H^2M-Net: Hierarchical Hyperbolic Memory Network for Pathological Report Generation

Shuilian Yao, Qi Jia*, Shirui Li, Yu Liu, and Xin Fan

Dalian University of Technology, Dalian, China
shuilian_yao@mail.dlut.edu.cn,{jiaqi,liuyu8824,xin.fan}@dlut.edu.cn

Abstract. Automatically generating pathology reports from whole slide images (WSIs) is a critical yet highly challenging task, aiming to transform gigapixel-level visual information into structured diagnostic text. WSIs inherently exhibit hierarchical semantic characteristics, spanning from local cellular morphology to global tissue architecture. However, existing methods predominantly rely on Euclidean space modeling and treat multi-scale features as flat representations, which limits their ability to effectively capture the intrinsic hierarchical structure of pathological data, thereby affecting the logical coherence of generated reports. To address this limitation, we propose H^2M-Net, a hyperbolic geometry-based hierarchical generation framework that explicitly models the hierarchical relationships between visual and textual modalities in the Lorentz space. Specifically, we embed both pathological images and report texts into a hyperbolic manifold, and introduce a radial position loss together with an entailment loss to enforce geometric and semantic constraints between parent and child nodes. On this basis, we construct a bottom-up cascaded generation architecture that consists of a visual encoder, memory encoder, and text decoder to enable cross-level information propagation. Experimental results demonstrate that our method outperforms current state-of-the-art approaches. The code is available at here.

Keywords: Whole Slide Image · Report Generation · Hyperbolic Space.

1 Introduction

To accurately support clinical decision-making and effectively alleviate the workload of pathologists, pathology report generation from whole slide images (WSIs), which are regarded as the gold standard for disease diagnosis, has become a research hotspot in medical artificial intelligence [3,11,26].

Distinguished from conventional medical report generation tasks [13,25], WSIs feature gigapixel-level resolutions and inherently hierarchical visual and diagnostic structures. A standardized pathology report follows a rigorous hierarchical workflow: it begins by identifying fine-grained morphological features such as nuclear atypia and mitotic figures at the microenvironment scale, integrates these into regional histological patterns such as glandular structures and tumor

* Corresponding author, jiaqi@dlut.edu.cn.

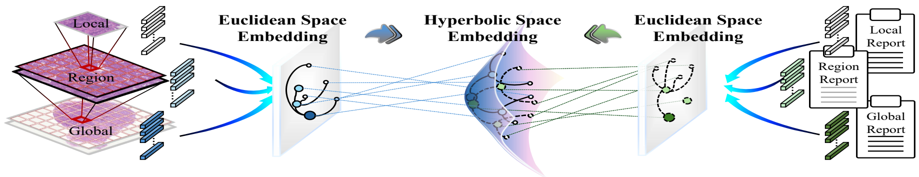

Fig. 1. Illustration of the Euclidean space embedding and hyperbolic space embedding.

stroma, and ultimately derives global diagnostic conclusions including cancer grading and subtype determination based on multi-scale information [4,28].

However, current research paradigms still predominantly rely on Euclidean vector spaces for feature representation[3,11,26,21,27,14,24]. Due to the inherent linear properties of Euclidean geometry, it suffers from mapping distortion when representing highly branched, tree-like hierarchical structures [12,2]. Consequently, models struggle to maintain logical consistency between micro-morphological descriptions and high-level diagnostic conclusions, leading to a semantic disconnection across different hierarchical levels, as shown in Fig. 1.

To overcome the above limitations, we argue that the geometric structure of the feature space must align with the inherent hierarchical organization of pathological data. As a non-Euclidean geometry with constant negative curvature, hyperbolic geometry has the unique property of exponential volume expansion with radius, which has been proven to be an ideal mathematical model for embedding complex hierarchies [7,19,15,17,8,9,18,10]. By mapping the natural hierarchy of WSIs onto a hyperbolic manifold, the core local-region-global inclusion relationships are naturally preserved, providing a geometrically congruent foundation for hierarchical diagnostic reasoning.

