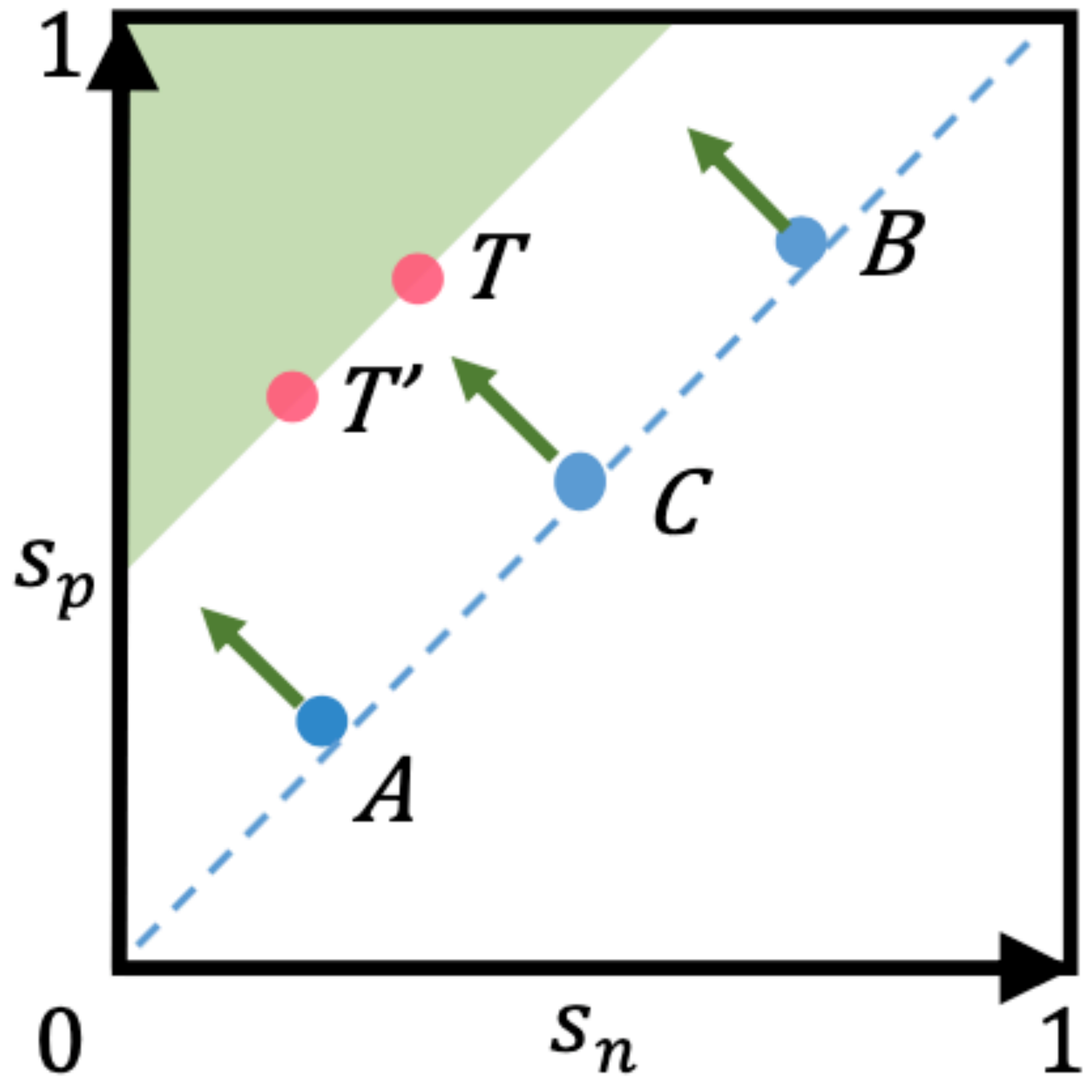
Problems:

- Lack of flexibility for optimization. The penalty strength on s_n and s_p is equal.
- Ambiguous convergence status. The decision boundary allows ambiguity for convergence.

Problems:

- Lack of flexibility for optimization. The penalty strength on s_n and s_p is equal.
- Ambiguous convergence status. The decision boundary allows ambiguity for convergence.

