# Embedding Transfer with Label Relaxation for Improved Metric Learning —Supplementary Material—

Sungyeon Kim[1] Dongwon Kim[1] Minsu Cho[1,2] Suha Kwak[1,2]
Dept. of CSE, POSTECH[1] Graduate School of AI, POSTECH[2]
{sungyeon.kim, kdwon, mscho, suha.kwak}@postech.ac.kr

This supplementary material presents deeper analyses of the proposed method, its implementation details, and additional experimental results, all of which are omitted from the main paper due to the space limit. Section 1 first introduces relaxed MS loss, an integration of the label relaxation and Multi-Similarity (MS) loss [3]. The generalization capability of our model is then illustrated in terms of the spectral decay metric [2] in Section 2. Section 3 describes details of the multi-view data augmentation strategy. In Section 4, we investigate the effect of hyperpameters on the performance of our loss. Finally, in Section 5, we present more qualitative examples for image retrieval before and after applying the proposed method on the three metric learning benchmarks.

## 1. Relaxed MS Loss

The proposed label relaxation technique can be applied to other metric learning losses based on pairwise relations of data. In this section, we present *relaxed MS loss* that is a combination of the label relaxation and Multi-Similarity (MS) loss [3], the state-of-the-art loss for pair-based metric learning. Specifically, relaxed MS loss is obtained by using relaxed relation labels instead of the binary class equivalence indicator and replacing cosine similarity with the relative Euclidean distance, like relaxed contrastive loss in the main paper. The relaxed MS loss is then formulated as

$$\mathcal{L}(X) = \frac{1}{n}\sum_{i=1}^{n}\left\{\frac{1}{\alpha}\log\left[1+\sum_{j\neq i} w_{ij}^{s}\exp\left(\alpha\frac{d_{ij}^{t}}{\mu_i}\right)\right] + \frac{1}{\beta}\log\left[1+\sum_{j\neq i}(1-w_{ij}^{s})\exp\left(\beta(\delta-\frac{d_{ij}^{t}}{\mu_i})\right)\right]\right\}, \tag{1}$$

where $n$ is the number of samples in the batch, $\delta$ is a margin, and $\alpha > 0$ and $\beta > 0$ are scaling factors. Also, $\mu_i = \frac{1}{n}\sum_{k=1}^{n} d_{ij}^{t}$ is the average distance of all pairs associated with $f_i^t$ in the batch.

Table 1 compares relaxed constrastive loss and relaxed MS loss on the three benchmarks for deep metric learning in the *self-transfer* and *dimensionality reduction* setting. The details for training are the same as those for relaxed contrastive loss. We set $\alpha$ and $\beta$ to 1 and 4 respectively in the self-transfer setting and 1 and 2 respectively in the dimension reduction setting. As shown in the table, relaxed MS loss achieves performance comparable to a relaxed contrastive loss in both settings. This result demonstrates the universality of our label relaxation method.

However, performance of relaxed MS loss is worse than that of the relaxed contrastive loss in most cases, and it demands careful tuning of hyper-parameters for each setting. In contrast, relaxed contrastive loss is overall better in terms of performance, more robust against hyper-parameter setting, and more interpretable due to its simplicity; this is the reason why we choose relaxed contrastive loss as the representative loss of our framework.

## 2. Generalization Effect of Label Relaxation

Spectral decay $\rho$ [2] is a recently proposed generalization measure for deep metric learning. It measures KL-divergence between the singular value spectrum of training data embeddings and a uniform distribution. Lower $\rho$ value means that a larger number of directions with significant variance exists in the embedding space, thus indicates better generalization [2].

