Hierarchical Hyperbolic Self-Attention Network for Medical Image Segmentation

Shaocheng Jin[1], Rui Wang[1], Chen Hu[1], Tianyang Xu[1], Xiao-Jun Wu[1], Tao Zhou[1,2](✉)

[1] School of Artificial Intelligence and Computer Science, Jiangnan University, Wuxi 214122, Jiangsu, China.
[2] The Institute of Translational Medicine, Wuxi School of Medicine, Jiangnan University, Wuxi 214122, Jiangsu, China.
taozhou.ai@jiangnan.edu.cn

Abstract. Medical image segmentation is critical for disease monitoring and treatment planning, with U-Net serving as a fundamental architecture in this domain. Building on its effectiveness, recent advancements have integrated mechanisms such as Transformers and Mamba into U-shaped structures, achieving promising results. However, these methods typically operate in Euclidean space, limiting their ability to capture complex geometric structures. Additionally, they often apply uniform attention strategies across scales, failing to appropriately model fine-grained spatial details in shallow layers and high-level semantic dependencies in deeper layers. To address these limitations, we propose a novel framework, Hierarchical Hyperbolic Self-Attention Network (H^{2}SA-Net), which integrates geometric representation learning with a dual-granularity attention mechanism. Our key contribution is a depth-adaptive stratified Hyperbolic Self-Attention (HSA) module. In shallow layers, pixel-wise HSA refines local spatial context, while in deeper layers, channel-wise HSA enhances global semantic correlations. To maintain geometric consistency, we employ the Einstein Midpoint for feature aggregation. Enhanced by a wavelet-based backbone and a Hyperbolic Multinomial Logistic Regression (MLR) head, the proposed framework effectively balances local texture preservation with global semantic understanding. Extensive experimental results on diverse datasets demonstrate the effectiveness of our method. The code is available at https://github.com/jjscc/H2SA-Net.

Keywords: Medical image segmentation, Hyperbolic geometry, Self-attention

1 Introduction

Medical image segmentation is a fundamental research direction in medical image analysis [10,11,30,23], providing crucial information for disease monitoring and treatment planning. Numerous deep learning methods have been developed for this task [11,22,26,35,36]. Notably, U-Net [26] is a widely adopted CNN architecture with a symmetric encoder-decoder design. However, CNNs struggle to capture global context, leading to inaccuracies in segmenting small targets. While U-Net variants have improved feature fusion [19,36], global dependency modeling remains difficult. The advent of Transformers [6,29] has addressed this by enabling long-range dependency modeling through self-attention, leading to high-performing hybrids like TransUNet [8] and Swin-UNet [5]. More recently, State-Space Models (SSMs) [17,18] have emerged as efficient alternatives, exemplified by U-Mamba [22]. Additionally, wavelet transform [12,13] has also shown promise for medical image segmentation [33] by providing explicit multi-resolution decomposition that helps preserve high-frequency details. A notable example is UWT-Net [35], which integrates wavelet convolution to exploit multi-frequency representations and achieve accurate segmentation performance.

However, most existing segmentation architectures perform feature modeling in Euclidean space, which may be suboptimal for organizing complex medical image representations with multi-scale and structured patterns. Hyperbolic geometry provides a promising alternative for modeling hierarchical or structured relationships with low distortion [7,24]. For medical image segmentation, this geometry can serve as an inductive bias to organize dense features across network depths, from shallow boundary and texture cues to deeper region-level semantic dependencies. While hyperbolic neural networks have shown promise in tasks like semantic segmentation [1] and domain adaptation [14], their application to fully supervised medical image segmentation remains largely underexplored. Moreover, existing hyperbolic methods typically adopt a uniform interaction strategy across network layers, overlooking the depth-dependent nature of visual representations. In medical segmentation networks, shallow layers primarily encode high-frequency spatial details, while deep layers encode abstract semantic information [11,32]. Applying the same attention mechanism uniformly across layers thus leads to suboptimal feature refinement.