Example: T has $\{s\_n, s\_p\} = \{0.2, 0.5\}$, T' has $\{s\_n', s\_p'\} = \{0.4, 0.7\}$, $m = 0.3$



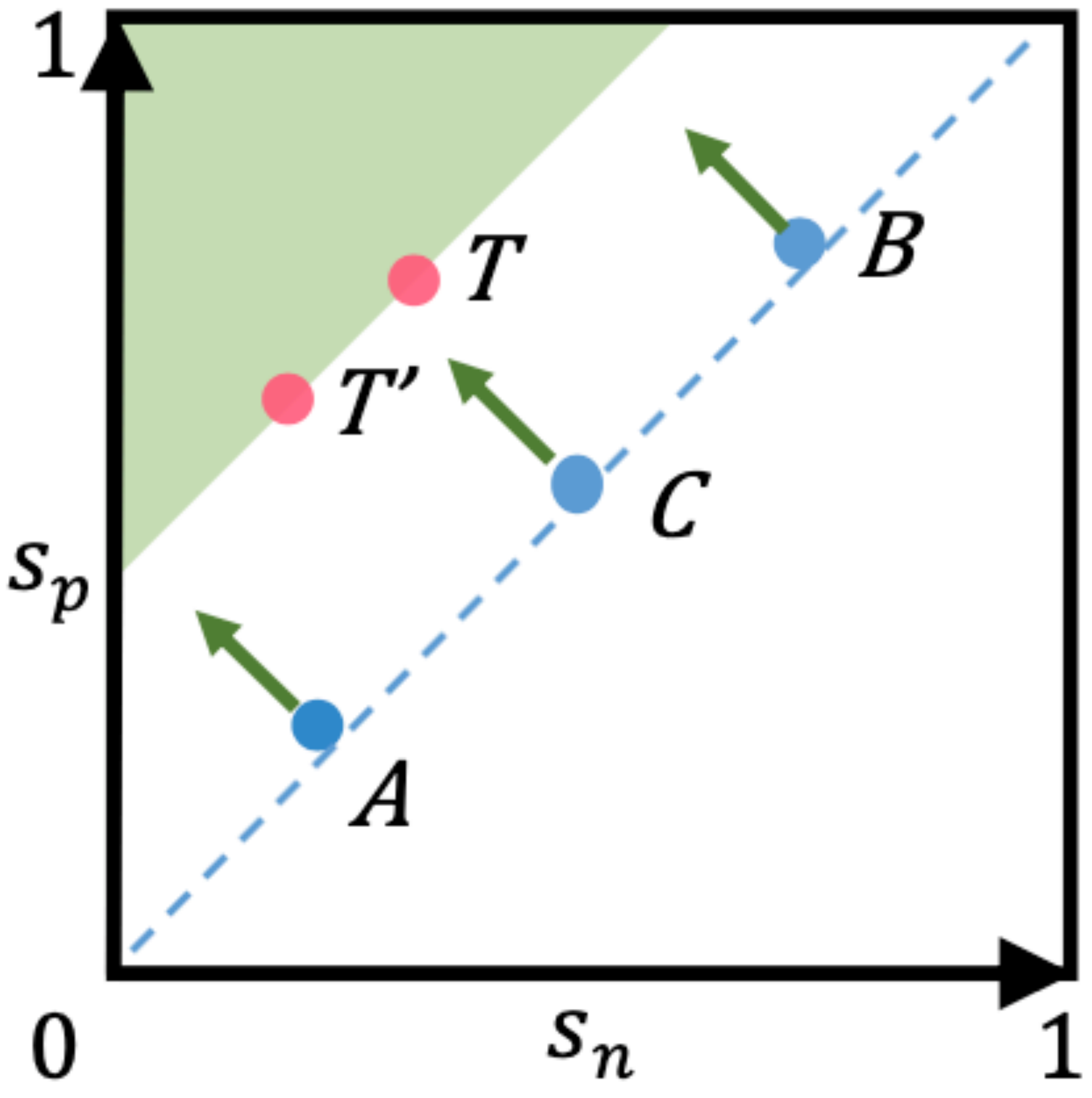
(a)

Problems:

- Lack of flexibility for optimization. The penalty strength on s_n and s_p is equal.
- Ambiguous convergence status. The decision boundary allows ambiguity for convergence.

