Semantic Alignment in Hyperbolic Space for Open-Vocabulary Semantic Segmentation

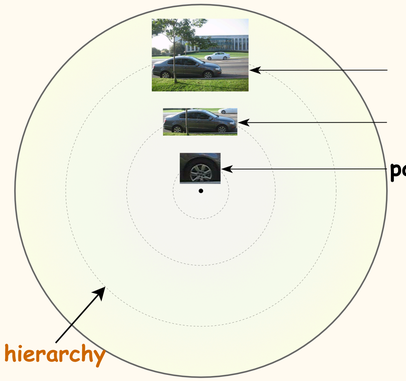
Hoang M. Truong Hai Nguyen-Truong Dang Huynh

Fulbright University Việt Nam

tmhoanggg.github.io/HyRo

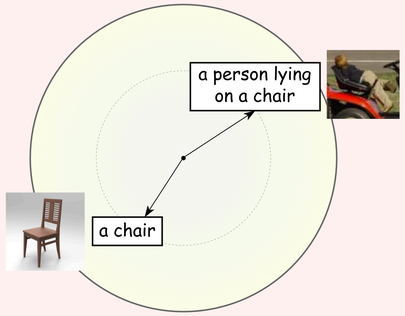
CVPR JUNE 3-7, 2026 DENVER COLORADO

Why Hyperbolic?



- Open-vocabulary segmentation must bridge image-level from CLIP and pixel-level prediction
- The task is hierarchical: scene → object → part/pixel, while text concepts are often more abstract
- Hyperbolic space naturally represents hierarchy, radial distance captures granularity.

Gap in prior hyperbolic adaption



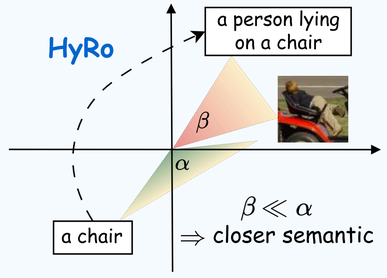
Radius only (hierarchy aligned)

Semantics still ambiguous

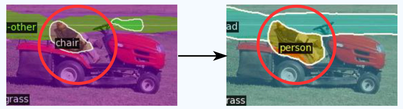
Same level ≠ same meaning

Failure: "a person lying on a chair" gets confused with "a chair"

Our method



- HyRo decouples radius (hierarchy) and angle (semantics) in hyperbolic space
- Theoretically proven to preserve hierarchical structure while correcting semantic misalignment



Quantitative Results & Ablation Study

Model	VLM	Additional Backbone	Fine-tuning Space	A-847	PC-459	A-150	PC-59	PAS-20	PAS-20^b
ZS3Net [2]	-	ResNet-101	E	-	-	-	19.4	38.3	-
LSeg [21]	CLIP ViT-B/32	ResNet-101	E	-	-	-	-	47.4	-
ZegFormer [12]	CLIP ViT-B/16	ResNet-101	E	4.9	9.1	16.9	42.8	86.2	62.7
ZSseg [39]	CLIP ViT-B/16	ResNet-101	E	7.0	-	20.5	47.7	88.4	-
OpenSeg [16]	ALIGN	ResNet-101	E	4.4	7.9	17.5	40.1	-	63.8
OVSeg [22]	CLIP ViT-B/16	ResNet-101c	E	7.1	11.0	24.8	53.3	92.6	-
ZegCLIP [46]	CLIP ViT-B/16	-	E	-	-	-	41.2	93.6	-
SED [37]	CLIP ConvNeXt-B	-	E	11.4	18.6	31.6	**57.3**	94.4	-
SAN [41]	CLIP ViT-B/16	Side Adapter	E	10.1	12.6	27.5	53.8	94.0	-
HyperCLIP* [31]	CLIP ViT-B/16	-	H	11.9	18.2	**31.7**	57.1	94.9	**77.1**
HyRo (Ours)	CLIP ViT-B/16	-	H	**12.0**	**18.9**	31.2	**57.3**	**95.0**	76.7

Radius	Rotation	A-847	PC-459	A-150	PC-59	PAS-20	PAS-20^b
✗	✗	11.4	17.6	29.8	56.2	94.8	75.9
✓	✗	11.9	18.2	**31.7**	57.1	94.9	76.4
✗	✓	11.6	18.3	30.6	56.5	**95.4**	**76.7**
✓	✓	**12.0**	**18.9**	31.2	**57.3**	95.0	**76.7**

Attention Visualization

(a) Image (b) Without HyRo (c) With HyRo (Ours)

Qualitative Results

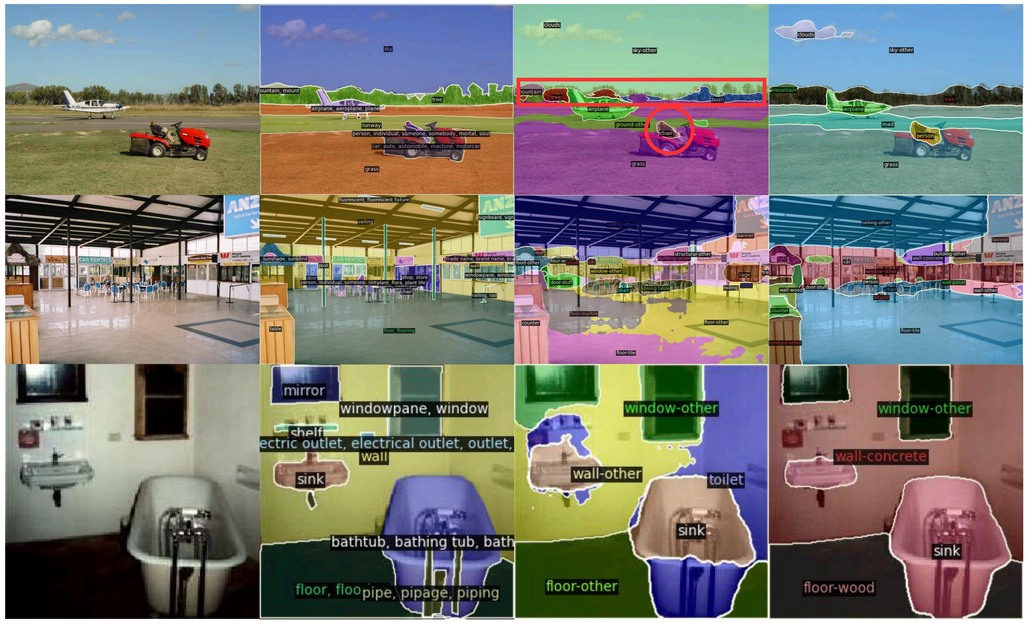


Image Ground truth HyperCLIP HyRo (Ours)

⊗c Mobius Multiplication Exponential Map Logarithmic Map S_k Radius Scaling Matrices R_k Orthogonal Rotation Matrices