A. Things we tried but they did not work

Our initial experiments were focused on self-supervised learning (SSL) in hyperbolic space. However, we noticed that the head output, which is usually ignored in SSL formulation, shows high performance, thus we switched to a more suitable metric learning formulation. During preliminary experiments, we tried our method with the ResNet-50 [2] backbone. In this case, the hyperbolic version also outperforms the sphere-based version. However, we do not publish and compare CNN-based architecture with other methods, because the transformer backbone performs clearly better without drawbacks, and we focus on it. We had a modification of our method with the MoCo [1] loss. However, it performed similarly to plain cross-entropy loss, so we decided not to include it in the final version. Also, we tried our method with ProxyNCA [4] loss, which performed worse.

B. Datasets visualization

Figures 1 to 4 illustrate how learned embeddings are arranged on the Poincaré disk. We use UMAP [3] method with the “hyperboloid” distance metric to reduce the dimensionality to 2D for visualization. Embeddings are obtained with Hyp-DINO configuration for CUB-200-2011 and Cars-196 datasets. Each point inside the disk corresponds to a sample, different colors indicate different classes.

Figures 5 and 6 demonstrate actual images of the first 4000 samples of the evaluation split of CUB-200-2011 and Cars-196 datasets. We use the layout from Figures 2 and 4 projected to a uniform 2D grid, preserving neighborhood relations of samples.
Figure 1. CUB-200-2011 train set

Figure 2. CUB-200-2011 test set

Figure 3. Cars-196 train set

Figure 4. Cars-196 test set
Figure 6. Cars-196 test subset (4000 images)
References