Based on this insight, we propose a hyperbolic hierarchical memory network (H^2M-Net). Unlike traditional flat Euclidean modeling approaches, we adopt the Lorentz model in hyperbolic geometry to construct a hierarchy-aware embedding space. We introduce a memory mechanism as a bridge across multi-level embeddings, aiming to progressively extract diagnostic information from local image patches and propagate it to the global representation level. Furthermore, to ensure the effectiveness of the hyperbolic hierarchical structure, we design strict geometric constraint losses: the entailment loss enforces the inclusion relationship between parent and child nodes via entailment cones, while the radial position loss precisely aligns feature hierarchy with their radial distances from the manifold origin. The main contributions are summarized as follows:

- We propose H^2M-Net, a hyperbolic hierarchical pathology report generation model, which introduces the non-Euclidean modeling capability of hyperbolic geometry into the pathology report generation task.
- We design a hierarchical memory module that constructs a transmission bridge across multi-levels, ensuring that diagnostic information is preserved during the progressive abstraction from local to global levels.

– Experimental results on public benchmark datasets demonstrate that the proposed method achieves state-of-the-art performance.

2 Method

As illustrated in Fig. 2, we first organize pathological images and report texts into hierarchical structural representations and embed them into the hyperbolic Lorentz space. The hierarchical semantic relationships are explicitly constrained by a radial position loss and an entailment loss. On this basis, we design a network unit to progressively generate reports in a bottom-up manner, where a memory encoder is introduced to aggregate and propagate cross-level information.

2.1 Preliminaries: Lorentz Model

We adopt the n-dimensional Lorentz model $\mathbb{H}^n_\rho$ with curvature $-\rho$ ($\rho > 0$) to represent the hyperbolic space. This model is defined as a manifold embedded in the $(n+1)$-dimensional Minkowski space $\mathbb{R}^{n,1}$.

For two vectors $\mathbf{u}, \mathbf{v} \in \mathbb{R}^{n+1}$, the Lorentzian inner product is defined as $\langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}} = -u_0 v_0 + \sum_{i=1}^{n} u_i v_i$. The hyperbolic manifold corresponds to the upper sheet of a two-sheeted hyperboloid: $\mathbb{H}^n_\rho = \{\mathbf{u} \in \mathbb{R}^{n+1} \mid \langle \mathbf{u}, \mathbf{u} \rangle_{\mathcal{L}} = -1/\rho, u_0 > 0\}$. The geodesic distance between $\mathbf{u}, \mathbf{v} \in \mathbb{H}^n_\rho$ is:$d_{\mathcal{L}}(\mathbf{u}, \mathbf{v}) = \frac{1}{\sqrt{\rho}} \operatorname{arcosh}\left(-\rho \langle \mathbf{u}, \mathbf{v} \rangle_{\mathcal{L}}\right)$.

In practical WSI modeling, the visual encoder extracts Euclidean features $\hat{\mathbf{z}} \in \mathbb{R}^n$. To map these into the hyperbolic space, we first regard them as vectors in the tangent space at the origin $\mathbf{o} = (\sqrt{1/\rho}, 0, \ldots, 0)$ by setting the 0-th dimension to zero, i.e., $\mathbf{z} = (0, \hat{\mathbf{z}}) \in T_{\mathbf{o}}\mathbb{H}^n_\rho$. They are then mapped onto the manifold via the exponential map $\exp_{\mathbf{o}}(\mathbf{z}) : T_{\mathbf{o}}\mathbb{H}^n_\rho \to \mathbb{H}^n_\rho$:

$$\exp_{\mathbf{o}}(\mathbf{z}) = \left(\frac{1}{\sqrt{\rho}} \cosh(\sqrt{\rho} \|\mathbf{z}\|_{\mathcal{L}}), \frac{\sinh(\sqrt{\rho} \|\mathbf{z}\|_{\mathcal{L}})}{\sqrt{\rho} \|\mathbf{z}\|_{\mathcal{L}}} \mathbf{z} \right), \tag{1}$$

where $\|\mathbf{z}\|_{\mathcal{L}} = \sqrt{\langle \mathbf{z}, \mathbf{z} \rangle_{\mathcal{L}}}$.

2.2 Hierarchical Geometric Constraints

We decompose pathological images and reports into hierarchical representations, map them into the Lorentz space, and impose geometric constraints to preserve their inherent taxonomy.