To this end, we propose the Hierarchical Hyperbolic Self-Attention Network (H^2SA-Net), a geometry-aware framework that integrates hyperbolic representation learning with a depth-adaptive attention mechanism. Our approach begins with a wavelet transform-based encoder [35] that decomposes the input image into multi-scale frequency components, thereby preserving high-frequency boundary details often lost during standard downsampling. Building on this representation, we propose a Hyperbolic Self-Attention (HSA) module tailored to different network depths. Specifically, pixel-wise HSA is applied to shallow features for spatial context modeling, while channel-wise HSA operates on deep features for semantic recalibration. Finally, the segmentation head incorporates a Hyperbolic Multinomial Logistic Regression (MLR) layer [1], which performs per-pixel classification via class-specific hyperplanes in hyperbolic space, ensuring geometric consistency throughout the prediction process.

Our contributions are highlighted as follows: i) We propose H^2SA-Net, a novel medical image segmentation framework that integrates wavelet-based encoding with hyperbolic geometry to better capture hierarchical anatomical structures. **ii)** We present a stratified hyperbolic self-attention module, which applies pixel-wise attention to shallow features and channel-wise attention to

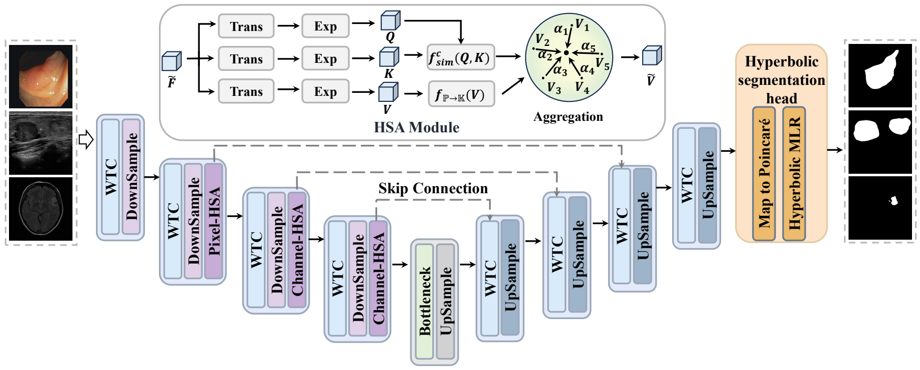

Fig. 1. Architecture of H^{2}SA-Net (WTC denotes the wavelet transform convolution).

deep features. **iii)** We employ a Hyperbolic MLR layer to enable geometry-consistent per-pixel classification directly in hyperbolic space. **iv)** Experimental results on three medical image segmentation tasks (polyp, ultrasound thyroid, and brain tumor) validate the effectiveness of the proposed H^{2}SA-Net.

2 Proposed Method

2.1 Overview of H^{2}SA-Net

Fig. 1 illustrates the overall architecture of H^{2}SA-Net. To effectively capture high-frequency texture details and preserve edge information, we utilize Wavelet Transform Convolution (WTC) as the encoder backbone [35]. A key innovation of our framework is the stratified Hyperbolic Self-Attention (HSA) module, which projects features into the Poincaré ball to model hierarchical dependencies. Specifically, we deploy pixel-wise HSA at intermediate encoder stages (Layers 2 and 3) to rigorously capture spatial contexts, while transitioning to channel-wise HSA at the deepest stage (Layer 4) to explicitly recalibrate semantic feature channels. Bridging the encoder and decoder is a standard convolutional bottleneck, which refines deep feature representations without altering feature dimension or resolution. Subsequently, WTC is employed for signal reconstruction in the decoder. Finally, instead of conventional Euclidean segmentation heads, we design a Hyperbolic segmentation head based on Hyperbolic Multinomial Logistic Regression (MLR) to perform geometry-consistent pixel-wise classification.

2.2 Hyperbolic Self-Attention

We propose a novel HSA module to extract the correlation of pixels in the low-level feature map and channels in the high-level feature map. We first briefly review the hyperbolic geometry. Hyperbolic space is a Riemannian manifold with constant negative curvature, which can be described by several isometric models [4]. In this work, we adopt the Poincaré ball model $(\mathbb{P}_c^n, g_c^{\mathbb{P}})$, where $\mathbb{P}_c^n = \{x \in \mathbb{R}^n : c\|x\|^2 < 1\}$ denotes the open ball of radius $\frac{1}{\sqrt{c}}$ and $g_c^{\mathbb{P}} = (\lambda_x^c)^2 I_n$ is the associated Riemannian metric with conformal factor $\lambda_x^c = \frac{2}{1-c\|x\|^2}$, and I_n represents the $n \times n$ identify matrix. The tangent space of the Poincaré ball can be viewed as an approximation of Euclidean space. Given a tangent vector $v \in T_x\mathbb{P}_c^n$, the Riemannian exponential map projects v onto the manifold by

