Lorentz Direct Concatenation for Stable Training in Hyperbolic Neural Networks

NEURAL INFORMATION PROCESSING SYSTEMS

Eric Zhonghang Qu [1] Dongmian Zou [1]

[1]Division of Natural and Applied Sciences, Duke Kunshan University.

昆山杜克大学 DUKE KUNSHAN UNIVERSITY

Duke UNIVERSITY

Introduction

We proposed the Lorentz direct concatenation that concatenates vectors in the Lorentz space and illustrate that it is much more stable than concatenating in the tangent space. We provide some insights and show superiority of performing direct concatenation in real tasks.

Methods

Suppose the input vectors are $\{\boldsymbol{x}_i\}_{i=1}^N$ where $\boldsymbol{x}_i \in \mathbb{L}_K^{n_i}$. Goal: concat them into $\boldsymbol{y} \in \mathbb{L}_K^M$ where $M = \Sigma_{i=1}^N n_i$.

Lorentz Tangent Concatenation

1. Logarithmic map $\boldsymbol{x}_i$ to the tangent space of the origin $\boldsymbol{o}$: $\boldsymbol{v}_i = \log_{\boldsymbol{o}}^K(\boldsymbol{x}_i) = \left[\begin{smallmatrix} v_{i_t} \\ \boldsymbol{v}_{i_s} \end{smallmatrix}\right] \in \mathbb{R}^{n_i+1}$
2. Perform the Euclidean concat to get $\boldsymbol{v} := \left(0, \boldsymbol{v}_{1_s}^\top, \ldots, \boldsymbol{v}_{N_s}^\top\right)^\top$.
3. Exponential map it back as $\boldsymbol{y} = \exp_{\boldsymbol{o}}^K(\boldsymbol{v}) \in \mathbb{L}_K^M$

Lorentz Direct Concatenation

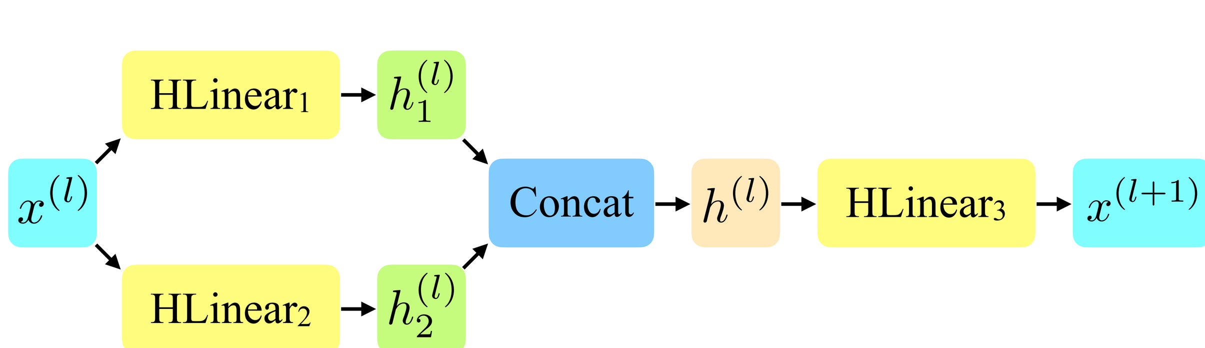
$$\boldsymbol{y} = \mathrm{HCat}(\{\boldsymbol{x}_i\}_{i=1}^N) = \left[\sqrt{\sum_{i=1}^N x_{i_t}^2 + \frac{N-1}{K}}, \boldsymbol{x}_{1_s}^\top, \cdots, \boldsymbol{x}_{N_s}^\top\right]^\top.$$

where $\boldsymbol{x}_{i_s}$ is the spatial component. Essentially, we direct concat the $\boldsymbol{x}_{i_s}$ in the Euclidean space and view it as $\boldsymbol{y}_s$.

Numerical Stability

We design a hyperbolic neural network consists of a cascading of L blocks as follows: for $l = 0, \cdots, L-1$,

$$h_1^{(l)} = \mathrm{Hlinear}_{d,d}(x^{(l)}), \qquad h_2^{(l)} = \mathrm{Hlinear}_{d,d}(x^{(l)});$$
$$h^{(l)} = \mathrm{HCat}(h_1^{(l)}, h_2^{(l)}); \qquad x^{(l+1)} = \mathrm{Hlinear}_{2\times d,d}(h^{(l)}).$$



- Input/Output data: sampled from two wrapped normal dist.
- The average gradient norm of the HLinear in each block is recorded, and shown in Figure 1.
- The increased norm is due to Log/Exp operations.
- Large norm ⇒ Far from origin ⇒ Unstable Float Repr.