Hierarchical Aggregation. For pathological images, each WSI is divided into non-overlapping patches at the local level. We perform attention-based aggregation in the Euclidean tangent space to progressively form region and global level representations. Specifically, each patch feature is denoted as $\mathbf{z}_i \in \mathbb{R}^d$ with spatial coordinate (x_i, y_i). We group $M \times M$ neighboring patches into a region $\{\mathbf{z}_1, \ldots, \mathbf{z}_{M^2}\}$, and apply attention-based aggregation to obtain the region

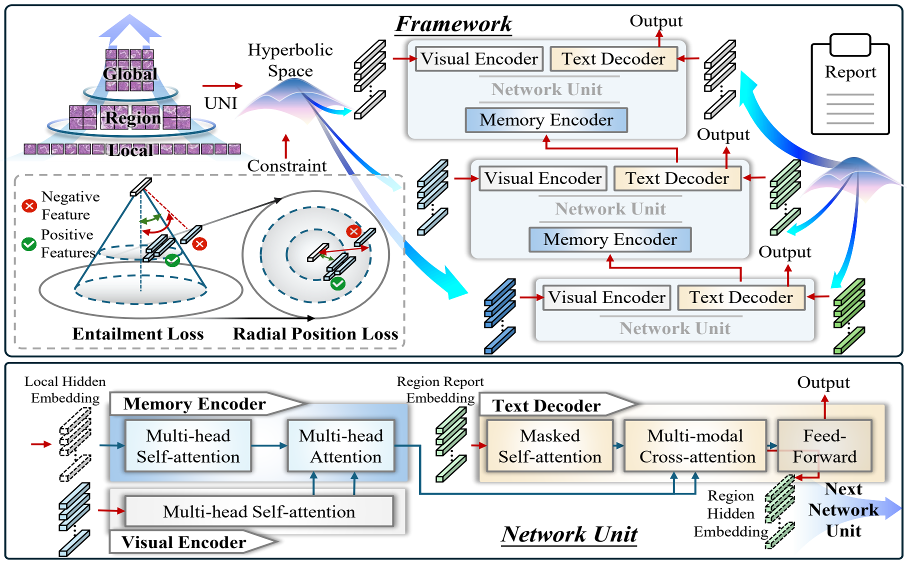

Fig. 2. Overview of the proposed H2M-Net. The upper part illustrates the bottom-up cross-level fusion process. The lower part presents the architecture of a single network unit, which contains a memory encoder.

feature $\mathbf{z}' = \sum_{j=1}^{M^2} \alpha_j \mathbf{z}_j$, where α_j denotes learnable attention weights. The aggregation from region to global level follows the same procedure. For pathological reports, reports are decomposed using a 'low-level priority' regex matching principle designed by our pathologists. Specifically, individual sentences are categorized via hierarchical lexicons: Global (diagnosis, carcinoma, stage, grade, subtype, conclusion, overall, status, malignant, benign), Region (size, diameter, infiltration, invasion, margin, boundary, multifocal, location, gross findings, lymphovascular invasion), and Local (atypia, differentiation, mitosis, necrosis, calcification, stroma, pleomorphism, papilloma, proliferative, in situ). Unmatched sentences default to the Global level, and empty levels are backfilled with the full report. During training, the supervision of the Global level is nested, encompassing both the global report and all sub-level reports. This design ensures multi-level semantic consistency, preventing error propagation even if minor matching deviations occur. After this hierarchical partitioning, features at different levels for both images and reports are extracted and subsequently embedded into the Lorentz manifold using Eq. 1.

Hierarchical Geometric Loss. To explicitly enforce hierarchical structure, inspired by [15,8,19], we exploit the hyperbolic property that more abstract nodes lie closer to the origin. We therefore constrain each parent node to be radially closer to the origin $\mathbf{o}$ than its child. Accordingly, we introduce a radial position loss $\mathcal{L}_{pos}$ and an entailment loss $\mathcal{L}_{ent}$ to enforce hierarchical depth and semantic inclusion, respectively. The radial position loss encourages a semantically more abstract parent node $\mathbf{x}$ to lie closer to the origin $\mathbf{o}$ than its child node $\mathbf{y}$:

