Understanding Fine-tuning CLIP for Open-vocabulary Semantic Segmentation in Hyperbolic Space
-Supplementary Material-

Zelin Peng[1], Zhengqin Xu[1], Zhilin Zeng[1], Changsong Wen[1], Yu Huang[1],
Menglin Yang[2], Feilong Tang[3,4], and Wei Shen[1(✉)]
[1]MoE Key Lab of Artificial Intelligence, AI Institute, Shanghai Jiao Tong University
[2]Hong Kong University of Science and Technology (Guangzhou) [3]MBZUAI, [4]Monash University.

A. Derivation of the Definition

Theorem 1 (Scaling). According to the definitions of the classical Poincaré Ball model and the hyperbolic tangent, for a point $\mathbf{x} \in \mathbb{D}_c^n$ in hyperbolic space, its hyperbolic radius (i.e., hyperbolic induced norm) is defined as:

$$\mathrm{Rad}_{\mathbf{x}} := d_c^{\mathbb{D}}(\mathbf{x}, \mathbf{0}) = (\frac{2}{\sqrt{c}})\tanh^{-1}(\sqrt{c}\|\mathbf{x}\|), \tag{S-1}$$

where $\mathbf{0}$ is the center of the hyperbolic space. Let s is a scaling parameter, combine with the Mobiüs scalar multiplication operation described in Section 3.2 and Eq. S-1, the radius of the point $\mathbf{x}_s = \mathcal{T}_{\mathbf{x}}(s) := s \otimes_c \mathbf{x}$ which changed by the scaling transformation $\mathcal{T}_{\mathbf{x}}(s)$ is obtained by:

$$\begin{aligned} \mathrm{Rad}_{\mathbf{x},s}(s) &= \frac{2}{\sqrt{c}}\tanh^{-1}(\sqrt{c}\|\mathbf{x}_s\|) \\ &= \frac{2}{\sqrt{c}}\tanh^{-1}(\sqrt{c}\left\|\frac{\tanh(\frac{s\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}})\mathbf{x}}{\sqrt{c}\|\mathbf{x}\|}\right\|). \end{aligned} \tag{S-2}$$

Notably, $\mathrm{Rad}_{\mathbf{x},s}(s)$ satisfies the monotonicity property with respect to s, allowing the hyperbolic radius to be adjusted by varying s. Furthermore, according to the Möbius matrix multiplication operation defined in [1], a scaling matrix $\mathbf{S}$ can similarly modify the hyperbolic radius through the transformation $\mathcal{T}_{\mathbf{x}}(\mathbf{S})$.

Proof. In a hyperbolic space, considering a point $\mathbf{x} \in \mathbb{D}_c^n$ with hyperbolic radius $\mathrm{Rad}_{\mathbf{x}} = (2/\sqrt{c})\tanh^{-1}(\sqrt{c}\|\mathbf{x}\|)$, we scale the hyperbolic radius $\mathrm{Rad}_{\mathbf{x}}$ by utilizing the Möbius scalar multiplication operation to change the point $\mathbf{x}$ by a scaling parameter s, which can be formulated as:

$$\mathbf{x}_s = \mathcal{T}_{\mathbf{x}}(s) := s \otimes_c \mathbf{x}. \tag{S-3}$$

The hyperbolic radius $\mathrm{Rad}_{\mathbf{x},s}(s)$ of the changed point $\mathbf{x}_s$ is obtained:

$$\begin{aligned}
\mathrm{Rad}_{\mathbf{x},s}(s) &= \frac{2}{\sqrt{c}}\tanh^{-1}\left(\sqrt{c}\|\mathbf{x}_s\|\right) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}\left(\sqrt{c}\left\|(1/\sqrt{c})\tanh\left(s\tanh^{-1}\left(\sqrt{c}\|\mathbf{x}\|\right)\right)\frac{\mathbf{x}}{\|\mathbf{x}\|}\right\|\right) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}(\sqrt{c}\left\|(1/\sqrt{c})\tanh\left(s\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)\frac{\mathbf{x}}{\|\mathbf{x}\|}\right\|) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}(\|\tanh\left(s\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)\frac{\mathbf{x}}{\|\mathbf{x}\|}\|) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}\left(\frac{\tanh\left(s\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)}{\|\mathbf{x}\|}\|\mathbf{x}\|\right) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}(\tanh\left(s\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)) \\
&= s\mathrm{Rad}_{\mathbf{x}}.
\end{aligned} \tag{S-4}$$