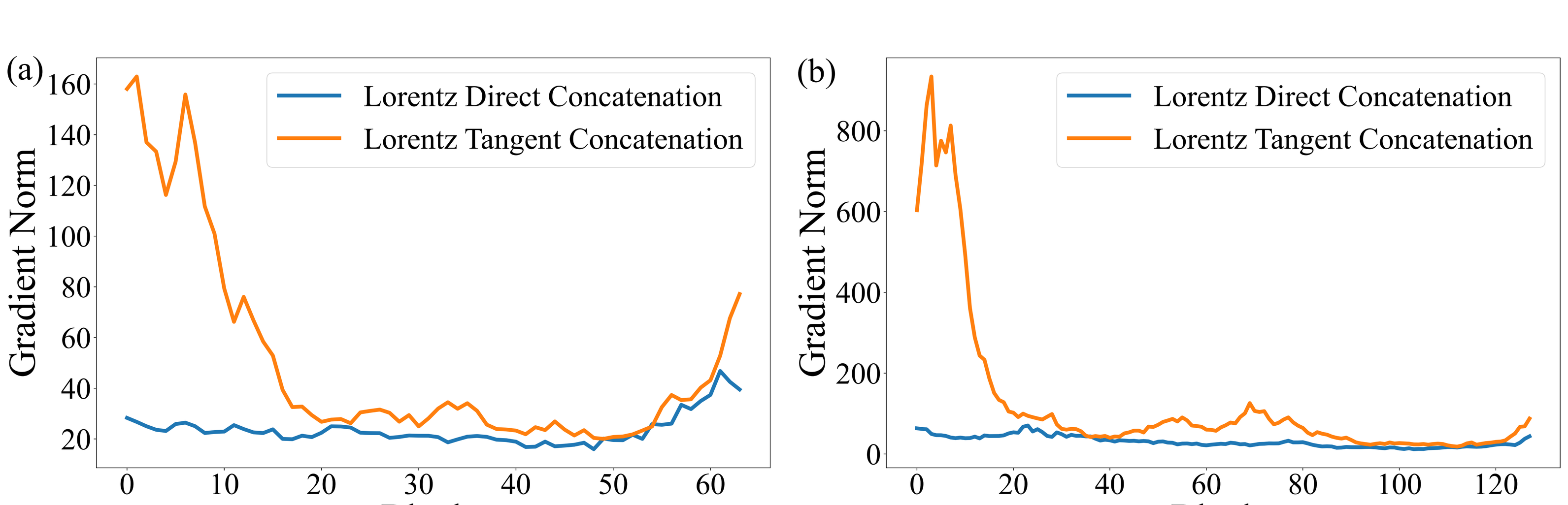


Figure 1. Average gradient norm in training. (a) 64 blocks. (b) 128 blocks.

Acknowledgement

The research results of this article are sponsored by the Kunshan Municipal Government research funding.

Hyperbolic Distances

- We study how concatenation influence pairwise distance.
- Given $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{c} \in \mathbb{L}_K^n$
- Let $\boldsymbol{x}_c = \mathrm{HCat}(\boldsymbol{x}, \boldsymbol{c})$, $\boldsymbol{y}_c = \mathrm{HCat}(\boldsymbol{y}, \boldsymbol{c})$ be Direct Concat
- Let $\boldsymbol{x}_c' = \mathrm{HTCat}(\boldsymbol{x}, \boldsymbol{c})$, $\boldsymbol{y}_c' = \mathrm{HTCat}(\boldsymbol{y}, \boldsymbol{c})$ be Tangent Concat
- We hope $d_{\mathcal{L}}(\boldsymbol{x}_c, \boldsymbol{y}_c)$ / $d_{\mathcal{L}}(\boldsymbol{x}_c', \boldsymbol{y}_c')$ be close to $d_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{y})$.
- In the experiment, we randomly sample $\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{c}$ from wrapped noraml distribution in different dimensions and report $|d_{\mathcal{L}}(\boldsymbol{x}_c, \boldsymbol{y}_c) - d_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{y})|$ and $|d_{\mathcal{L}}(\boldsymbol{x}_c', \boldsymbol{y}_c') - d_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{y})|$ as well as their differences in Figure 2.