$$\exp_x^c(v) = x \oplus_c \left(\tanh\left(\sqrt{c}\frac{\lambda_x^c \|v\|}{2}\right) \frac{v}{\sqrt{c}\|v\|} \right), \tag{1}$$

where $\oplus_c$ denotes the Möbius addition, and it is defined as follows:

$$x \oplus_c y = \frac{(1 + 2c\langle x, y\rangle + c\|y\|^2)x + (1 - c\|x\|^2)y}{1 + 2c\langle x, y\rangle + c^2\|x\|^2\|y\|^2}. \tag{2}$$

Conversely, the Riemannian logarithmic map projects points on the manifold back to the tangent space, which is defined as follows:

$$\log_x^c(y) = \frac{2}{\sqrt{c}\lambda_x^c} \tanh^{-1}(\sqrt{c}\| - x \oplus_c y\|) \frac{-x \oplus_c y}{\| - x \oplus_c y\|}. \tag{3}$$

The geodesic distance between two points $x, y \in \mathbb{P}_c^n$ can be deduced as follows:

$$d_{\mathbb{P}}^c(x, y) = \frac{2}{\sqrt{c}} \tanh^{-1}(\sqrt{c}\| - x \oplus_c y\|). \tag{4}$$

Since the query, key, and value are defined on the Poincaré ball, we first apply linear transformation in the tangent space at the origin, and then map them onto $\mathbb{P}_c^n$ via the exponential map. For the input feature map $\tilde{F}$, we obtain

$$Q = \exp_0^c(W_q\tilde{F}), K = \exp_0^c(W_k\tilde{F}), V = \exp_0^c(W_v\tilde{F}), \tag{5}$$

where W_q, W_k, W_v are learnable parameters. In Euclidean space, the inner product is commonly used to measure the similarity between queries and keys. Here, we adopt a distance-based similarity defined on hyperbolic space [20] and extend it to the Poincaré ball:

$$\alpha = f_{sim}^c(Q, K) = f(-\tau d_{\mathbb{P}}^c(Q, K) - \mathcal{G}), \tag{6}$$

where $d_{\mathbb{P}}^c(\cdot, \cdot)$ is the Poincaré distance, τ is an inverse temperature and $\mathcal{G}$ is a bias parameter. We use the Softmax function as the activation $f(\cdot)$. Considering numerical stability and computational complexity, we map the points from Poincaré ball to the Klein model and use the Einstein midpoint to compute aggregation [28]. Since the Poincaré ball model and the Klein model are isometric, we can achieve this process by

$$\mathbb{P}_c^n \to \mathbb{K}_c^n : \tilde{x} = f_{\mathbb{P}\to\mathbb{K}}(x) \frac{x}{1 + \sqrt{1 - c\|x\|^2}}. \tag{7}$$

The aggregated output can be achieved by the Einstein midpoint, which is calculated as follows:

$$\tilde{V}_i = f_{agg}(\{\alpha_{ij}\}_j, \{V_{ij}\}_j) = \sum_j \left[\frac{\alpha_{ij}\gamma(V_{ij})}{\sum_l \alpha_{il}\gamma(V_{il})}\right] V_{ij}, \tag{8}$$

where $\gamma(V_{ij}) = \frac{1}{\sqrt{1-\|V_{ij}\|^2}}$ is the Lorentz factor.

2.3 Hyperbolic Segmentation Head

Unlike traditional Euclidean segmentation heads, we leverage a hyperbolic MLR-based segmentation head to perform per-pixel classification for image segmentation [1,15]. Given the output feature map $\hat{F}$ from the decoder, we first apply the exponential map to project each pixel-wise feature onto the Poincaré ball, obtaining a hyperbolic embedding z. In this hyperbolic space, each class-specific decision boundary can be represented by a gyroplane [15], defined as follows:

$$H^c = \{z_{ij} \in \mathbb{P}_c^n, \langle -p \oplus_c z_{ij}, w\rangle = 0\}, \tag{9}$$

where $p \in \mathbb{P}_c^n$ is the offset, and $w \in T_p\mathbb{P}_c^n$ is the orientation of the gyroplane. The hyperbolic distance of z_{ij} to the gyroplane of class y is given as follows:

$$d_c(z_{ij}, H_y^c) = \frac{1}{\sqrt{c}} \sinh^{-1}\left(\frac{2\sqrt{c}\langle -p_y \oplus_c z_{ij}, w_y\rangle}{(1 - c\| - p_y \oplus_c z_{ij}\|^2)\|w_y\|}\right). \tag{10}$$

