Sampling Wisely: Deep Image Embedding by Top-k Precision Optimization Supplementary Materials

Jing Lu[1*] Chaofan Xu[2,3*] Wei Zhang[2] Lingyu Duan[4] Tao Mei[2]
[1]Business Growth BU, JD [2] JD AI Research [3]Harbin Institute of Technology [4]Peking University
lvjing12@jd.com, xuchaofan1994@126.com, wzhang.cu@gmail.com, lingyu@pku.edu.cn, tmei@live.com

1. Proof for Theorems

This section provides the proofs for the 4 theorems.

Theorem 1. ***Upper bounding***: *For any $n_+ < k$ and* $\mathbf{s}$,

$$\ell_k(\mathbf{s},\mathbf{y}) \geq \gamma\ell_{Prec@k}(\mathbf{s},\mathbf{y}) - \gamma(k-n_+) \tag{1}$$

Proof. We rewrite the loss function by adding the scores in set $\mathcal{K}\backslash\mathcal{N}$ to both of the two terms.

$$\begin{aligned}\ell_k(\mathbf{s},\mathbf{y}) &= \sum_{z_i\in\mathcal{N}} \hat{s}_i - \sum_{z_i\in\mathcal{P}} \hat{s}_i \\ &= \sum_{z_i\in\mathcal{K}} \hat{s}_i - \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} \hat{s}_i\end{aligned} \tag{2}$$

Note that $\mathcal{K}$ was defined as the top k ranked according to $\hat{s}_i$. We also define set $\mathcal{K}'$ as the top k ranked according to s_i, i.e. $\mathcal{K}' = \{z_i \in \mathcal{C} : s_i \geq s_{[k]}\}$. So, the first term,

$$\sum_{z_i\in\mathcal{K}} \hat{s}_i \geq \sum_{z_i\in\mathcal{K}'} \hat{s}_i = \gamma\ell_{\text{Prec@}k}(\mathbf{s},\mathbf{y}) + \sum_{z_i\in\mathcal{K}'} s_i \tag{3}$$

We further consider the second term. In set $\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}$ there are k images including n_+ positive images. So

$$\sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} \hat{s}_i = \gamma(k-n_+) + \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} s_i \tag{4}$$

By definition $\sum_{z_i\in\mathcal{K}'} s_i$ is the maximum for the sum of s_i over k images and $|\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}| = k$ so,

$$\sum_{z_i\in\mathcal{K}'} s_i \geq \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} s_i \tag{5}$$

Combining the three formulas above concludes this proof. □

Theorem 2. ***Consistency***: *For any $n_+ < k$, when there is a large margin γ between positive images and negative images that should be ranked out of $\mathcal{K}$ (the $k-n_++1$-th ranked negative image), i.e. $s^+_{[n_+]} - s^-_{[k-n_++1]} \geq \gamma$, we have $\ell_k(\mathbf{s},\mathbf{y}) = \ell_{Prec@k}(\mathbf{s},\mathbf{y}) - (k-n_+) = 0$. Here $\mathbf{s}^+ \in \mathbb{R}^{n_+}$ and $\mathbf{s}^- \in \mathbb{R}^{n-n_+}$ are two sub-vectors of $\mathbf{s}$ containing the similarity scores of positive and negative images.*

Proof. If $\mathcal{K}$ contains all n_+ positive images, $\mathcal{P} = \mathcal{N} = \emptyset$, obviously, $\ell_k(\mathbf{s},\mathbf{y}) = 0$.

We now assume $\mathcal{K}$ contains $n'_+ < n_+$ positive images, $\mathcal{P} = \mathcal{N} \neq \emptyset$. We have,

$$\begin{aligned}\ell_k(\mathbf{s},\mathbf{y}) &= \sum_{z_i\in\mathcal{K}} \hat{s}_i - \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} \hat{s}_i \\ =&\gamma(k-n'_+) + \sum_{z_i\in\mathcal{K}} s_i - \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} \hat{s}_i \\ =&\gamma(k-n'_+) + \sum_{z_i\in\mathcal{K}\backslash\mathcal{N}} s_i + \sum_{z_i\in\mathcal{N}} s_i - \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} \hat{s}_i\end{aligned} \tag{6}$$

$|\mathcal{N}| = n_+ - n'_+$. Based on the definition of $\mathcal{N}$ and the large margin condition,

$$\sum_{z_i\in\mathcal{N}} s_i \leq \gamma(n'_+ - n_+) + \sum_{z_i\in\mathcal{P}} s_i \tag{7}$$