$$\mathcal{L}_{pos}(\mathbf{x},\mathbf{y}) = \max\left(0, d_{\mathcal{L}}(\mathbf{x},\mathbf{o}) - d_{\mathcal{L}}(\mathbf{y},\mathbf{o}) + \epsilon\right), \tag{2}$$

where ϵ is a safety margin. The entailment loss enforces hierarchical inclusion via hyperbolic cones defined in the Lorentz model. For a point $\mathbf{u} = (u_0, u_1, \ldots, u_n) \in \mathbb{H}^n_\rho$, let $\mathbf{u}_s = (u_1, \ldots, u_n) \in \mathbb{R}^n$ denote its spatial components in the subspace. The half-aperture angle of the cone centered at $\mathbf{u}$ is defined as $\phi(\mathbf{u}) = \arcsin\left(\frac{r}{\sqrt{\rho}\|\mathbf{u}_s\|_2}\right)$, where $r > 0$ is a fixed radius parameter controlling the cone width. The angular distance between two nodes $\mathbf{x}$ and $\mathbf{y}$ is defined as the Euclidean angle between their spatial projections: $\theta(\mathbf{x},\mathbf{y}) = \arccos\left(\frac{\mathbf{x}_s \cdot \mathbf{y}_s}{\|\mathbf{x}_s\|_2\|\mathbf{y}_s\|_2}\right)$. A child node $\mathbf{y}$ is considered to be entailed by its parent $\mathbf{x}$ if $\theta(\mathbf{x},\mathbf{y}) \leq \phi(\mathbf{x})$. To softly enforce this constraint, we define the entailment loss as:

$$\begin{aligned}\mathcal{L}_{ent} = &\exp(\theta(\mathbf{x},\mathbf{y})/\phi(\mathbf{x}) - 1)\max(0,\, \theta(\mathbf{x},\mathbf{y}) - \alpha\phi(\mathbf{x})) \\ &+ \exp(\phi(\mathbf{x})/\theta(\mathbf{x},\mathbf{y}) - 1)\max(0,\, \phi(\mathbf{x}) - \beta\theta(\mathbf{x},\mathbf{y})),\end{aligned} \tag{3}$$

where α and β control the strength of the inclusion and exclusion constraints.

2.3 Cross-Level Fusion

To effectively propagate information across different hierarchical levels, we construct a bottom-up information flow composed of three cascaded network units. Within each unit, all multi-head attention operations follow the design principles of HypFormer [23], ensuring that multi-head attention computations are performed directly on the Lorentz manifold $\mathbb{H}^n_\rho$, thereby avoiding geometric distortion introduced by projections to Euclidean tangent spaces.

At the l-th level, the Lorentz visual features $\mathbf{V}^{(l)} = \{\mathbf{v}_1, \ldots, \mathbf{v}_N\}$ are fed into the visual encoder. The encoder consists of multi-head hyperbolic self-attention layers. The enhanced output visual representations are denoted as $\hat{\mathbf{V}}^{(l)} = \{\hat{\mathbf{v}}_1, \ldots, \hat{\mathbf{v}}_N\}$. To enable cross-level information interaction, we introduce a memory encoding mechanism. The hidden embedding from the previous level, $\mathbf{H}^{(l-1)}$, serves as the Query, while the enhanced visual features $\hat{\mathbf{V}}^{(l)}$ act as the Keys and Values. Through multi-head attention, cross-level fusion is performed to obtain the memory representation $\mathbf{M}^{(l)}$. The text decoder generates the report sequence $\mathbf{Y}^{(l)}$ at the l-th level in an autoregressive manner. It consists of masked self-attention, multi-modal cross-attention, and feed-forward networks. During the multimodal cross-attention stage, the decoder uses the output of masked self-attention as the Query, while the fused memory $\mathbf{M}^{(l)}$ serves as the Keys and Values. The resulting hyperbolic hidden states are then transformed as follows: (1) They are mapped to the vocabulary probability space via a feed-forward to generate the current textual reports; (2) The hidden state at the final time step is extracted as $\mathbf{H}^{(l)}$ and propagated to the $(l+1)$-th cascaded unit, enabling hierarchical progressive generation.