The logit of class y for pixel output z_{ij} is deduced as follows:

$$p(\hat{y} = y|z_{ij}) \propto \exp\left(\frac{\lambda_{p_y}^c\|w_y\|}{\sqrt{c}} \sinh^{-1}\left(\frac{2\sqrt{c}\langle -p_y \oplus_c z_{ij}, w_y\rangle}{(1 - c\| - p_y \oplus_c z_{ij}\|^2)\|w_y\|}\right)\right), \tag{11}$$

which can be optimized with the cross-entropy loss and gradient descent. Following [1], we employ an efficient formulation by rewriting the Möbius addition as a linear combination, formulated as follows:

$$p_y \oplus_c z_{ij} = \beta_1 p_y + \beta_2 z_{ij}, \tag{12}$$

where $\beta_1 = \frac{1+2c\langle p_y, z_{ij}\rangle + c\|z_{ij}\|^2}{1+2c\langle p_y, z_{ij}\rangle + c^2\|p_y\|^2\|z_{ij}\|^2}$, and $\beta_2 = \frac{1-c\|p_y\|^2}{1+2c\langle p_y, z_{ij}\rangle + c^2\|p_y\|^2\|z_{ij}\|^2}$.

Therefore, the inner product and squared norm of the Möbius addition can be computed as follows:

$$\langle -p_y \oplus_c z_{ij}, w_y\rangle = \beta_1\langle -p_y, w\rangle + \beta_2\langle z_{ij}, w\rangle, \tag{13}$$

$$\| - p_y \oplus_c z_{ij}\|^2 = \beta_1^2\| - p_y\|^2 + 2\beta_1\beta_2\langle -p_y, z_{ij}\rangle + \beta_2^2\|z_{ij}\|^2. \tag{14}$$

2.4 Loss Function

We combine the cross entropy loss $\mathcal{L}_{BCE}$ and the Intersection over Union (IoU) loss $\mathcal{L}_{IoU}$ [31] as the final loss function, which is defined as:

$$\mathcal{L}_{Seg} = \mathcal{L}_{BCE}(P, G) + \mathcal{L}_{IoU}(P, G), \tag{15}$$

where P and G denote the semantic prediction maps and the ground truths, respectively.

Table 1. Performance comparison on the different segmentation tasks.

Model	Param (M)	Kvasir [21]		ClinicDB [2]		ETIS [27]		Brain [3]		TN3K [16]	
		Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
U-Net [26] (MICCA'15)	31.04	0.821	0.756	0.824	0.767	0.406	0.343	0.812	0.735	0.820	0.728
Att-UNet [25] (MIDL'18)	35.75	0.769	0.683	0.866	0.809	0.382	0.308	0.817	0.734	0.825	0.738
U-Net++ [36] (MICCA'18)	36.63	0.824	0.753	0.797	0.741	0.413	0.342	0.815	0.737	0.828	0.740
DeepLabv3 [9] (ECCCV'18)	39.76	0.857	0.817	0.873	0.821	0.575	0.487	0.814	0.733	0.827	0.735
Swin-UNet [5] (ECCV'22)	27.17	0.869	0.816	0.847	0.798	0.573	0.516	0.816	0.728	0.828	0.740
U-Mamba [22] (Arxiv'24)	11.26	0.836	0.767	0.865	0.810	0.480	0.407	0.837	0.749	0.811	0.731
KMUNet [34] (MICCA'25)	7.01	0.850	0.748	0.846	0.733	0.576	0.475	0.836	0.720	0.830	0.711
UWT-Net [35] (MICCA'25)	29.34	0.869	0.807	0.856	0.805	0.561	0.495	0.823	0.733	0.823	0.738
H^{2}SA-Net (Ours)	34.84	**0.885**	**0.823**	**0.884**	**0.840**	**0.580**	**0.513**	**0.844**	**0.759**	**0.832**	**0.750**