Example: T has {s_n, s_p} = {0.2, 0.5}, T' has {s_n', s_p'} = {0.4, 0.7}, m = 0.3

{s_n', s_p} = {0.4, 0.5}



(a)

Circle Loss

Reweight the pair:

- different similarity scores should have different penalty strength.
- if a similarity score deviates far from the optimum, it should receive strong penalty.
- the weights are linear functions w.r.t similarity scores.

$$(\alpha_n s_n - \alpha_p s_p)$$

$$\begin{cases} \alpha_p = [O_p - s_p]_+, \\ \alpha_n = [s_n - O_n]_+, \end{cases}$$

Circle Loss

Reweight the pair:

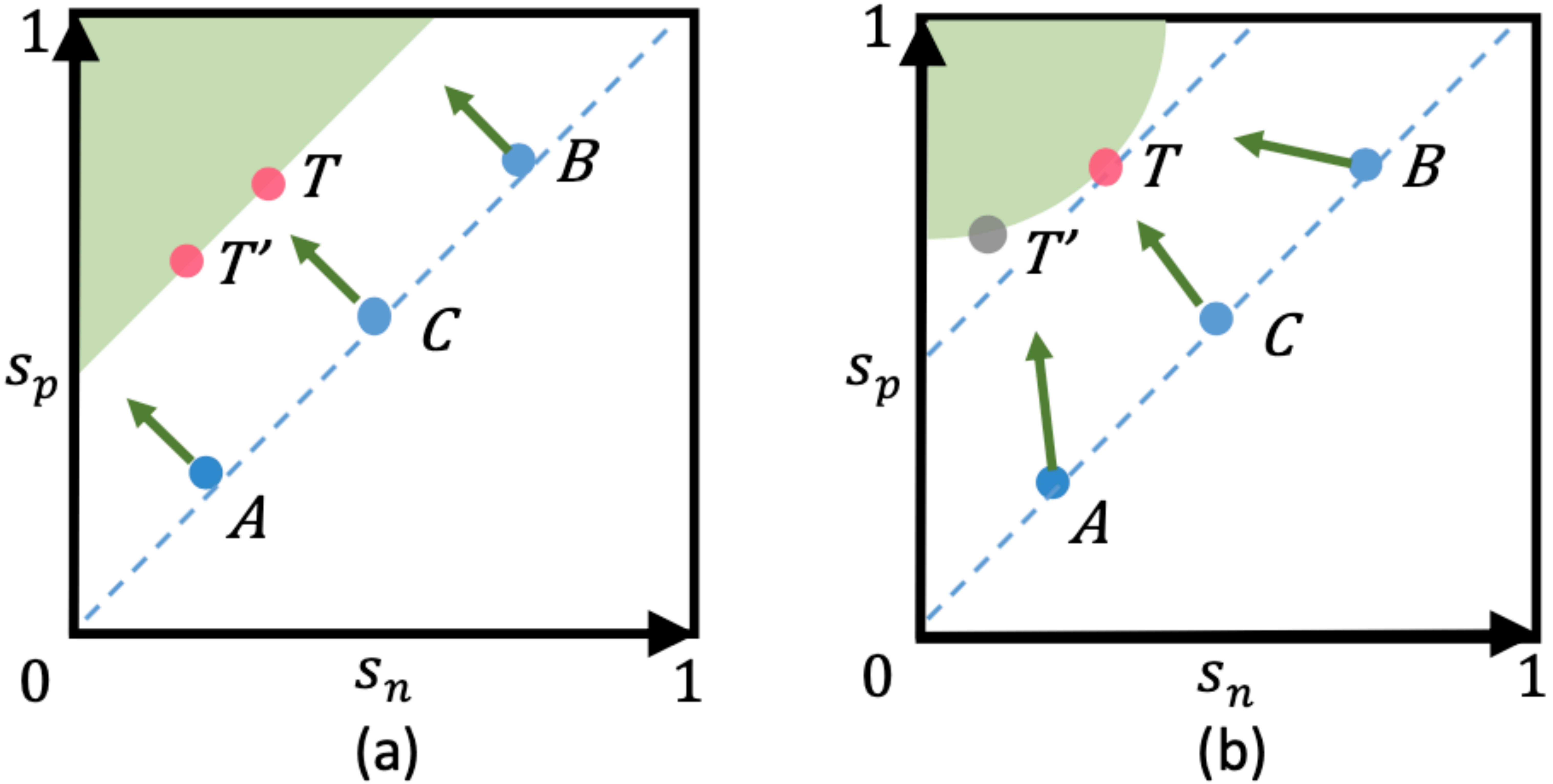
- different similarity scores should have different penalty strength.
- if a similarity score deviates far from the optimum, it should receive strong penalty.
- the weights are linear functions w.r.t similarity scores.

