Appendix

Hong Xuan[1][0000−0002−4951−3363], Abby Stylianou[2], Xiaotong Liu[1], and Robert Pless[1]

[1] The George Washington University, Washington DC 20052
{xuanhong,liuxiaotong2017,pless}@gwu.edu
[2] Saint Louis University, St. Louis MO 63103
abby.stylianou@slu.edu

1 Similarity after gradient updating(NCA-based triplet loss)

The following derivations show how to get S_{ap}^{new} and S_{an}^{new} in equation 7 and 8 with updated and unnormalized $\mathbf{f_a}^{new}$,$\mathbf{f_p}^{new}$,$\mathbf{f_n}^{new}$ in equation 4, 5 and 6.

$$\begin{aligned} S_{ap}^{new} &= \mathbf{f_a}^{new\intercal}\mathbf{f_p}^{new} \\ &= (1+\beta^2)\mathbf{f_a}^\intercal\mathbf{f_p} + \beta\mathbf{f_a}^\intercal\mathbf{f_a} + \beta\mathbf{f_p}^\intercal\mathbf{f_p} - \beta\mathbf{f_n}^\intercal\mathbf{f_p} - \beta^2\mathbf{f_n}^\intercal\mathbf{f_a} \\ &= (1+\beta^2)S_{ap} + 2\beta - \beta S_{pn} - \beta^2 S_{an} \end{aligned} \tag{1}$$

$$\begin{aligned} S_{an}^{new} &= \mathbf{f_a}^{new\intercal}\mathbf{f_n}^{new} \\ &= (1+\beta^2)\mathbf{f_a}^\intercal\mathbf{f_n} - \beta\mathbf{f_a}^\intercal\mathbf{f_a} - \beta\mathbf{f_n}^\intercal\mathbf{f_n} + \beta\mathbf{f_p}^\intercal\mathbf{f_n} - \beta^2\mathbf{f_p}^\intercal\mathbf{f_a} \\ &= (1+\beta^2)S_{ap} - 2\beta + \beta S_{pn} - \beta^2 S_{ap} \end{aligned} \tag{2}$$

For the calculation of S_{ap}^{new}, we construct two hyper-planes: P_{ap} spanned by $\mathbf{f_a}$ and $\mathbf{f_p}$, and P_{an} spanned by $\mathbf{f_a}$ and $\mathbf{f_n}$. On P_{ap}, $\mathbf{f_p}$ can be decomposed into two components: $\mathbf{f_p^{a\|}}$(the direction along $\mathbf{f_a}$) and $\mathbf{f_p^{a\perp}}$(the direction vertical to $\mathbf{f_a}$). On P_{an}, $\mathbf{f_n}$ can be decomposed into two components: $\mathbf{f_n^{a\|}}$(the direction along $\mathbf{f_a}$) and $\mathbf{f_n^{a\perp}}$(the direction vertical to $\mathbf{f_a}$). Then the S_{pn} should be:

$$\begin{aligned} S_{pn} &= \mathbf{f_p}^\intercal\mathbf{f_n} \\ &= (\mathbf{f_p^{a\|}} + \mathbf{f_p^{a\perp}})(\mathbf{f_n^{a\|}} + \mathbf{f_n^{a\perp}}) \\ &= S_{ap}S_{an} + \gamma\sqrt{1-S_{ap}^2}\sqrt{1-S_{an}^2} \end{aligned} \tag{3}$$

where $\gamma = \frac{\mathbf{f_p^{a\perp}}\mathbf{f_n^{a\perp}}}{\|\mathbf{f_p^{a\perp}}\|\|\mathbf{f_n^{a\perp}}\|}$, which represents the projection factor between P_{ap} and P_{an}. When $\mathbf{f_a}$, $\mathbf{f_n}$, and $\mathbf{f_p}$ are close enough so that locally the hypersphere is a plane, then γ is the dot-product of normalized vector from $\mathbf{f_a}$ to $\mathbf{f_p}$ and $\mathbf{f_a}$ to $\mathbf{f_n}$. If $\mathbf{f_p}$,$\mathbf{f_a}$,$\mathbf{f_n}$ are co-planer then $\gamma = 1$, and if moving from $\mathbf{f_a}$ to $\mathbf{f_p}$ is orthogonal to the direction from $\mathbf{f_a}$ to $\mathbf{f_n}$, then $\gamma = 0$.