3 Experiment

3.1 Experimental Setup

Datasets. We conduct experiments on three medical image segmentation tasks: polyp, thyroid, and brain tumor segmentation. • Colonoscopy datasets: We evaluate our method on three publicly available datasets: Kvasir [21], CVC-ClinicDB [2], and ETIS [27]. Following the setting in [11], we use 900 images from Kvasir and 550 images from CVC-ClinicDB for training. The remaining images from Kvasir and CVC-ClinicDB, and the entirety of ETIS are taken as the test set. • The TN3K [16] dataset is a challenging dataset for the thyroid nodule segmentation task, with $2,879$ training and 614 test images. • Brain tumor dataset [3] includes brain MRI scans with FLAIR abnormality segmentation masks from 110 patients with lower-grade glioma in The Cancer Genome Atlas. We adopt a patient-level split: all slices from the same patient are assigned exclusively to either the training or test set, avoiding leakage in slice-based evaluation. We use $1,090$ images for training and the remaining 283 images for testing.

Implementation Details. All experiments are conducted using PyTorch 2.5.1 with CUDA 12.4 on an NVIDIA RTX 5880 GPU. Our method is trained with the Adam optimizer using a learning rate of $1e^{-3}$ and weight decay of $1e^{-4}$. We only apply standard data augmentations, including random horizontal flipping and random rotation. The batch size is set to 10, and all models can be converged within 150 epochs. In our experiments, all input images are uniformly resized to the dimension of 352×352. For a fair comparison, we do not use any pre-trained weights or post-processing strategies within all comparison methods.

3.2 Performance Comparison

We compare H^{2}SA-Net with several representative segmentation methods, including U-Net [26], Att-UNet [25], U-Net++ [36], DeepLabv3 [9], Swin-UNet [5], U-Mamba [22], KMUNet [34] and HWT-Net [35]. Table 1 reports the quantitative results comparison with state-of-the-art methods on five datasets. It can be

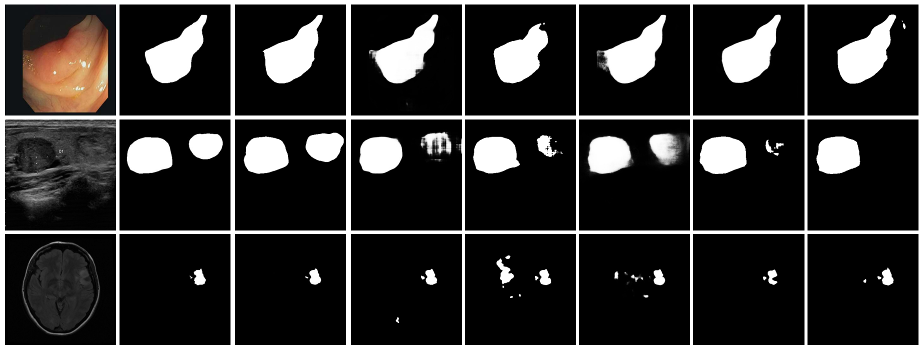

Fig. 2. Qualitative visualization of our model and other comparison methods.

Table 2. Ablation study on the key components. "HSA-C" and "HSA-P" denote channel-wise HSA and pixel-wise HSA. "ESA" denotes Euclidean self-attention, "E-Cls" denotes Euclidean segmentation head, and "Euc" denotes the Euclidean counterpart, where all hyperbolic components are replaced with their Euclidean counterpart.

Settings	**Kvasir** [21]		**ClinicDB** [2]		**ETIS** [27]		**TN3K** [16]	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
H^{2}SA-Net (HSA-C)	0.870	0.810	0.862	0.810	0.572	0.502	0.827	0.743
H^{2}SA-Net (HSA-P)	0.872	0.811	0.867	0.820	0.571	0.510	0.830	0.747
H^{2}SA-Net (ESA-(C+P))	0.853	0.786	0.855	0.807	0.556	0.490	0.814	0.726
H^{2}SA-Net (E-Cls)	0.877	0.816	0.867	0.821	0.552	0.492	0.828	0.743
H^{2}SA-Net (Euc)	0.864	0.804	0.860	0.808	0.556	0.494	0.820	0.736
H^{2}SA-Net	**0.885**	**0.823**	**0.884**	**0.840**	**0.580**	**0.513**	**0.832**	**0.750**

seen that our model outperforms other methods across different segmentation tasks. Compared with UWT-Net, which integrates wavelet convolution into a U-Net backbone, our approach achieves steady performance improvements without substantially increasing the number of model parameters. Qualitative results for various medical image segmentation tasks are presented in Fig. 2. As observed, our method yields segmentation outputs that closely align with the ground truth, preserving fine-grained structures and capturing subtle details, particularly in challenging regions and along object boundaries.