$$(\alpha_n s_n - \alpha_p s_p)$$

$$\begin{cases} \alpha_p = [\overset{1}{O_p} - s_p]_+, \\ \alpha_n = [s_n - \underset{0}{O_n}]_+, \end{cases}$$

Problems:

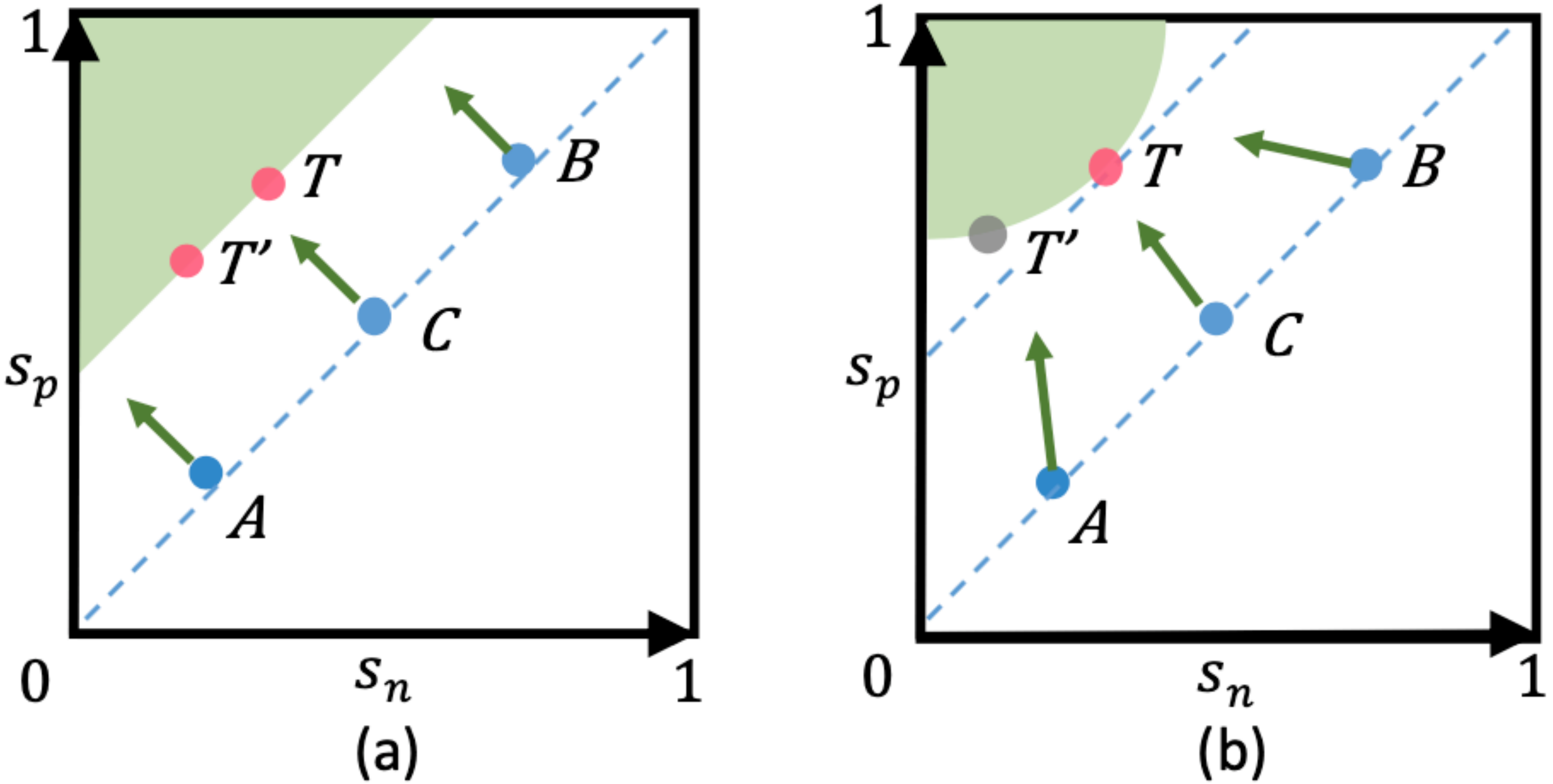
- Lack of flexibility for optimization. The penalty strength on s_n and s_p is equal.
- Ambiguous convergence status. The decision boundary allows ambiguity for convergence.