Dataset	CUB				CAR			
Method	R@1	R@2	R@4	R@8	R@1	R@2	R@4	R@8
Triplet-Semihard[64]	42.6	55.03	66.4	77.2	51.5	63.8	73.5	82.4
Lifted[64]	43.6	56.6	68.6	79.6	53.0	65.7	76.0	84.3
Clustering[64]	49.8	61.4	71.8	81.9	58.1	70.6	80.3	87.8
SmartMining[64]	49.8	62.3	74.1	83.3	64.7	76.2	84.2	90.2
ProxyNCA[64]	49.2	61.9	67.9	72.4	73.2	82.4	86.4	88.7
N-pair[64]	51.9	64.3	74.9	83.2	71.1	79.7	86.5	91.6
SHN[64]	56.2	67.0	78.0	86.1	68.2	77.8	85.5	91.0
SCT[64]	**57.6**	**69.8**	**79.3**	**87.2**	**73.2**	**82.4**	**88.4**	**93.0**

Table 1. Retrieval Performance on the CUB and CAR datasets comparing to the best reported results with embedding dimension 64 trained on ResNet50.

2 Norm of updated features(NCA-based triplet loss)

The following derivation shows how to derive $\|\mathbf{f_a}^{new}\|$, $\|\mathbf{f_p}^{new}\|$ and $\|\mathbf{f_n}^{new}\|$ in equation 9 and 10. On P_{ap}, $\mathbf{g_p}$ can be decomposed into the direction along $\mathbf{f_p}$ and the direction vertical to $\mathbf{f_p}$. On P_{an}, $\mathbf{g_n}$ can be decomposed into the direction along $\mathbf{f_n}$ and the direction vertical to $\mathbf{f_n}$. Then,

$$\|\mathbf{f_p}^{new}\|^2 = (1+\beta S_{ap})^2 + \beta^2(1-S_{ap}^2) \tag{4}$$

$$\|\mathbf{f_n}^{new}\|^2 = (1-\beta S_{an})^2 + \beta^2(1-S_{an}^2) \tag{5}$$

On P_{ap}, $\mathbf{g_a}$ can be decomposed into 3 components: component in the plane and along $\mathbf{f_a}$, component in the plane and vertical $\mathbf{f_a}$, and component vertical to P_{ap}. Then,

$$\begin{aligned}\|\mathbf{f_a}^{new}\|^2 = &(1+\beta S_{ap}-\beta S_{an})^2 \\ &+(\beta\sqrt{1-S_{ap}^2}-\gamma\beta\sqrt{1-S_{an}^2})^2 \\ &+(\beta\sqrt{1-\gamma^2}\sqrt{1-S_{an}^2})^2\end{aligned} \tag{6}$$

3 Gradient updates(margin-based triplet loss)

The main paper uses NCA-based triplet loss. Another margin-based triplet-loss function is derived to guarantee a specific margin. This can be expressed in terms of $\mathbf{f_a}$,$\mathbf{f_p}$,$\mathbf{f_n}$ as:

$$L = \max(\|\mathbf{f_a}-\mathbf{f_p}\|^2 - \|\mathbf{f_a}-\mathbf{f_n}\|^2 + \alpha, 0) = \max(D, 0) \tag{7}$$

Dataset	SOP			In-shop		
Method	R@1	R@10	R@100	R@1	R@10	R@20
BIER [4]^{512}	72.7	86.5	94.0	76.9	92.8	95.2
ABE [2]^{512}	76.3	88.4	94.8	87.3	96.7	97.9
FastAP [1]^{512}	76.4	89.0	95.1	**90.9**	**97.7**	**98.5**
MS [6]^{512}	78.2	90.5	96.0	89.7	97.9	98.5
EasyPositive [7]^{512}	78.3	90.7	96.3	87.8	95.7	96.8
SHN^{512}	80.8	92.0	**96.9**	90.1	97.3	98.2
SCT^{512}	**81.6**	**92.3**	96.7	90.0	97.3	98.1