In Fig. 1, Spectral decay $\rho$ of the source and target embedding models is presented. Target embedding models are trained with our method and its variants. As shown in the figure, target embedding models after embedding transfer have lower $\rho$ value than the source, and our method significantly reduces $\rho$ value compared to its unrelaxed or L2-normalized version. We argue that our method reduces $\rho$ value and improves generalization performance since the relaxed relation labels and the relative pairwise distance helps target embedding space to encode rich pairwise relation without restriction on the manifold.

| Recall@$K$ | | | CUB-200-2011 | | | Cars-196 | | | SOP | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 1 | 2 | 4 | 1 | 2 | 4 | 1 | 10 | 100 |
| (a) | *Source*: PA [1] | $BN^{512}$ | 69.1 | 78.9 | 86.1 | 86.4 | 91.9 | 95.0 | 79.2 | 90.7 | 96.2 |
| | Relaxed contrastive | $BN^{512}$ | 72.1 | 81.3 | 87.6 | **89.6** | **94.0** | **96.5** | **79.8** | **91.1** | **96.3** |
| | Relaxed MS | $BN^{512}$ | **72.3** | **81.3** | **88.3** | 89.2 | 93.9 | 96.4 | 79.3 | 90.8 | 96.1 |
| (b) | *Source*: PA [1] | $BN^{512}$ | 69.1 | 78.9 | 86.1 | 86.4 | 91.9 | 95.0 | 79.2 | 90.7 | 96.2 |
| | Relaxed contrastive | $BN^{64}$ | 67.4 | **78.0** | **85.9** | **86.5** | **92.3** | **95.3** | **76.3** | **88.6** | **94.8** |
| | Relaxed MS | $BN^{64}$ | **67.5** | 77.9 | 85.9 | 86.0 | 91.5 | 94.8 | 75.4 | 87.9 | 94.6 |

Table 1. Image retrieval performance of two types of relaxed losses in the two different settings: (a) Self-transfer and (b) dimensionality reduction. Embedding networks of the methods are denoted by abbreviations: BN–Inception with BatchNorm. Superscripts indicate embedding dimensions of the networks.

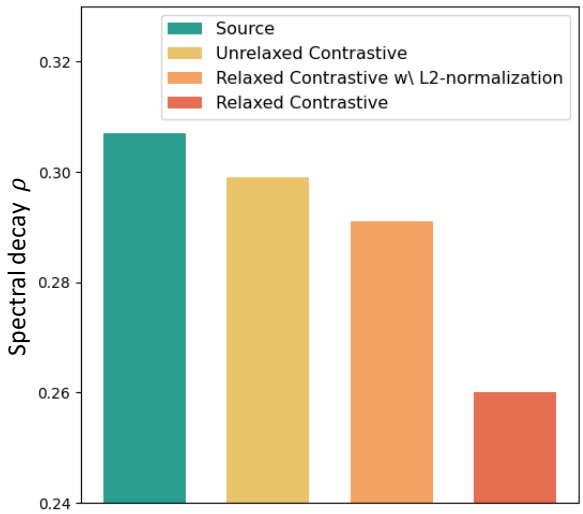


Figure 1. Spectral decay $\rho$ of source and target embedding models trained on the Cars-196 dataset.

| Methods | Recall@1 | |
|---|---|---|
| | CUB | Cars |
| Ours | 72.1 | 89.6 |
| Ours w/o augmentation | 71.5 | 89.0 |

Table 2. Effect of multi-view augmentation strategy on the CUB-200-2011 (CUB) and Cars-196 (Cars) datasets.

## 3. Details of Multi-view Data Augmentation

In recent approaches to self-supervised representation learning, the use of multi-view samples produced from the same image plays an important role for performance improvement. We present a simple yet effective multi-view augmentation strategy for enhancing the effect of embedding transfer. It helps transfer knowledge by considering relations between multiple views of individual samples, such as relations between different parts of an object.

The overall procedure of our multi-view augmentation is as follows. We first apply the standard random augmentation technique multiple times to images of input batch. Then, all augmented multi-view images are passed through the source and target embedding networks. Note that the source and target model take the same augmented image as input. The output embedding vectors are concatenated and used as the inputs of the embedding transfer loss. Fig. 2 illustrates this procedure where the number of views is two. The top and bottom of the figure describe the standard augmentation technique and our strategy, respectively. When using standard augmentation, only relations between different samples are considered.