So

$$\ell_k(\mathbf{s},\mathbf{y}) \leq \quad \gamma(k-n_+) + \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} s_i - \hat{s}_i \tag{8}$$

The set $\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}$ contains $k-n_+$ negative images. So $\ell_k(\mathbf{s},\mathbf{y}) \leq 0$. Since we already known $\ell_k(\mathbf{s},\mathbf{y}) \geq 0$ from Theorem 1, we conclude $\ell_k(\mathbf{s},\mathbf{y}) = 0$. □

We now prove the two properties of Case 2 in the following 2 Theorems. [1]

Theorem 3. ***Upper bounding***: *For any $n_+ > k$, and* $\mathbf{s}$,

$$\ell_k(\mathbf{s},\mathbf{y}) \geq \gamma\ell_{Prec@k}(\mathbf{s},\mathbf{y}) \tag{9}$$

*Equal Contribution

[1]A special case of was proven in [13]

Proof. The set $\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}$ contains only k positive images.

$$\sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} \hat{s}_i = \sum_{z_i\in\mathcal{P}\bigcup\mathcal{K}\backslash\mathcal{N}} s_i \tag{10}$$

which is different from Eq 4 in Theorem 1. Other steps of this proof is straight forward, so we omit them for concise. □

Theorem 4. ***Consistency****: For $n_+ > k$, when there is a large margin γ between the top k positive and the top negative images, i.e. $s^+_{[k]} - s^-_{[1]} \geq \gamma$, we have $\ell_{Prec@k} = \ell_k = 0$.*

Proof. We also assumes $n'_+ < n_+$. So

$$\ell_k(\mathbf{s},\mathbf{y}) = \sum_{z_i\in\mathcal{N}} \hat{s}_i - \sum_{z_i\in\mathcal{P}} \hat{s}_i = \gamma|\mathcal{N}| + \sum_{z_i\in\mathcal{N}} s_i - \sum_{z_i\in\mathcal{P}} s_i \tag{11}$$

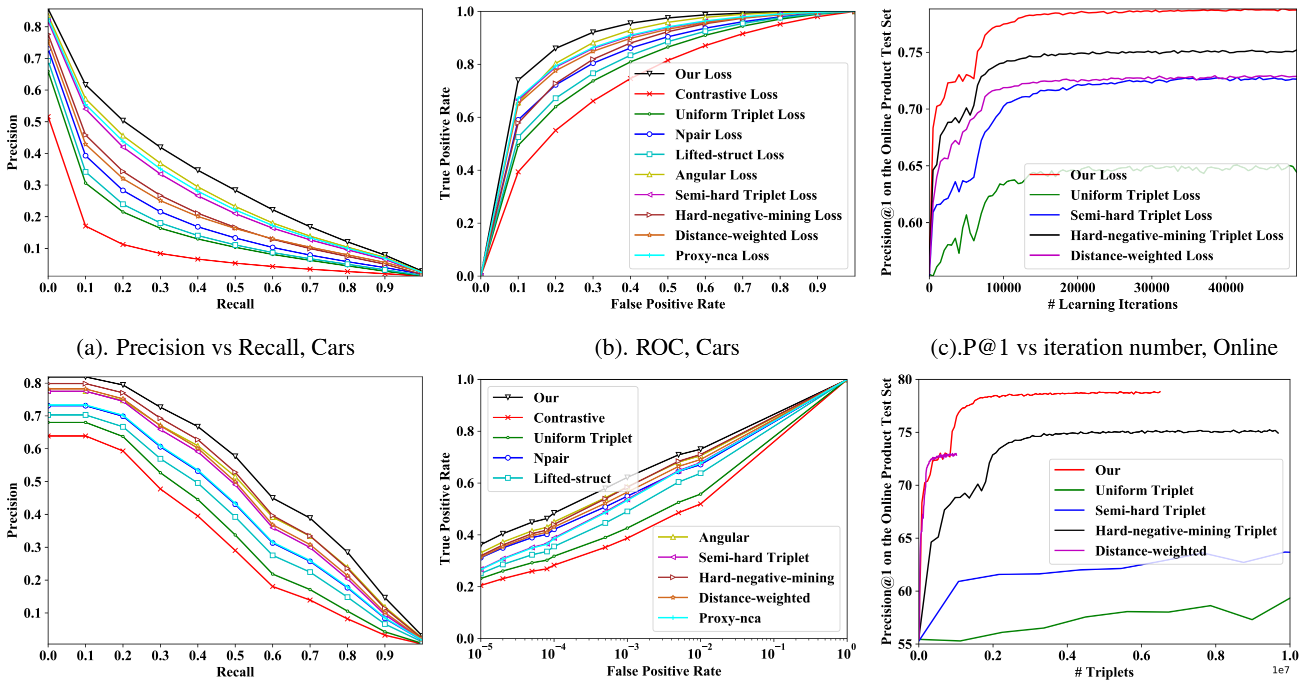
Given the large margin condition, $\sum_{z_i\in\mathcal{P}} s_i - \sum_{z_i\in\mathcal{N}} s_i \geq \gamma|\mathcal{N}|$, We have $\ell_k(\mathbf{s},\mathbf{y}) \leq 0$. Combining with the above theorem $\ell_k(\mathbf{s},\mathbf{y}) \geq 0$, we have $\ell_k(\mathbf{s},\mathbf{y}) = 0$. □

