Supplementary

A. Derivation of Gradient in Eq. 8

$$
\begin{aligned}
&\nabla \mathcal{L}_{NCE_i} \\
&= \frac{1}{\left(1 + \sum_{k \neq i} \exp\left(\frac{1}{\tau}\left(s_{i,k} - s_{i,i}\right)\right)\right)} \nabla \sum_{k \neq i} \exp\left(\frac{1}{\tau}\left(s_{i,k} - s_{i,i}\right)\right) \\
&= \frac{1}{\tau} \frac{\sum_{k \neq i} \exp\left(\frac{1}{\tau}\left(s_{i,k} - s_{i,i}\right)\right) \nabla \left(s_{i,k} - s_{i,i}\right)}{\left(1 + \sum_{k \neq i} \exp\left(\frac{1}{\tau}\left(s_{i,k} - s_{i,i}\right)\right)\right)} \\
&= \frac{1}{\tau} \sum_{k \neq i} \frac{\exp\left(\frac{1}{\tau} s_{i,k}\right)}{\exp\left(\frac{1}{\tau} s_{i,i}\right)\left(1 + \sum_{k \neq i} \exp\left(\frac{1}{\tau}\left(s_{i,k} - s_{i,i}\right)\right)\right)} \nabla \left(s_{i,k} - s_{i,i}\right) \\
&= \frac{1}{\tau} \sum_{k \neq i} \frac{\exp\left(s_{i,k}/\tau\right)}{\exp\left(s_{i,i}/\tau\right) + \sum_{k \neq i} \exp\left(s_{i,k}/\tau\right)} \nabla \left(s_{i,k} - s_{i,i}\right)
\end{aligned}
\tag{11}
$$

B. Comparison with Different ViTs

In Figure 4, we demonstrate that temperature τ has a consistent impact on various ViT models when equipped with general contrastive loss and hyperbolic contrastive loss. Specifically, we evaluate ViT-s, DINO, and DeiT-s. Across different backbone transformer settings, hyperbolic embeddings consistently outperform Euclidean embeddings when $\tau > 0.2$. For DINO hyperbolic embeddings show similar performance when $\tau = 0.2$ and $\tau = 0.3$. When τ increases, the performance of both Euclidean and hyperbolic embeddings drops. However, hyperbolic embeddings are always superior to the Euclidean case.

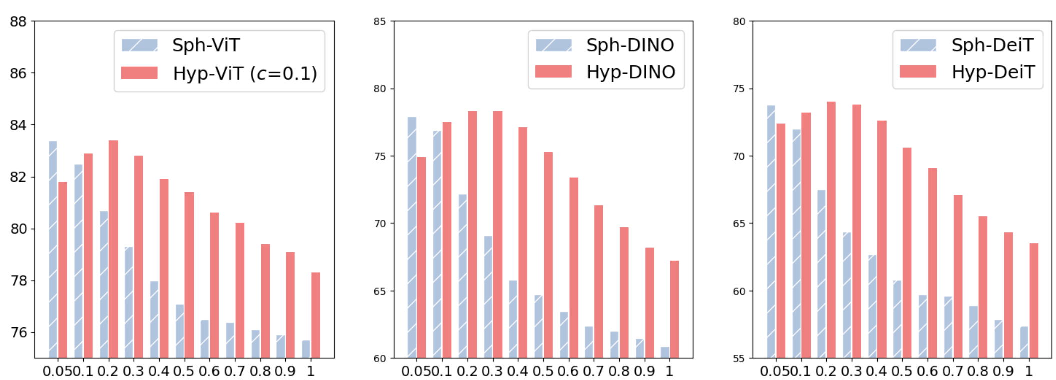

Figure 4. Recall of 1K metric comparison of models trained with different temperatures τ using CUB-200-2011 dataset. The x-axis indicates different τ. "Sph-" are versions with hypersphere embeddings optimized using D_{cos}, "Hyp-" are versions with hyperbolic embeddings optimized using D_{hyp}. "ViT", "DINO", "DeiT" indicates the pretraining for the vision transformer encoders. For "Hyp-" we fix the curvature parameter $c = 0.1$