Applying a multi-view augmentation strategy for embedding transfer allows knowledge transfer to consider more diverse and detailed relations between samples produced from the same image. The empirical advantage of the multi-view augmentation is verified in Table 2, where it improves the stability and convergence of embedding transfer as well as the performance of target embedding models.

## 4. Impact of Hyperparameters

We empirically investigated the effect of the hyperparameters $\delta$ and $\sigma$ on performance. We examine Recall@1 in accuracy of relaxed contrastive loss by varying the values of the hyperparameters $\sigma \in \{0.25, 0.5, 1, 2, 4\}$ and $\delta \in \{0.8, 0.9, 1, 1.1, 1.2\}$. As summarized in Figure 3, the accuracy of our method was consistently high and outperformed state of the art in most cases when $\sigma$ is greater than 0.25 and less than 4. Note that the values of $\delta$ and $\sigma$ used in the paper are not optimal as we did not tune them using the test set.

## 5. Additional Qualitative Results

More qualitative results of image retrieval on the CUB-200-2011, Cars-196, and SOP datasets are presented in Fig. 4, 5, and 6, respectively. We prove the positive effect of the proposed method by showing qualitative results before and after applying the proposed method in the *self-transfer* setting; the source embedding model is Inception-BatchNorm with 512 embedding dimension and trained with the proxy-anchor loss [1]. The overall results indicate that the proposed method significantly improves the source embedding model. From the examples of the 2nd, 3rd, and

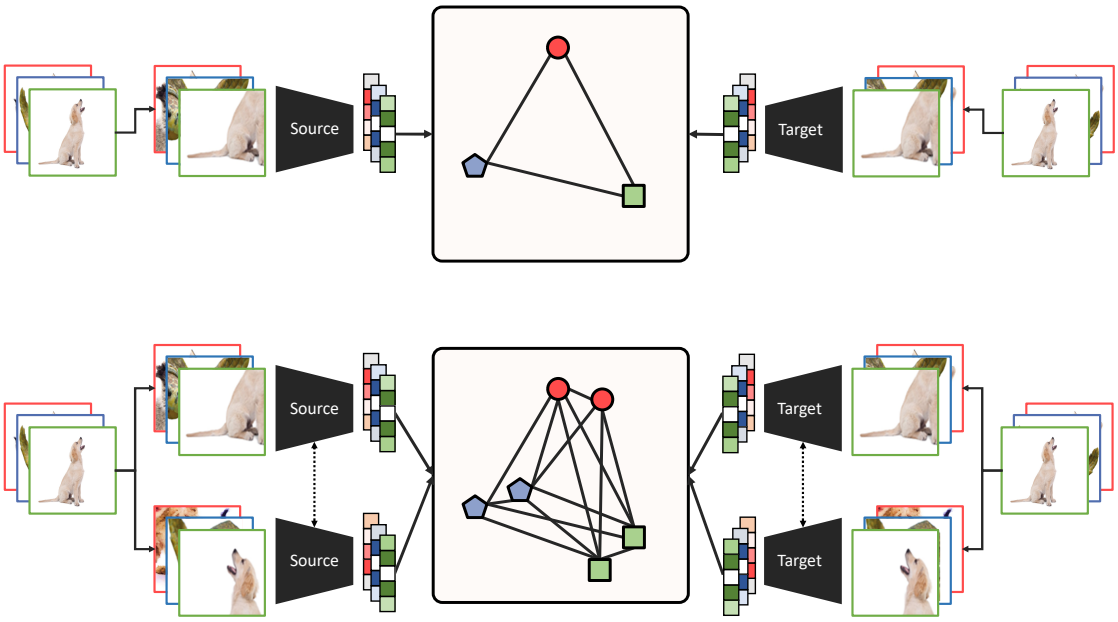


Figure 2. Comparison of standard data augmentation (*top*) and our multi-view augmentation (*bottom*). Different colors and shapes represent distinct samples.