The overall loss is defined as $\mathcal{L}_{total} = \sum_l \mathcal{L}^{(l)}_{LM} + \mathcal{L}_{ent} + \mathcal{L}_{pos}$, where $\mathcal{L}^{(l)}_{LM}$ denotes the autoregressive language modeling loss at level l.

3 Experiments and Results

3.1 Experimental Setup

Datasets. We follow the data protocol of [3] and conduct experiments on the PathText datasets (BRCA and LUAD) using the same train, validation, and test splits. Specifically, the BRCA dataset (breast cancer) comprises 1,041 whole-slide images, partitioned into 845 for training, 98 for validation, and 98 for testing. The LUAD dataset (lung adenocarcinoma) contains 541 WSIs, with 325, 108, and 108 cases allocated to the training, validation, and test sets, respectively.

Implementation Details. In the experimental setting, we follow the preprocessing protocols of [3][26]. WSIs are processed using CLAM [16]. Specifically, non-overlapping 256×256 patches are extracted at $20\times$ magnification, and patch-level features are obtained using the pretrained UNI model [4]. For optimization, we employ the Adam optimizer with an initial learning rate of 1e-4 and a weight decay of 5e-5. The maximum number of training epochs is set to 60, with early stopping applied using a patience of 20 epochs. All experiments are conducted on a single NVIDIA RTX 5090 GPU.

Evaluation Metrics. We adopt six text generation metrics for quantitative evaluation. BLEU scores are computed based on n-gram precision. BLEU-1 measures word-level matching accuracy, while BLEU-2, BLEU-3, and BLEU-4 (2-gram to 4-gram) further evaluate phrase-level coherence, fluency, and structural consistency with the reference reports. METEOR considers both precision and recall and incorporates stemming and synonym matching, providing a more semantically sensitive evaluation of the generated text. ROUGE-L is based on the Longest Common Subsequence and measures how well the generated report preserves the overall structure and key information of the reference text.

3.2 Experimental Results

In this section, we first compare our method with state-of-the-art approaches, followed by ablation studies on key modules and hyperparameters. Additionally, we provide visualization analyses of feature distributions in the hyperbolic space.

Comparison with SOTA. To validate the proposed method, we perform comprehensive comparisons with recent pathology report generation approaches on the BRCA and LUAD subsets of the PathText dataset. For fairness, all methods use identical data splits and preprocessing, with UNI as the unified feature extractor. As shown in Tab. 1, conventional LSTM- and Transformer-based baselines show limited performance due to restricted modeling capacity. While recent methods (MI-Gen, Hist-Gen, and Bi-Gen) improve generation quality via multi-scale modeling, our approach consistently achieves the best results across all metrics on both subsets. Compared with the second-best method, Bi-Gen, our model improves BLEU-1 by 2.71% and 1.83% on BRCA and LUAD, respectively. ROUGE scores increase by 2.89% and 0.63%, indicating stronger semantic consistency and better structural coherence in the generated reports.

Table 1. Quantitative results of pathology report generation on the PathText (BRCA and LUAD) datasets. The best performance is highlighted in yellow.

Method	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE
PathText-BRCA						
Att-LSTM [22]	0.3773	0.2460	0.1617	0.1090	0.1534	0.2810
Vanilla Transformer [20]	0.3825	0.2518	0.1686	0.1155	0.1581	0.2784
R2Gen [6]	0.3753	0.2465	0.1630	0.1095	0.1525	0.2769
R2GenCMN [5]	0.3806	0.2433	0.1601	0.1071	0.1645	0.2724
MI-Gen [3]	0.4136	0.2754	0.1805	0.1209	0.1663	0.2795
Hist-Gen [11]	0.4223	0.2788	0.1756	0.1165	0.1603	0.2785
Bi-Gen [26]	0.4230	0.2820	0.1891	0.1294	0.1736	0.2773
Ours	0.4501	0.3040	0.2131	0.1460	0.1805	0.3062
PathText-LUAD						
Att-LSTM [22]	0.3707	0.2348	0.1528	0.1006	0.1407	0.2764
Vanilla Transformer [20]	0.3896	0.2516	0.1641	0.1083	0.1451	0.2854
R2Gen [6]	0.3939	0.2522	0.1673	0.1126	0.1475	0.2900
R2GenCMN [5]	0.3855	0.2467	0.1631	0.1120	0.1487	0.2904
MI-Gen [3]	0.3947	0.2538	0.1686	0.1154	0.1518	0.2914
Hist-Gen [11]	0.4009	0.2600	0.1747	0.1219	0.1563	0.2809
Bi-Gen [26]	0.4075	0.2633	0.1766	0.1234	0.1558	0.2929
Ours	0.4258	0.2750	0.1851	0.1289	0.1606	0.2992