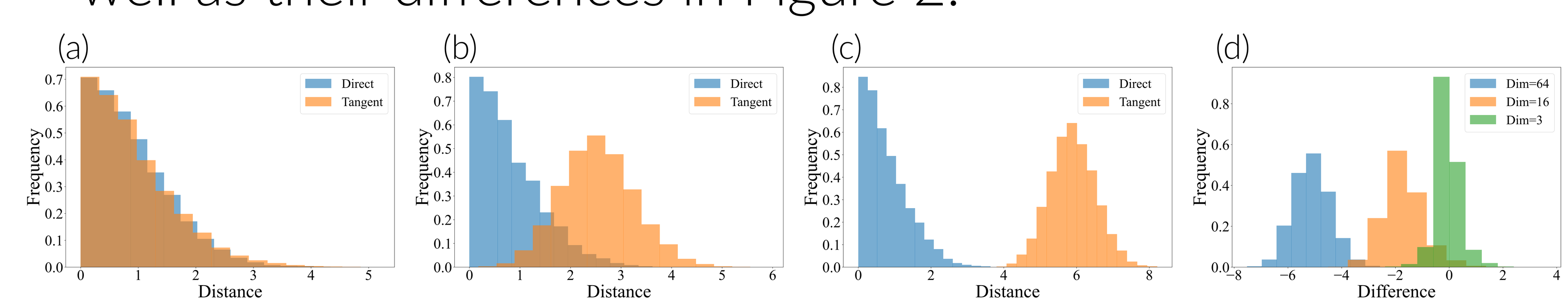


Figure 2. (a, b, c): Difference between concated and original distances with $n = 3, 6, 64$, respectively. (d): $|d_{\mathcal{L}}(\boldsymbol{x}_c, \boldsymbol{y}_c) - d_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{y})| - |d_{\mathcal{L}}(\boldsymbol{x}_c', \boldsymbol{y}_c') - d_{\mathcal{L}}(\boldsymbol{x}, \boldsymbol{y})|$.

Applications

We compare the performance of Direct / Tangent concatenation on the HAEGAN framework as illustrated in Figure 3.

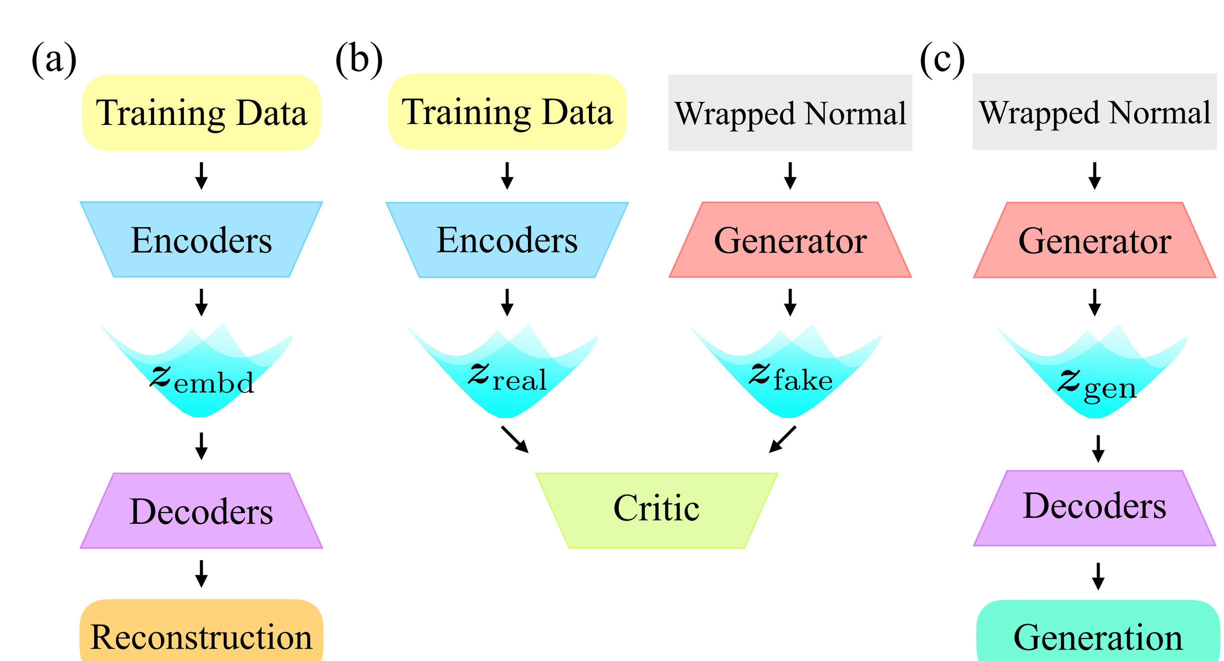


Figure 3. The HAEGAN framework.

We use the HAEGAN with a tree AE on random trees to test concatenation methods:

Table 1. Results of the tree generation experiments.

Concat	MMD		Average Difference			Time
	Degree	Orbit	Orbit	Betweenness	Closeness	(s/step)
Beta	0.000470	**0.000001**	0.129896	0.026102	0.022375	1.7529
Tangent	0.000314	0.000052	0.131563	0.024171	0.021858	1.6385
Direct	**0.000156**	0.000005	**0.123286**	**0.023706**	**0.021740**	1.3146

We also test molecular generation on the MOSES dataset with a hyperbolic version of JTVAE as the AE:

Table 2. Results of the molecular generation experiments.

Concat	Validity (↑)	Unique (↑)	Novelty (↑)	SNN (↑)	Scaf (↑)
Beta	NaN	NaN	NaN	NaN	NaN
Tangent	NaN	NaN	NaN	NaN	NaN
Direct	1.0±0.0	1.0±0.0	0.905±0.006	0.593±0.002	0.113±0.007

Our Related Works

Our paper HAEGAN HKConv