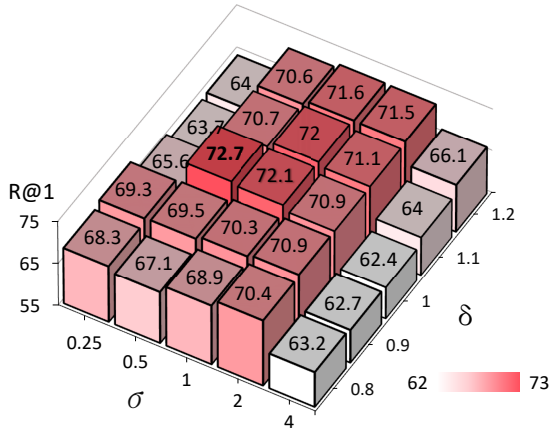


Figure 3. Recall@1 versus hyperparameters $\delta$ and $\sigma$ on the CUB dataset.

5th rows of Fig. 4, both models retrieve birds visually similar to the query, but only the models after embedding transfer successfully retrieved birds of the same species. Meanwhile, the examples of the 2nd, 3rd, and 5th rows of Fig. 5 show that the model trained with our method provides accurate results regardless of the color changes of the cars. Also, in the examples of the 2nd and 3rd rows of Fig. 6, the source model makes mistakes easily since the false positives are similar to the query in terms of appearance, yet it becomes more accurate after applying to embedding transfer with the proposed method.

## References

[1] Sungyeon Kim, Dongwon Kim, Minsu Cho, and Suha Kwak. Proxy anchor loss for deep metric learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2020. 2, 4, 5

[2] Karsten Roth, Timo Milbich, Samarth Sinha, Prateek Gupta, Bjoern Ommer, and Joseph Paul Cohen. Revisiting training strategies and generalization performance in deep metric learning. In *Proc. International Conference on Machine Learning (ICML)*, 2020. 1

[3] Xun Wang, Xintong Han, Weilin Huang, Dengke Dong, and Matthew R Scott. Multi-similarity loss with general pair weighting for deep metric learning. In *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019. 1