3.3 Ablation Studies

Effectiveness of Key Components. We conduct experiments to access the effectiveness of the main components. As shown in Table 2, applying pixel-wise HSA to shallow features and channel-wise HSA to deep features consistently improves segmentation performance. Moreover, HSA shows a significant improvement over Euclidean self-attention, further validating its effectiveness. In

Table 3. Effect of the curvature parameter c.

Settings	Kvasir [21]		ClinicDB [2]		ETIS [27]		TN3K [16]	
	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
H^2SA-Net ($c = 0.5$)	0.876	0.819	0.861	0.814	0.565	0.498	0.829	0.746
H^2SA-Net ($c = 1.0$)	0.879	0.817	0.880	0.832	0.573	0.502	0.827	0.743
H^2SA-Net ($c = 1.5$)	0.872	0.810	0.865	0.820	0.570	0.498	0.822	0.742
H^2SA-Net (c learnable)	**0.885**	**0.823**	**0.884**	**0.840**	**0.580**	**0.513**	**0.832**	**0.750**

Table 4. Effectiveness of our HSA applied to other architectures.

Model	Efficiency		Kvasir [21]		ClinicDB [2]		ETIS [27]		TN3K [16]	
	Param(M)	FLOPs(G)	Dice	IoU	Dice	IoU	Dice	IoU	Dice	IoU
U-Mamba [22]	11.26	233.08	0.836	0.767	0.865	0.810	0.480	0.407	0.811	0.731
U-Mamba-H²SA	12.64	237.74	**0.850**	**0.785**	**0.875**	**0.826**	**0.495**	**0.428**	**0.821**	**0.737**
TransUNet [8]	105.50	60.98	0.913	0.857	0.895	0.847	0.728	0.658	0.841	0.754
TransUNet-H²SA	106.83	65.58	**0.916**	**0.867**	**0.910**	**0.866**	**0.735**	**0.663**	**0.856**	**0.774**

addition, replacing the MLR-based hyperbolic segmentation head with a traditional Euclidean segmentation head leads to a performance drop. Furthermore, replacing all hyperbolic components with their Euclidean counterparts consistently underperforms the full model. This indicates that embedding features in hyperbolic space can better separate pixel representations, dovetailing with previous work [1]. In terms of efficiency, the Euclidean baseline requires 61.34 GFLOPs and 46.846 ms inference latency. The Euclidean variant of our design requires 67.44 GFLOPs and 47.747 ms, while the full hyperbolic model requires 67.94 GFLOPs and 51.219 ms. Compared with the Euclidean variant, hyperbolic modeling introduces only 0.50 GFLOPs and 3.472 ms additional latency, while yielding consistent performance gains.

Effect of Curvature Parameter c. Compared with Euclidean space, hyperbolic space introduces an additional curvature parameter c that controls the curvature and radius of the Poincaré ball. In our model, c is a learnable parameter initialized to 1 and optimized jointly with the network. We conduct a series of experiments to further investigate the effect of different curvature settings. The corresponding results are presented in Table 3. The results indicate that allowing the model to autonomously adjust the manifold curvature can better match the inherent geometry of the data, thereby enhancing geometric consistency.

Generality of HSA. To evaluate the generality of our model, we integrate HSA into two representative network architectures: TransUNet [8] and U-Mamba [22]. Table 4 reports the results on the polyp segmentation task. It can be observed that incorporating our HSA consistently improves the performance of both Transformer and Mamba-based architectures, further validating its effectiveness and plug-and-play compatibility.