Table 2. Retrieval Performance on the SOP and In-shop datasets comparing to the best reported results with embedding dimension 512 trained on ResNet50

$$\mathbf{g_p} = \frac{\partial L}{\partial \mathbf{f_p}} = \begin{cases} -\beta(\mathbf{f_a}-\mathbf{f_p}) & \text{if } D>0 \\ 0 & \text{otherwise} \end{cases} \tag{8}$$

$$\mathbf{g_n} = \frac{\partial L}{\partial \mathbf{f_n}} = \begin{cases} \beta(\mathbf{f_a}-\mathbf{f_n}) & \text{if } D>0 \\ 0 & \text{otherwise} \end{cases} \tag{9}$$

$$\mathbf{g_a} = \frac{\partial L}{\partial \mathbf{f_a}} = \begin{cases} \beta(\mathbf{f_n}-\mathbf{f_p}) & \text{if } D>0 \\ 0 & \text{otherwise} \end{cases} \tag{10}$$

where $D = (\|\mathbf{f_a} - \mathbf{f_p}\|^2 - \|\mathbf{f_a} - \mathbf{f_n}\|^2 + \alpha)$ and $\beta = 2$. For simplicity, in the following discussion, we set $D > 0$ for margin-based triplet loss. Then we can get the $\mathbf{f_a}^{new}$, $\mathbf{f_p}^{new}$ and $\mathbf{f_n}^{new}$ and their norm:

$$\mathbf{f_p}^{new} = \mathbf{f_p} + \beta(\mathbf{f_a} - \mathbf{f_p}) = (1-\beta)\mathbf{f_p} + \beta\mathbf{f_a} \tag{11}$$

$$\mathbf{f_n}^{new} = \mathbf{f_n} - \beta(\mathbf{f_a} - \mathbf{f_n}) = (1+\beta)\mathbf{f_n} - \beta\mathbf{f_a} \tag{12}$$

$$\mathbf{f_a}^{new} = \mathbf{f_a} - \beta\mathbf{f_n} + \beta\mathbf{f_p} \tag{13}$$

$$\|\mathbf{f_p}^{new}\|^2 = (1-\beta+\beta S_{ap})^2 + \beta^2(1-S_{ap}^2) \tag{14}$$

$$\|\mathbf{f_n}^{new}\|^2 = (1+\beta-\beta S_{an})^2 + \beta^2(1-S_{an}^2) \tag{15}$$

$$\begin{aligned} \|\mathbf{f_a}^{new}\|^2 = &(1+\beta S_{ap} - \beta S_{an})^2 \\ &+ (\beta\sqrt{1-S_{ap}^2} - \gamma\beta\sqrt{1-S_{an}^2})^2 \\ &+ (\beta\sqrt{1-\gamma^2}\sqrt{1-S_{an}^2})^2 \end{aligned} \tag{16}$$

The updated similarity S_{ap}^{new} and S_{an}^{new} will be:

$$S_{ap}^{new} = (1-\beta+\beta^2)S_{ap} + 2\beta - \beta^2 - \beta(1-\beta)S_{pn} - \beta^2 S_{an} \tag{17}$$

$$S_{an}^{new} = (1+\beta+\beta^2)S_{an} - 2\beta - \beta^2 + \beta(1+\beta)S_{pn} - \beta^2 S_{ap} \tag{18}$$