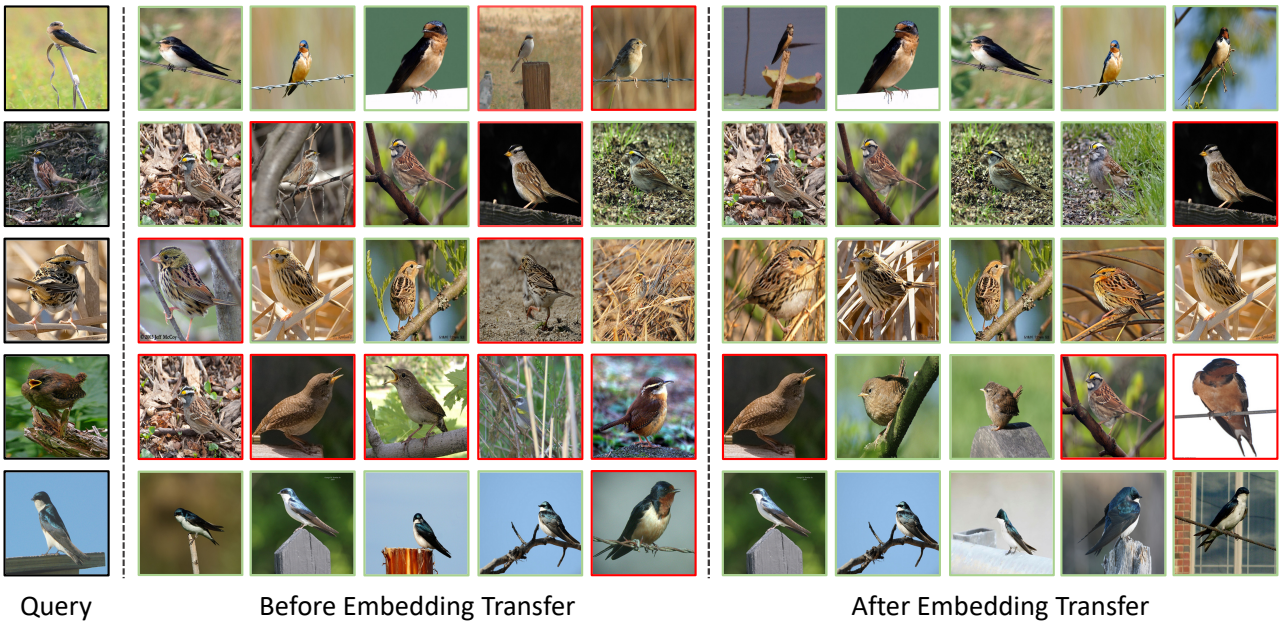


Figure 4. Top 5 image retrievals of the state of the art [1] before and after the proposed method is applied on the CUB-200-2011 dataset. Images with green boundary are success cases and those with red boundary are false positives.

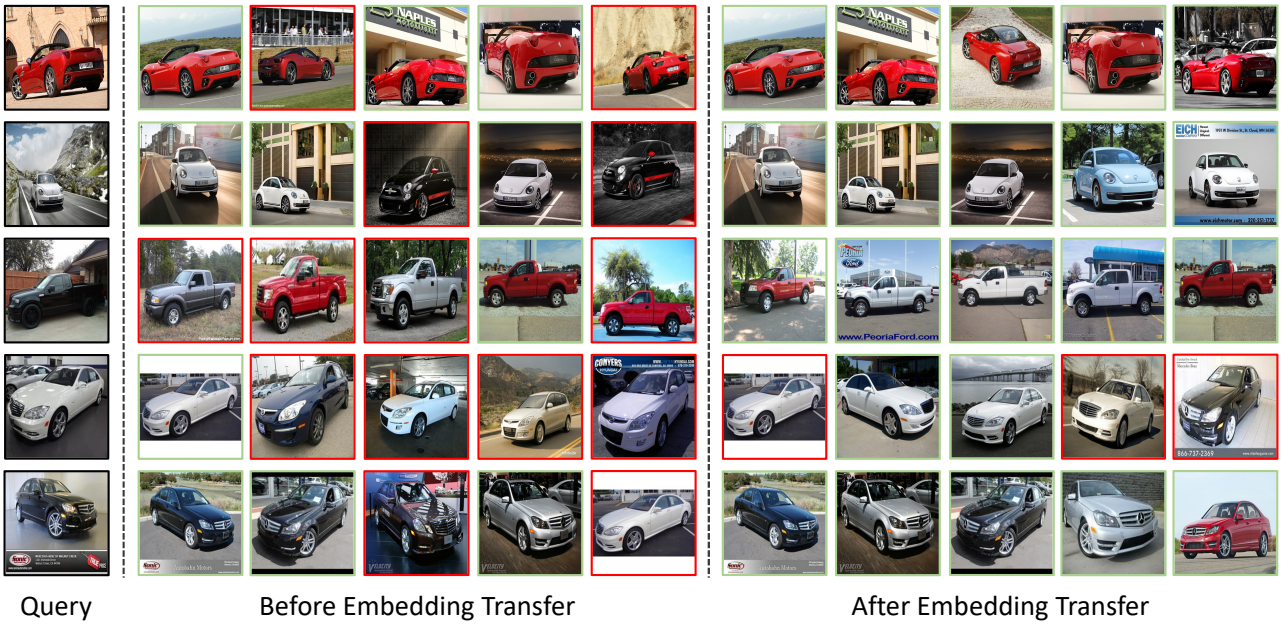


Figure 5. Top 5 image retrievals of the state of the art [1] before and after the proposed method is applied on the Cars-196 dataset. Images with green boundary are success cases and those with red boundary are false positives.


Figure 6. Top 5 image retrievals of the state of the art [1] before and after the proposed method is applied on the SOP dataset. Images with green boundary are success cases and those with red boundary are false positives.