	P/R@1	P@3	P@5	P@10	R@3	R@5	R@10	NMI	mAP	F1
CUB-200-2011										
Uniform Triplet	44.53	40.64	38.83	35.45	64.84	73.80	83.12	54.96	20.75	19.42
Hard Mining Triplet	53.88	50.01	47.64	43.92	72.64	79.74	87.51	62.17	27.14	30.02
Semi Hard Tripltet	51.87	48.62	46.58	43.18	71.44	79.03	86.75	61.14	27.16	27.01
Distance Weighted	50.49	46.70	44.18	40.64	70.44	77.94	86.44	60.41	24.59	27.87
Contrastive Loss	39.69	36.19	34.01	31.01	59.06	68.11	80.17	53.09	18.33	20.48
Lifted Struct Loss	45.19	41.37	39.10	35.94	66.00	74.11	83.59	58.07	21.86	23.58
N-Pair Loss	50.61	47.75	45.22	41.56	69.68	76.86	85.62	59.56	25.79	25.97
Angular Loss	51.98	47.58	45.14	41.03	71.42	78.93	86.80	60.99	24.32	27.83
Proxy NCA Loss	52.70	48.79	46.34	42.42	71.48	78.54	86.14	61.64	26.13	28.52
Ours ℓ_k	**54.12**	**50.17**	**47.90**	**44.43**	**72.69**	**80.30**	**87.98**	**63.53**	**27.79**	**31.70**
Standford Cars										
Uniform Triplet	52.97	45.46	41.16	34.61	70.51	77.17	85.02	44.73	12.97	12.00
Hard Mining Triplet	69.12	62.05	57.74	51.00	83.20	87.76	92.19	57.00	22.38	25.29
Semi Hard Triplet	62.35	56.00	52.19	46.35	77.92	83.68	89.19	54.19	21.87	22.32
Distance Weighted	59.02	52.75	48.80	42.95	75.55	81.29	87.52	52.36	19.98	20.42
Contrastive Loss	38.00	30.71	27.19	22.39	54.32	62.61	73.23	34.93	7.49	7.00
Lifted Struct Loss	56.56	49.04	44.71	37.97	73.31	79.34	86.16	46.27	14.47	13.25
N-Pair Loss	61.75	53.70	49.18	42.27	77.01	82.60	88.53	49.47	16.56	15.63
Angular Loss	71.44	64.73	60.70	53.57	84.28	88.65	92.81	57.40	23.48	25.28
Proxy NCA Loss	72.39	66.14	62.05	54.98	85.46	89.71	93.40	59.00	24.18	27.21
Ours ℓ_k	**73.3**4	**67.37**	**63.34**	**56.17**	**86.29**	**90.38**	**94.12**	**59.64**	**24.79**	**27.73**
Online Product										
Uniform Triplet t	61.82	45.97	36.30	23.19	70.65	74.08	78.30	27.36	44.35	24.27
Hard Mining Triplet	72.94	57.65	46.87	30.51	80.62	83.58	86.97	36.54	37.45	33.79
Semi Hard Triplet	67.46	51.88	41.58	26.81	75.68	79.02	83.11	32.05	49.52	27.85
Distance Weighted	67.21	51.69	41.55	26.88	75.50	78.74	82.56	27.65	49.55	25.50
Contrastive Loss	58.14	41.97	32.56	20.52	66.71	70.33	74.83	26.98	41.14	25.63
Lifted Struct Loss	64.45	48.60	38.63	24.73	72.89	76.31	80.34	37.84	46.72	33.52
N-Pair Loss	65.51	49.76	39.70	25.51	73.84	77.29	81.39	35.86	47.74	31.09
Angular Loss	68.43	52.66	42.33	27.37	76.66	79.79	83.61	30.04	50.43	27.77
Proxy NCA Loss	67.21	51.50	41.25	26.26	75.43	78.73	82.61	36.37	49.32	31.90
Ours ℓ_k	**74.95**	**59.90**	**48.89**	**31.89**	**82.40**	**85.24**	**88.45**	**38.03**	**52.34**	**35.27**