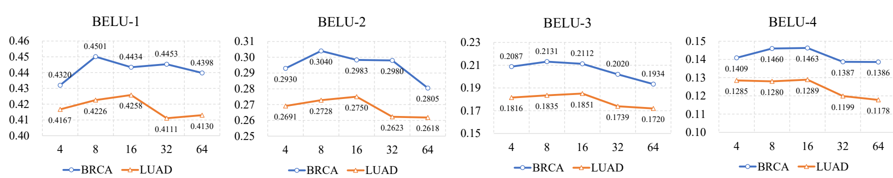

Fig. 3. Ablation on hyperparameter M for hierarchical aggregation.

Ablation Analysis. To evaluate the contribution of each component, we performed systematic ablation studies in Tab. 2. 'B' denotes the baseline without hierarchical structures or hyperbolic embeddings; 'Hier.', 'Lorentz', 'L_{ent}, L_{pos}', and 'Memory' represent hierarchical modeling, hyperbolic embedding, entailment and radial position loss, and the memory encoder, respectively. Introducing hierarchical modeling (b) improves BLEU-4 by 0.46%, confirming the benefit of explicit structural design. Adding hyperbolic embedding (c) further boosts BLEU-1 by 1.44%, demonstrating its effectiveness for hierarchical semantics. Incorporating structural losses (d,e) increases BLEU-1 to 0.4480 and 0.4465, enhancing embedding stability and semantic consistency. Combining hierarchy with Memory alone (f vs. b) yields a 0.64% BLEU-1 gain, but remains limited without hyperbolic geometry and structural constraints. The full model (g) achieves the best performance across all metrics.

Table 2. Ablation study of different model components.

	B	Hier.	Lorentz	L_{ent}	L_{pos}	Memory	BLEU-1	BLEU-2	BLEU-3	BLEU-4	METEOR	ROUGE
(a)	✓						0.4223	0.2788	0.1756	0.1165	0.1603	0.2785
(b)	✓	✓					0.4278	0.2890	0.1823	0.1211	0.1724	0.2799
(c)	✓	✓	✓				0.4422	0.2908	0.1934	0.1314	0.1757	0.2893
(d)	✓	✓	✓	✓			0.4480	0.2935	0.1940	0.1313	0.1775	0.2952
(e)	✓	✓	✓		✓		0.4465	0.2971	0.1902	0.1360	0.1778	0.2849
(f)	✓	✓				✓	0.4342	0.2845	0.1860	0.1294	0.1739	0.2807
(g)	✓	✓	✓	✓	✓	✓	0.4501	0.3040	0.2131	0.1460	0.1805	0.3062

We conduct an ablation study in Fig. 3 to evaluate the impact of the hyperparameter, which defines the $M \times M$ block grid in the hierarchical aggregation module. Experiments are performed on the BRCA and LUAD datasets, and performance is assessed using BLEU-1 to BLEU-4. On the BRCA dataset, the model achieves its best performance at $M = 8$, while performance degrades when $M > 8$. In contrast, the optimal value on the LUAD dataset is $M = 16$, suggesting that a larger receptive field is more suitable for capturing the hierarchical structural characteristics of this dataset. Moreover, for both datasets, $M = 64$ yields the worst performance, confirming that an excessively large receptive field introduces over-smoothing during aggregation.

Hyperbolic Embedding Visualization. Figure 4 visualizes the hyperbolic embeddings using HoroPCA [1], a dimensionality reduction method specifically designed for hyperbolic space. Subfigure (a) presents the feature distribution without loss constraints, where the hierarchical organization appears disordered, while subfigure (b) presents the distribution after applying the proposed geometric losses. As illustrated in (b), a clear hierarchical organization emerges: features are arranged concentrically from the origin outward, with slide-level representations near the center, surrounded by region-level features, and local-level features distributed closer to the boundary.

