Searching for Actions on the Hyperbole

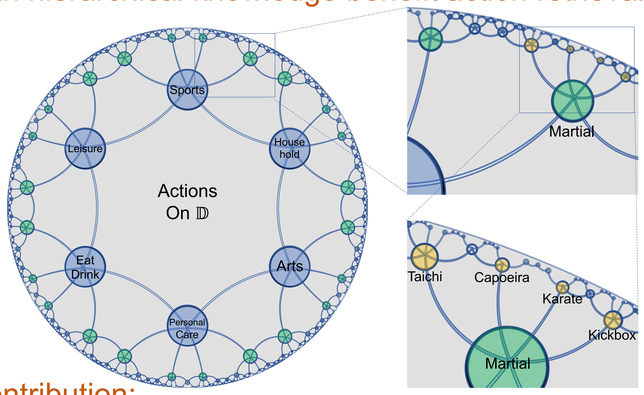
Teng Long, Pascal Mettes, Heng Tao Shen, Cees Snoek

UNIVERSITY OF AMSTERDAM

CVPR SEATTLE WASHINGTON JUNE 16-18 2020

Problem

Can hierarchical knowledge benefit action retrieval?



Contribution:

I. We position action hierarchy with discriminative hyperbolic embeddings
II. We propose hyperbolic matching between actions and videos on the hyperbole $\mathbb{D}$
III. We perform hierarchical action search by name/video, on supervised/zero-shot setting.

Preliminary of $\mathbb{D}$

Mobius Addition $\boldsymbol{a} \oplus_c \boldsymbol{b}$:

The vector addition in $\mathbb{D}$ with curvature c

$$\frac{\left(1+2c\langle \mathbf{a},\mathbf{b}\rangle + c\|\mathbf{b}\|^2\right)\mathbf{a} + \left(1-c\|\mathbf{a}\|^2\right)\mathbf{b}}{1+2c\langle \mathbf{a},\mathbf{b}\rangle + c^2\|\mathbf{a}\|^2\|\mathbf{b}\|^2}$$

Exponential Map $\exp_{\mathbf{x}}(\mathbf{v})$:

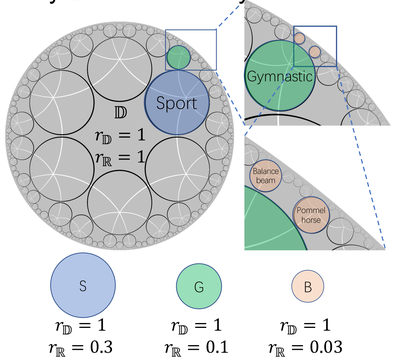
The transform from $\mathbb{R}$ to $\mathbb{D}$

$$\mathbf{x} \oplus_c \left(\tanh\left(\sqrt{c}\frac{\lambda_x^c\|\mathbf{v}\|}{2}\right)\frac{\mathbf{v}}{\sqrt{c}\|\mathbf{v}\|}\right)$$

Conformality: angular invariance for $\mathbb{D}/\mathbb{R}$

$$\frac{g_x^{\mathbb{D}}(u,v)}{\sqrt{g_x^{\mathbb{D}}(u,u)}\sqrt{g_x^{\mathbb{D}}(v,v)}} = \frac{\langle u,v\rangle}{\|u\|\|v\|}$$

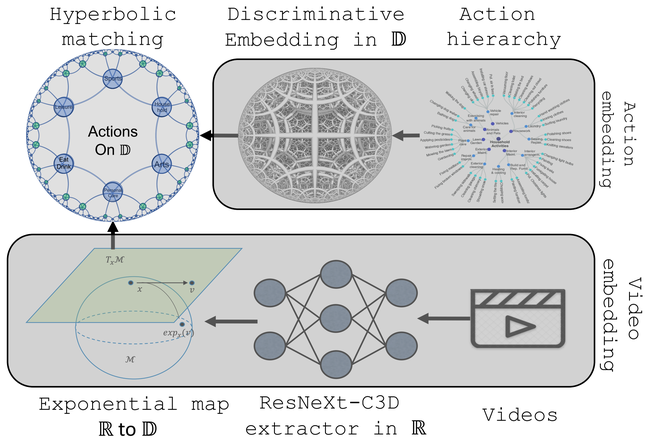
Why $\mathbb{D}$ suits hierarchy?