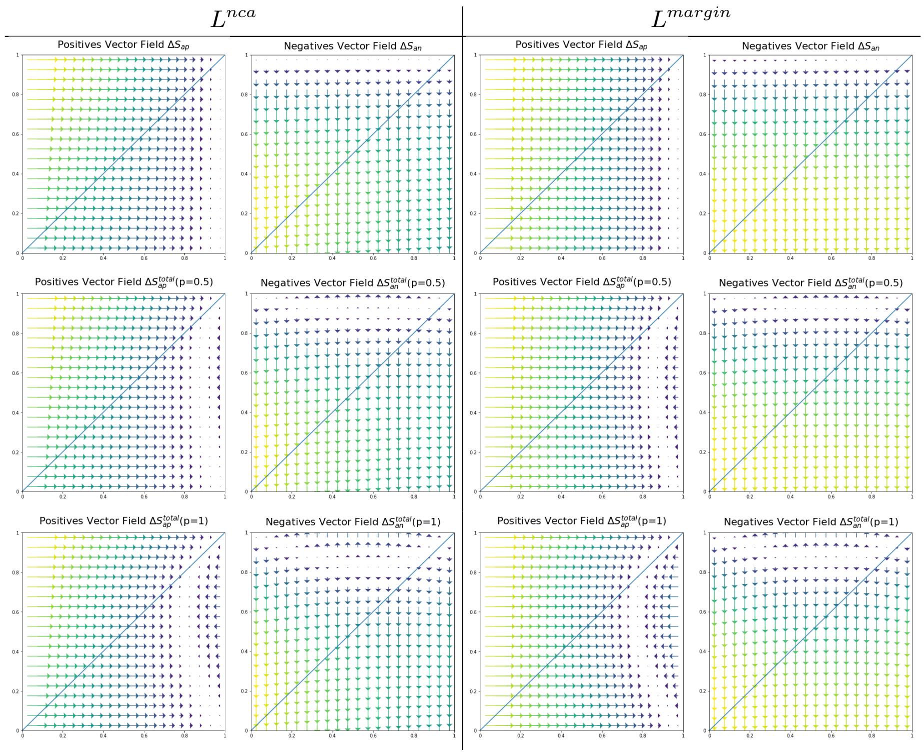

Fig. 1. Numerical simulation for ΔS_{ap}, ΔS_{an}, ΔS_{ap}^{total} and ΔS_{an}^{total} change of L^{nca} and L^{margin} with p=0.5,1.0.

Comparing to the S_{ap}^{new} and S_{an}^{new} in equation 7 and 8 of main paper, margin-based triplet loss behavior is similar to the NCA-based triplet loss. And we simulate the ΔS_{ap} and ΔS_{an} with the margin-based triplet loss in figure 1. These plots show that the behavior of problematic regions are qualitatively similar for both methods with different values of entanglement strength p.

4 More Results on other dataset

We also compare our results with state-of-the-art embedding approaches such as BIER [4],ABE [2],FaseAP [1] Multi-Similarity [6] and Easy Positive [7] on SOP [5] and In-shop [3] dataset. Tables 2 shows the SC-triplet loss outperforms the best previously reported results on the SOP dataset.

References

1. Cakir, F., He, K., Xia, X., Kulis, B., Sclaroff, S.: Deep metric learning to rank. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2019)

2. Kim, W., Goyal, B., Chawla, K., Lee, J., Kwon, K.: Attention-based ensemble for deep metric learning. In: Proc. European Conference on Computer Vision (ECCV) (September 2018)
3. Liu, Z., Luo, P., Qiu, S., Wang, X., Tang, X.: Deepfashion: Powering robust clothes recognition and retrieval with rich annotations. In: Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2016)
4. Opitz, M., Waltner, G., Possegger, H., Bischof, H.: Bier - boosting independent embeddings robustly. In: Proc. International Conference on Computer Vision (ICCV) (Oct 2017)
5. Song, H.O., Xiang, Y., Jegelka, S., Savarese, S.: Deep metric learning via lifted structured feature embedding. In: Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (2016)
6. Wang, X., Han, X., Huang, W., Dong, D., Scott, M.R.: Multi-similarity loss with general pair weighting for deep metric learning. In: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition. pp. 5022–5030 (2019)
7. Xuan, H., Stylianou, A., Pless, R.: Improved embeddings with easy positive triplet mining. In: The IEEE Winter Conference on Applications of Computer Vision (WACV) (March 2020)