OpenHype: Hyperbolic Embeddings for Hierarchical Open-Vocabulary Radiance Fields

Understanding 3D scenes requires modeling their inherent semantic hierarchies, where objects are composed of parts and grouped into larger contextual structures. While Neural Radiance Fields (NeRFs) [1] have emerged as a powerful representation for novel view synthesis and scene understanding, existing open-vocabulary methods often treat scenes as flat collections of elements. This limits their ability to reason about multi-scale semantics, handle compositional queries, and generalize to diverse real-world hierarchies. Current hierarchical NeRF approaches either rely on multiple rendering passes—leading to high inference cost—or impose fixed, discrete hierarchies that fail to capture the fluid structure of real-world environments.

We propose OpenHype, a novel framework that leverages hyperbolic geometry to encode scene hierarchies in a continuous latent space. By mapping language-aligned features into the Lorentz model of hyperbolic space, OpenHype naturally embeds hierarchical relationships along geodesic paths: object parts are placed near the boundary, whole objects lie closer to the origin, and higher-level scene groupings emerge along shared geodesics. This design enables continuous hierarchy traversal without predefined levels or additional rendering passes. At inference time, open-vocabulary queries can flexibly navigate these embeddings to segment objects and parts at multiple granularities.

We train OpenHype in two stages: (1) a hyperbolic auto-encoder maps CLIP [2] features from Semantic-SAM [3] mask crops into a continuous hierarchical space, and (2) a NeRF is trained to reconstruct both the scene appearance (color and density) and the corresponding hyperbolic embeddings. During inference, hierarchical traversal along geodesics in hyperbolic space produces multi-scale semantic responses, aggregated through a softmax-weighted scheme for robust open-vocabulary segmentation.

We evaluate OpenHype on the Search3D [4] benchmark and the LERF [5, 6] dataset. On Search3D, OpenHype improves mean IoU by +8.9 and accuracy by +12.1 on challenging part-level queries compared to state-of-the-art baselines (LERF [5], OpenNeRF [7], LangSplat [6]). On LERF, our method surpasses recent discrete-hierarchy approaches, including N2F2 [8] with an overall Iou of 54.6. Qualitative results further show superior segmentation of object parts and robustness to compositional queries, while ablation studies confirm the effectiveness of hyperbolic supervision, feature extrapolation, and geodesic aggregation.

In summary, OpenHype introduces continuous hierarchical vision-language embeddings for NeRFs, enabling efficient, flexible, and accurate open-vocabulary 3D scene understanding. Beyond NeRF, our formulation is compatible with alternative 3D representations such as Gaussian Splatting, opening avenues for broader integration of hyperbolic embeddings in 3D vision.

References

[1] Ben Mildenhall et al. "NeRF: Representing Scenes as Neural Radiance Fields for View Synthesis". In: *ECCV*. 2020.

[2] Alec Radford et al. "Learning transferable visual models from natural language supervision". In: *ICML*. 2021.

[3] Feng Li et al. "Segment and Recognize Anything at Any Granularity". In: *ECCV*. 2024.

[4] Ayca Takmaz et al. "Search3d: Hierarchical open-vocabulary 3d segmentation". In: *IEEE Robotics and Automation Letters* (2025).

[5] Justin Kerr et al. "Lerf: Language embedded radiance fields". In: *CVPR*. 2023.

[6] Minghan Qin et al. "Langsplat: 3d language gaussian splatting". In: *CVPR*. 2024.

[7] Francis Engelmann et al. "OpenNeRF: Open Set 3D Neural Scene Segmentation with Pixel-Wise Features and Rendered Novel Views". In: *ICLR*. 2024.

[8] Yash Bhalgat et al. "N2F2: Hierarchical Scene Understanding with Nested Neural Feature Fields". In: *ECCV*. 2024.