Table 1. Comparison with state-of-the-art sampling methods and loss functions on three benchmark datasets. The network backbone is Inception with batch normalization layer. "P" is for precision and "R" is for Recall. Note that NMI and F1 in Online Product Dataset are computed by 12 super categories for time efficiency.

Backbone	Uniform	Hard	Semi-hard	Distance	Contrastive	Lifted	Npair	Angular	Proxy	Ours
Inception	88.20	89.53	89.49	89.57	87.21	88.69	88.94	89.77	86.91	**90.07**
Dense201	87.89	90.95	90.98	90.23	87.19	88.76	89.48	91.03	89.54	**91.94**

Table 2. In our previous tables on Online Product, we reported the NMI and F1 on 12 super classes for time efficiency. For easy comparison with that in literature, we also report the NMI for $11k$ fine-grained classes.



(a). Precision vs Recall, Cars (b). ROC, Cars (c).P@1 vs iteration number, Online

(d).Precision vs Recall, Online (e).ROC, Online (f). P@1 vs triplets number, Online

Figure 1. Precision vs Recall curve, ROC curve on Cars and Online Product dataset (a,b,d,e, shared legend). The top-1 precision on test data along the training process of Online Product dataset. (c, f, shared legend). Our algorithm outperforms all baselines.Other results in our main file. The steps in performance gain in figure (c) is due to the decrease in learning rate.

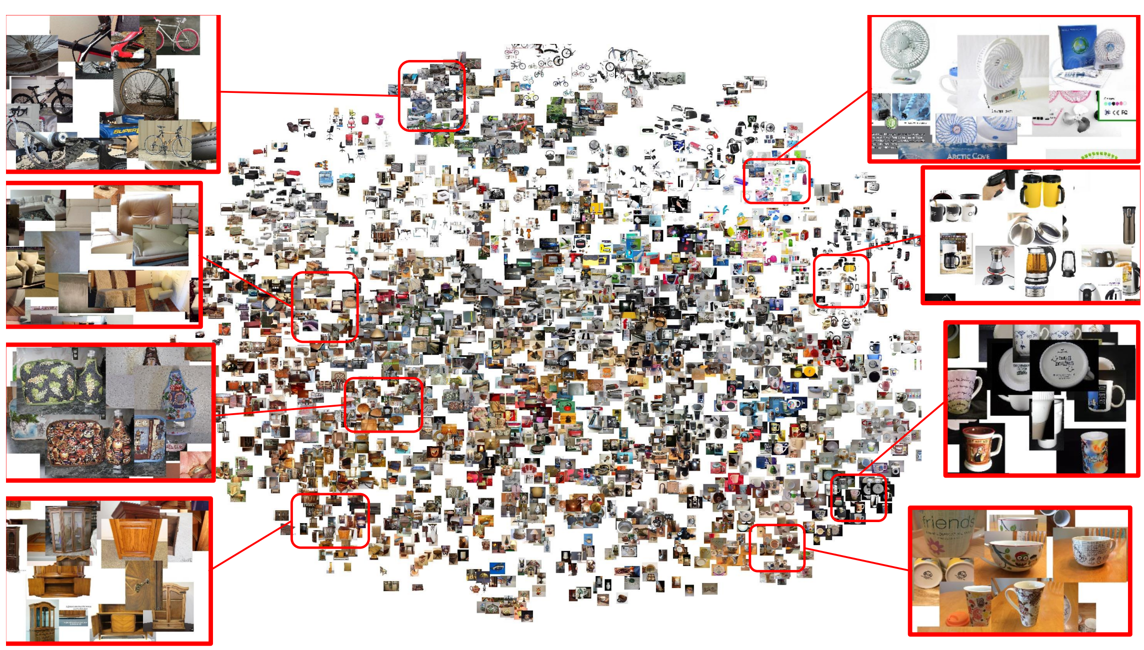


Figure 2. Barnes-Hut t-SNE visualization of our embedding on the test split of Online Product dataset. The embedding generated by the proposed algorithm put similar images in clusters. Best viewed on a monitor zoomed in.