Triplet loss ***Circle loss***

Problems:

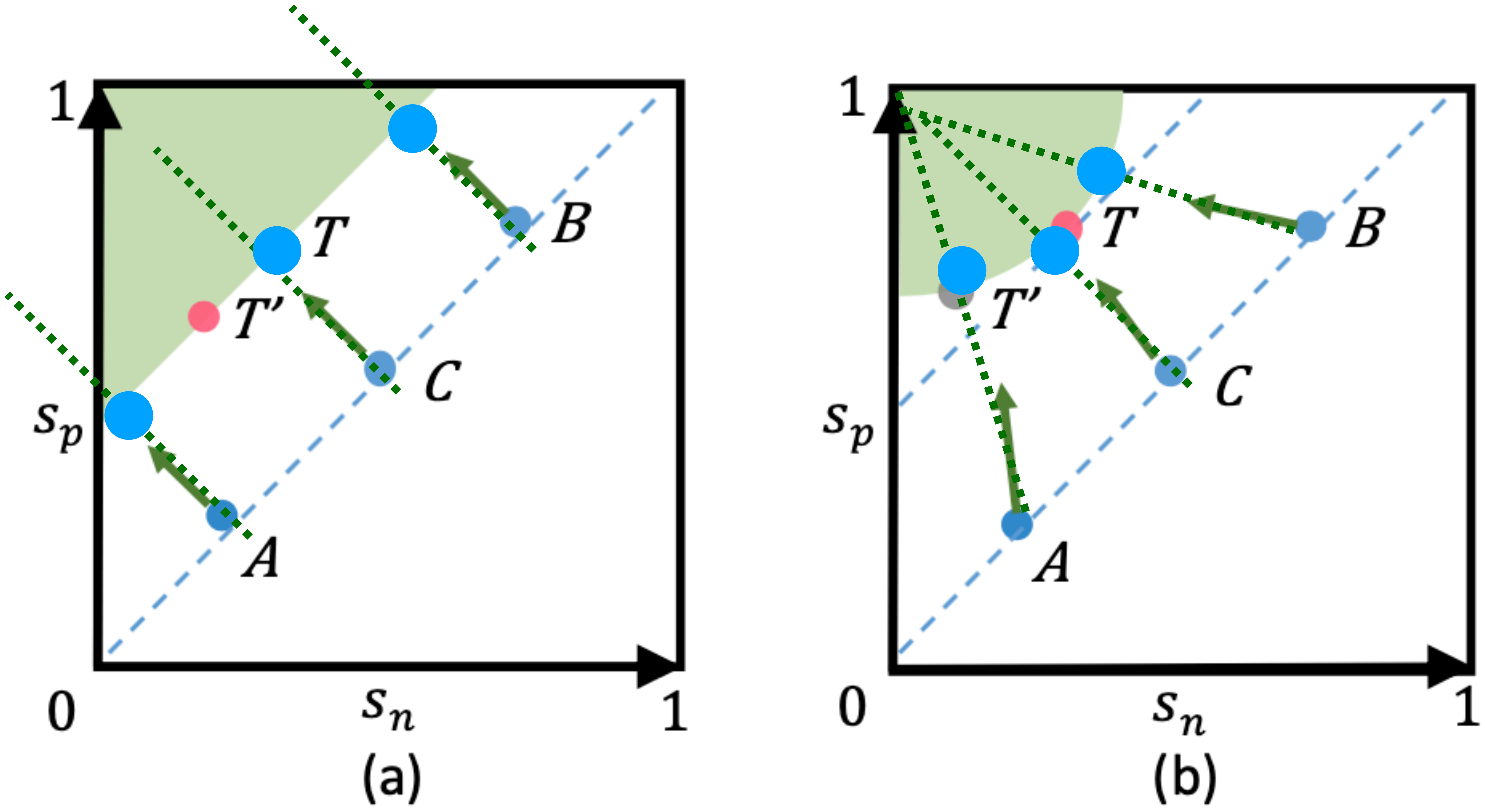
- ~~Lack of flexibility for optimization. The penalty strength on s_n and s_p is equal.~~
- Ambiguous convergence status. The decision boundary allows ambiguity for convergence.