$\mathbb{D}$: Area grows exponentially to r

Tree: #nodes grows exponentially to depth

Model



Action hierarchy embedding:

$\mathcal{L}_H$ preserve the parent-child relationship
$\mathcal{L}_P$ separate leaf node in $\mathbb{D}$
$\mathcal{L}_2$ further construct entailment cone

$$\mathcal{L}_1(\mathcal{P},\mathcal{N},\mathbf{P}) = \mathcal{L}_H(\mathcal{P},\mathcal{N}) + \lambda\cdot\mathcal{L}_S(\mathbf{P})$$

$$\mathcal{L}_2 = \sum_h E(\mathbf{u},\mathbf{v}) + \sum_{\neg h}\left(\gamma - E\left(\mathbf{u}',\mathbf{v}'\right)\right)^+$$

Matching videos to actions:

Softmax classifier based on hyperbolic distance

$$d_c(\mathbf{a},\mathbf{b}) := \frac{2}{\sqrt{c}}\operatorname{arctanh}\left(\sqrt{c}\|-\mathbf{b}\oplus_c\mathbf{a}\|\right)$$

$$p = \frac{\exp\left(-d_c\left(\Psi_e(v;\theta),\phi_c(k)\right)\right)}{\sum_{k'}\exp\left(-d_c\left(\Psi_e(v;\theta),\phi_c\left(k'\right)\right)\right)}$$

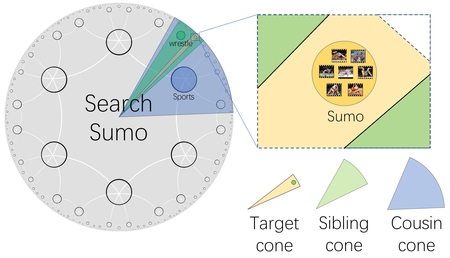
Riemannian optimization:

Gradient Descent based on Riemannian gradient ∇_R

$$\mathbf{P}_{t+1} = \mathbf{P}_t - \eta_t\nabla_R\mathcal{L}\left(\mathbf{P}_t\right)$$

Search

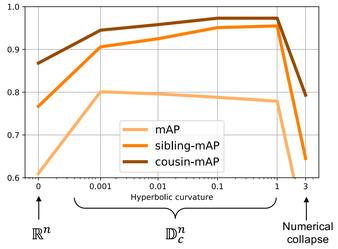
Cosine distance is a fast approximation for distance ranking in $\mathbb{D}$



$$d_{\mathbf{q}}\left(\mathbf{x}_i\right) = 1 - \cos\left(\mathbf{q},\mathbf{x}_i\right)$$

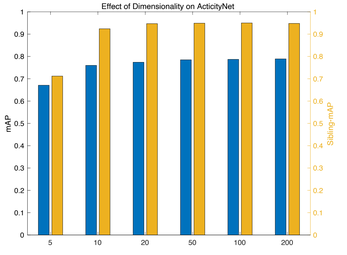
Code:

Results



Effect of curvature

$c \to 0$: degrade into $\mathbb{R}$
$c \to \infty$: extremely curved $\mathbb{D}$, numerical collapse



Effect of dimensionality

d $\to 20$: performance saturated at low dimensionality

$\mathbb{D}$ v.s. Δ^n, $\mathbb{R}^n$, $\mathbb{S}^{n-1}$:

Hyperbolic space suits hierarchical search while Benefit plain search too. Other spaces show little.

	space	H-ActivityNet mAP	H-ActivityNet S-mAP	H-Kinetics mAP	H-Kinetics S-mAP	H-Moments mAP	H-Moments S-mAP
ResNextC3D [18]	Δ^n	0.592	0.761	0.532	0.733	0.145	0.173
DeViSE [14]	$\mathbb{R}^n$	0.609	0.761	0.553	0.715	0.134	0.161
Li *et al.* [22]	Δ^n	0.583	0.760	0.552	0.753	-	-
Mettes *et al.* [25]	$\mathbb{S}^{n-1}$	0.587	0.760	0.551	0.754	0.142	0.172
Barz and Denzler [4]	$\mathbb{S}^{n-1}$	0.583	0.747	0.547	0.725	0.143	0.172
This Paper	$\mathbb{R}^n$	0.610	0.767	0.565	0.738	0.143	0.172
This Paper	$\mathbb{D}_c^n$	**0.678**	**0.843**	**0.593**	**0.824**	**0.163**	**0.201**

Qualitative results:

(a) Search results.

(b) Failure cases.

Zero-shot generalization:

	H-ActivityNet mAP	H-ActivityNet S-mAP	H-Moments mAP	H-Moments S-mAP
Zhang *et al.* [42]	0.397	0.449	0.026	0.027
Li *et al.* [22]	0.389	0.461	-	-
Mettes *et al.* [25]	0.235	0.281	0.169	0.171
Barz and Denzler [4]	0.453	0.527	0.216	0.219
This Paper	**0.543**	**0.627**	**0.222**	**0.225**

$\mathbb{D}$ not only benefit supervised search, but also promote zero-shot search.