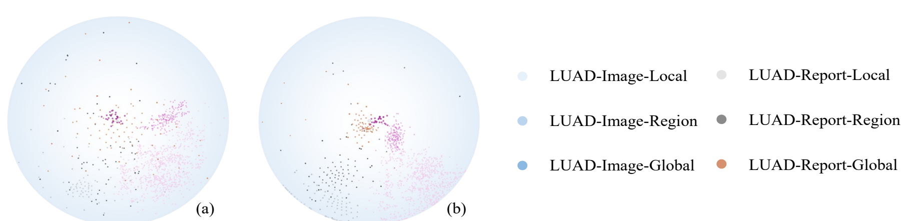

Fig. 4. The visualization of hyperbolic embeddings from different hierarchical levels.

4 Conclusion

In this paper, we proposed H^{2}M-Net, a hierarchical hyperbolic memory network for pathology report generation that addresses the geometric mismatch of Euclidean representations. By embedding multi-scale features on the Lorentz manifold and introducing a hierarchical memory mechanism with explicit geometric constraints, our model effectively preserves local-to-global diagnostic logic and enhances cross-level reasoning. Extensive experiments demonstrate that H^{2}M-Net achieves superior performance and generates diagnostically accurate reports.

Acknowledgments. This work is supported by the National Natural Science Foundation of China (62272083, 62495085, 62472066), the Key Research and Development Program of Liaoning Province (2023JH26/10200014), the Fundamental Research Funds for the Central Universities of China (LD2023019), the Liaoning Provincial Natural Science Foundation of China (2024-MS-271, 2025-MS-336), and the Liaoning Provincial Medical Education Research Project of China (2024-N002-04).

Disclosure of Interests. The authors have no competing interests to declare that are relevant to the content of this article.

References

1. Chami, I., Gu, A., Nguyen, D.P., Ré, C.: Horopca: Hyperbolic dimensionality reduction via horospherical projections. In: International Conference on Machine Learning. pp. 1419–1429. PMLR (2021)
2. Chami, I., Ying, Z., Ré, C., Leskovec, J.: Hyperbolic graph convolutional neural networks. Advances in neural information processing systems **32** (2019)
3. Chen, P., Li, H., Zhu, C., Zheng, S., Shui, Z., Yang, L.: Wsicaption: Multiple instance generation of pathology reports for gigapixel whole-slide images. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 546–556. Springer (2024)
4. Chen, R.J., Ding, T., Lu, M.Y., Williamson, D.F., Jaume, G., Song, A.H., Chen, B., Zhang, A., Shao, D., Shaban, M., et al.: Towards a general-purpose foundation model for computational pathology. Nature medicine **30**(3), 850–862 (2024)
5. Chen, Z., Shen, Y., Song, Y., Wan, X.: Cross-modal memory networks for radiology report generation. In: Proceedings of the 59th annual meeting of the association for computational linguistics and the 11th international joint conference on natural language processing (volume 1: long papers). pp. 5904–5914 (2021)
6. Chen, Z., Song, Y., Chang, T.H., Wan, X.: Generating radiology reports via memory-driven transformer. arXiv preprint arXiv:2010.16056 (2020)
7. Desai, K., Nickel, M., Rajpurohit, T., Johnson, J., Vedantam, S.R.: Hyperbolic image-text representations. In: International Conference on Machine Learning. pp. 7694–7731. PMLR (2023)
8. Duan, Y., Nie, F., Chen, H., Hu, Z., Wang, R., Li, X.: Hyperbolic hierarchical representation learning for generalized category discovery. IEEE Transactions on Neural Networks and Learning Systems (2025)