Different directions

Problems:

- ~~Lack of flexibility for optimization. The penalty strength on s_n and s_p is equal.~~
- ~~Ambiguous convergence status. The decision boundary allows ambiguity for convergence.~~



Closer on the decision boundary

The Gradients of Losses

Circle loss assigns different gradients to the similarity scores, depending on their distances to the optimum (e.g. A and B)

Experiment: Face recognition

Table 1: Identification rank-1 accuracy (%) on MFC1 dataset with different backbones and loss functions.

Loss function	MFC1 [12] rank-1		
	ResNet34	ResNet50	ResNet100
Softmax	92.36	93.91	95.04
NormFace [30]	92.62	94.12	95.27
AM-Softmax [29, 32]	97.54	97.86	98.31
ArcFace [2]	97.68	98.03	98.36
CircleLoss (ours)	**97.81**	**98.17**	**98.50**

Table 2: Face verification accuracy (%) on LFW, YTF and CFP-FP with ResNet34 backbone.

Loss function	LFW [10]	YTF [37]	CFP-FP [23]
Softmax	99.18	96.19	95.01
NormFace [30]	99.25	96.03	95.34
AM-Softmax [29, 32]	99.63	96.31	95.78
ArcFace [2]	99.68	96.34	95.84
CircleLoss(ours)	**99.73**	**96.38**	**96.02**

Experiment: Person re-ID

Table 4: Evaluation of Circle loss on re-ID task. We report R-1 accuracy (%) and mAP (%).

Method	Market-1501		MSMT17	
	R-1	mAP	R-1	mAP
PCB [26] (Softmax)	93.8	81.6	68.2	40.4
MGN [31] (Softmax+Triplet)	95.7	86.9	-	-
JDGL [42]	94.8	86.0	**77.2**	**52.3**
ResNet50 + AMSoftmax	92.4	83.8	75.6	49.3
ResNet50 + CircleLoss(ours)	94.2	84.9	76.3	50.2
MGN + AMSoftmax	95.3	86.6	76.5	51.8
MGN + CircleLoss(ours)	**96.1**	**87.4**	76.9	52.1

Experiment: Fine-grained image retrieval

Table 5: Comparison with state of the art on CUB-200-2011, Cars196 and Stanford Online Products. R@K(%) is reported.

Loss function	CUB-200-2011 [28]				Cars196 [14]				Stanford Online Products [19]			
	R@1	R@2	R@4	R@8	R@1	R@2	R@4	R@8	R@1	R@10	R@10^2	R@10^3
LiftedStruct [19]	43.6	56.6	68.6	79.6	53.0	65.7	76.0	84.3	62.5	80.8	91.9	97.4
HDC [18]	53.6	65.7	77.0	85.6	73.7	83.2	89.5	93.8	69.5	84.4	92.8	97.7
HTL [3]	57.1	68.8	78.7	86.5	81.4	88.0	92.7	95.7	74.8	88.3	94.8	98.4
ABIER [20]	57.5	71.5	79.8	87.4	82.0	89.0	93.2	96.1	74.2	86.9	94.0	97.8
ABE [13]	60.6	71.5	79.8	87.4	**85.2**	**90.5**	94.0	96.1	76.3	88.4	94.8	98.2
Multi-Simi [34]	65.7	77.0	**86.3**	91.2	84.1	90.4	94.0	96.5	78.2	90.5	96.0	**98.7**
CircleLoss(ours)	**66.7**	**77.4**	86.2	**91.2**	83.4	89.8	**94.1**	**96.5**	**78.3**	**90.5**	**96.1**	98.6

Conclusion

- *Circle loss allows the similarity scores to learn at different paces.*
- *High flexibility in optimization.*
- *A more definite convergence target.*

Comment

- *Simple but effective*
- *Easy to implement*
- *Good presentastion of motivation*