The $\mathrm{Rad}_{\mathbf{x},s}(s)$ satisfies the monotonicity criteria of s. For a scaling matrix $\mathbf{S}$, according to the relation definition of Möbius matrix multiplication in [1], the hyperbolic radius $\mathrm{Rad}_{\mathbf{x},\mathbf{S}}(\mathbf{S})$ is obtained as:

$$\begin{aligned}
\mathrm{Rad}_{\mathbf{x},\mathbf{S}}(\mathbf{S}) &= \frac{2}{\sqrt{c}}\tanh^{-1}\left(\sqrt{c}\|\mathcal{T}_{\mathbf{x}}(\mathbf{S})\|\right) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}\left(\sqrt{c}\|\mathbf{S}\otimes_c\mathbf{x}\|\right) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}\left(\sqrt{c}\left\|(1/\sqrt{c})\tanh\left(\frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|}\tanh^{-1}\left(\sqrt{c}\|\mathbf{x}\|\right)\right)\frac{\mathbf{S}\mathbf{x}}{\|\mathbf{S}\mathbf{x}\|}\right\|\right) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}(\sqrt{c}\left\|(1/\sqrt{c})\tanh\left(\frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|}\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)\frac{\mathbf{S}\mathbf{x}}{\|\mathbf{S}\mathbf{x}\|}\right\|) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}(\|\tanh\left(\frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|}\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)\frac{\mathbf{S}\mathbf{x}}{\|\mathbf{S}\mathbf{x}\|}\|) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}\left(\frac{\tanh\left(\frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|}\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)}{\|\mathbf{S}\mathbf{x}\|}\|\mathbf{S}\mathbf{x}\|\right) \\
&= \frac{2}{\sqrt{c}}\tanh^{-1}(\tanh\left(\frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|}\frac{\sqrt{c}}{2}\mathrm{Rad}_{\mathbf{x}}\right)) \\
&= \frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|}\mathrm{Rad}_{\mathbf{x}}.
\end{aligned} \tag{S-5}$$

It is noticed that $\frac{\|\mathbf{S}\mathbf{x}\|}{\|\mathbf{x}\|}$ is a scalar, and thus the $\mathrm{Rad}_{\mathbf{x},\mathbf{S}}(\mathbf{S})$ also satisfies the monotonicity criteria *w.r.t* $\mathbf{S}$. Consequently, according to Eq. (S-4) and Eq. (S-5), both the scaling scalar s and the scaling matrix $\mathbf{S}$ can directly adjust the hyperbolic radius $\mathrm{Rad}_{\mathbf{x}}$ of the point $\mathbf{x}$ via the scaling transformation $\mathcal{T}(\cdot)$. ■

B. Discussion with HyperLoRA [2]

Although both our approach and HyperLoRA [2] operate in hyperbolic space, HyperCLIP distinguishes itself from HyperLoRA in three key aspects. (1) The motivations are entirely different. HyperLoRA aims to modify the original hierarchical structure, enabling the model to adapt from generalized domains to specific domains. In contrast, our method preserves the original hierarchical structure and only adjusts its hierarchical level to enhance fine-grained capabilities within the same domain. (2) The analytic models for constructing hyperbolic space differ significantly. HyperLoRA uses the Lorentz model, which defines an open hyperbolic space for easy adaptation to target tasks, while we employ the Poincaré ball model, embedding hyperbolic space within an n-dimensional unit sphere with a fixed boundary. This makes the Poincaré ball model better suited for preserving the original hierarchical structure while adjusting the hierarchical level from image-level to pixel-level. (3) The operations in analytic models differ fundamentally. HyperLoRA employs Lorentz transformations to easily modify the original hierarchical structure for domain adaptation. In contrast, we utilize gyrovector space operations, specifically Möbius operations, to ensure that added vectors remain within the original Poincaré ball, thereby preserving the original hierarchical structure in the Poincaré ball without distortion. Experimental results show our method introduces fewer trainable parameters (5.6M vs. 7.5M) and achieves better performance (as shown in the Table 5), compared with HyperLoRA, which verify the significance of the differences.

References

[1] Valentin Khrulkov, Leyla Mirvakhabova, Evgeniya Ustinova, Ivan Oseledets, and Victor Lempitsky. Hyperbolic image embeddings. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 6418–6428, 2020. 1, 2

[2] Menglin Yang, Aosong Feng, Bo Xiong, Jihong Liu, Irwin King, and Rex Ying. Hyperbolic fine-tuning for large language models. *arXiv preprint arXiv:2410.04010*, 2024. 2