9. Gonzalez-Jimenez, A., Lionetti, S., Amruthalingam, L., Gottfrois, P., Gröger, F., Pouly, M., Navarini, A.A.: Is hyperbolic space all you need for medical anomaly detection? In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 312–322. Springer (2025)
10. Guo, Y., Liu, Y., Wang, W., Guo, Y., Jia, Q.: Hyphoi: Exploring hierarchical hyperbolic embeddings for human-object interaction detection. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 6517–6527 (2026)
11. Guo, Z., Ma, J., Xu, Y., Wang, Y., Wang, L., Chen, H.: Histgen: Histopathology report generation via local-global feature encoding and cross-modal context interaction. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 189–199. Springer (2024)
12. Huang, P., Huang, Y., Zhao, W., He, J., Yu, L.: Hyperpath: Knowledge-guided hyperbolic semantic hierarchy modeling for wsi analysis. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 262–272. Springer (2025)
13. Liu, K., Ma, Z., Kang, X., Li, Y., Xie, K., Jiao, Z., Miao, Q.: Enhanced contrastive learning with multi-view longitudinal data for chest x-ray report generation. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 10348–10359 (2025)
14. Liu, Y., Feng, L., Jia, Q., Liu, Z., Cao, Z.H.: Two teachers are better than one: Semi-supervised elliptical object detection by dual-teacher collaborative guidance. In: Proceedings of the 32nd ACM International Conference on Multimedia. pp. 6355–6363 (2024)
15. Liu, Y., He, Z., Han, K.: Hyperbolic category discovery. In: Proceedings of the Computer Vision and Pattern Recognition Conference. pp. 9891–9900 (2025)
16. Lu, M.Y., Williamson, D.F., Chen, T.Y., Chen, R.J., Barbieri, M., Mahmood, F.: Data-efficient and weakly supervised computational pathology on whole-slide images. Nature biomedical engineering **5**(6), 555–570 (2021)
17. Pal, A., van Spengler, M., di Melendugno, G.M.D., Flaborea, A., Galasso, F., Mettes, P.: Compositional entailment learning for hyperbolic vision-language models. arXiv preprint arXiv:2410.06912 (2024)
18. Qiao, Z., Han, L., Zhen, X., Gao, J., Qian, Z.: Hyden: Hyperbolic density representations for medical images and reports. In: Proceedings of the 31st International Conference on Computational Linguistics. pp. 6285–6297 (2025)
19. Ramasinghe, S., Shevchenko, V., Avraham, G., Thalaiyasingam, A.: Accept the modality gap: An exploration in the hyperbolic space. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 27263–27272 (2024)
20. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)
21. Wang, J., Sun, Y., Mei, Y., Xu, Z., Fan, X.: Personalizing federated learning guided by site-aggregated representation for multi-site one-shot medical image segmentation. In: 2025 IEEE International Conference on Bioinformatics and Biomedicine (BIBM). pp. 7113–7120. IEEE (2025)
22. Xu, K., Ba, J., Kiros, R., Cho, K., Courville, A., Salakhudinov, R., Zemel, R., Bengio, Y.: Show, attend and tell: Neural image caption generation with visual attention. In: International conference on machine learning. pp. 2048–2057. PMLR (2015)

23. Yang, M., Verma, H., Zhang, D.C., Liu, J., King, I., Ying, R.: Hypformer: Exploring efficient transformer fully in hyperbolic space. In: Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining. pp. 3770–3781 (2024)
24. Yao, S., Jia, Q., Liu, Y., Zhang, P., Sun, L., Wang, W., Zhu, Y., Zhang, B., Fan, X.: Dual-level hypergraph generation for addressing feature scarcity in whole-slide image classification. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 28328–28337 (2026)
25. Yun, H., Maeng, J., Kang, E., Suk, H.I.: Diff-rrg: Longitudinal disease-wise patch difference as guidance for llm-based radiology report generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 152–161. Springer (2025)
26. Zhang, L., Yun, B., Li, Q., Wang, Y.: Historical report guided bi-modal concurrent learning for pathology report generation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 343–352. Springer (2025)
27. Zhang, Y., Qiu, B., Jia, Q., Liu, Y., He, R.: Not just object, but state: compositional incremental learning without forgetting. Advances in Neural Information Processing Systems **37**, 123182–123206 (2024)
28. Zhang, Z., Chen, P., McGough, M., Xing, F., Wang, C., Bui, M., Xie, Y., Sapkota, M., Cui, L., Dhillon, J., et al.: Pathologist-level interpretable whole-slide cancer diagnosis with deep learning. Nature Machine Intelligence **1**(5), 236–245 (2019)