Supplementary Material:
Sampling Matters in Deep Embedding Learning

Chao-Yuan Wu
UT Austin
cywu@cs.utexas.edu

R. Manmatha
A9/Amazon
manmatha@a9.com

Alexander J. Smola
Amazon
smola@amazon.com

Philipp Krähenbühl
UT Austin
philkr@cs.utexas.edu

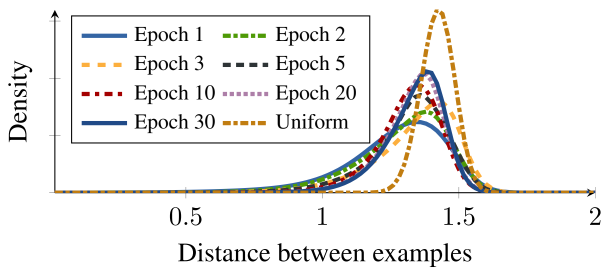

Figure 1: Empirical pairwise-distance distributions for negative pairs. They roughly follow a bell-shaped curve.

1. Empirical pairwise-distance distributions

To better understand the effects of distance weighted sampling during training, we analyze our learned embeddings. Specifically, we compute empirical pairwise distance distributions for negative pairs based on the embeddings of testing images. Figure 1 presents the results on Stanford Online Product dataset. We see that after the first epoch, the distribution already forms a bell shape, and in later epochs, it gradually concentrates. This justifies our motivation of using distance weighted sampling so that examples from all distances have a chance to be sampled.

2. Stability analysis

Here we measure the stability of different loss functions when using different batch construction. Specifically, we change the number of images m per class in a batch and see how it impacts the solutions. For this purpose, we experiment with face verification and use the optimal verification boundary on the validation set as a summary of the solution. The results are summarized in Figure 2. We see that the triplet loss converges to different solutions when using different batch constructions. In addition, we observe large fluctuations in the early stage, indicating unstable training. On the other hand, the margin based loss is robust, it always converges to the roughly the same geometry.

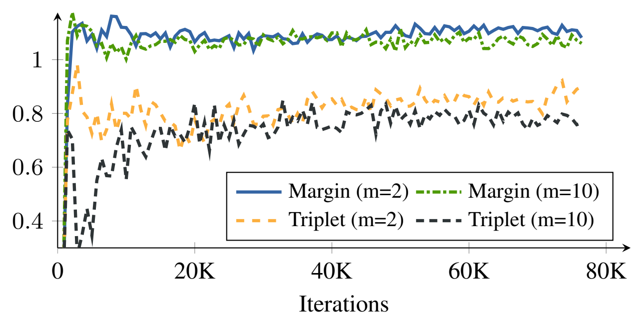

Figure 2: Optimal validation threshold for the LFW dataset. Triplet loss with different sampling strategies converges to different solutions. In addition, it has large fluctuations in the early stage, indicating unstable training. Margin based loss always converges stably to the same solution.

3. Ablation study for batch size

We analyze the sensitivity of our approach with respect to batch sizes. Table 1 presents the results. We see that distance weighted sampling consistently outperforms other sampling strategies, and margin based loss consistently outperforms triplet loss.

Loss, batch size	@1	@10	@100	@1000
Triplet ℓ_2, 40				
Semihard	44.3	63.7	79.7	92.2
Distance weighted	52.9	70.9	83.9	94.0
Triplet ℓ_2, 80				
Semihard	47.4	67.5	83.1	93.6
Distance weighted	54.5	72.0	85.4	94.4
Triplet ℓ_2, 120				
Semihard	48.8	67.7	82.7	93.3
Distance weighted	54.7	72.7	85.9	94.6
Margin, 40				
Random	41.9	60.2	76.3	89.6
Semihard	60.7	75.3	85.9	94.1
Distance weighted	61.1	75.8	86.5	94.2
Margin, 80				
Random	37.5	56.3	73.8	88.3
Semihard	61.0	74.6	85.3	93.6
Distance weighted	61.7	75.5	86.0	94.0
Margin, 120				
Random	37.7	56.6	73.7	88.3
Semihard	59.6	73.7	84.4	93.2
Distance weighted	60.5	74.7	85.5	93.8

Table 1: Recall@k evaluated on Stanford Online Products.