4 Conclusion

In this paper, we propose H^2SA-Net, a geometry-aware medical image segmentation framework integrating wavelet encoding with hyperbolic representation learning. We design a novel hyperbolic self-attention mechanism in the Poincaré ball and employ a stratified strategy that applies pixel-wise attention at intermediate layers and channel-wise attention at deeper layers. Experiments on three segmentation tasks demonstrate the effectiveness of H^2SA-Net, and ablation studies verify the contribution of each component. This study highlights the potential for further investigation into feature representation learning in hyperbolic space for medical imaging tasks. Moreover, the proposed HSA can be extended from 2D to volumetric segmentation by replacing pixel-wise HSA with voxel- or window-wise HSA while retaining channel-wise hyperbolic recalibration.

Acknowledgement . This work was in part supported by the National Natural Science Foundation of China (No. 62576153) and Postgraduate Research & Practice Innovation Program of Jiangsu Province (22671065).

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

References

1. Atigh, M.G., Schoep, J., Acar, E., Van Noord, N., Mettes, P.: Hyperbolic image segmentation. In: Proc. CVPR (2022)
2. Bernal, J., Sánchez, F.J., Fernández-Esparrach, G., Gil, D., Rodríguez, C., Vilariño, F.: Wm-dova maps for accurate polyp highlighting in colonoscopy: Validation vs. saliency maps from physicians. Computerized Med. Imag. Graph. pp. 99–111 (2015)
3. Buda, M., Saha, A., Mazurowski, M.A.: Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. Comput. Biol. Med. pp. 218–225 (2019)
4. Cannon, J.W., Floyd, W.J., Kenyon, R., Parry, W.R., et al.: Hyperbolic geometry. Flavors of geometry p. 2 (1997)
5. Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., Wang, M.: Swin-unet: Unet-like pure transformer for medical image segmentation. In: Proc. ECCV (2022)
6. Carion, N., Massa, F., Synnaeve, G., Usunier, N., Kirillov, A., Zagoruyko, S.: End-to-end object detection with transformers. In: Proc. ECCV (2020)
7. Chami, I., Gu, A., Chatziafratis, V., Ré, C.: From trees to continuous embeddings and back: Hyperbolic hierarchical clustering. In: Proc. NeurIPS (2020)
8. Chen, J., Lu, Y., Yu, Q., Luo, X., Adeli, E., Wang, Y., Lu, L., Yuille, A.L., Zhou, Y.: TransUNet: Transformers make strong encoders for medical image segmentation. arXiv preprint arXiv:2102.04306 (2021)
9. Chen, L.C., Zhu, Y., Papandreou, G., Schroff, F., Adam, H.: Encoder-decoder with atrous separable convolution for semantic image segmentation. In: Proc. ECCV (2018)

10. Dai, D., Dong, C., Xu, S., Yan, Q., Li, Z., Zhang, C., Luo, N.: Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation. Med. Image Anal. p. 102293 (2022)
11. Fan, D.P., Ji, G.P., Zhou, T., Chen, G., Fu, H., Shen, J., Shao, L.: Pranet: Parallel reverse attention network for polyp segmentation. In: Proc. MICCA (2020)
12. Finder, S.E., Amoyal, R., Treister, E., Freifeld, O.: Wavelet convolutions for large receptive fields. In: Proc. ECCV (2024)
13. Finder, S.E., Zohav, Y., Ashkenazi, M., Treister, E.: Wavelet feature maps compression for image-to-image cnns. In: Proc. NeurIPS (2022)
14. Franco, L., Mandica, P., Kallidromitis, K., Guillory, D., Li, Y.T., Darrell, T., Galasso, F.: Hyperbolic active learning for semantic segmentation under domain shift. In: Proc. ICML (2024)
15. Ganea, O., Bécigneul, G., Hofmann, T.: Hyperbolic neural networks. Proc. NeurIPS (2018)
16. Gong, H., Chen, J., Chen, G., Li, H., Li, G., Chen, F.: Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. Comput. Biol. Med. p. 106389 (2023)
17. Gu, A.: Modeling sequences with structured state spaces. Stanford University (2023)
18. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)
19. Gu, Z., Cheng, J., Fu, H., Zhou, K., Hao, H., Zhao, Y., Zhang, T., Gao, S., Liu, J.: CE-Net: Context encoder network for 2D medical image segmentation. IEEE Trans. Med. Imaging. pp. 2281–2292 (2019)
20. Gulcehre, C., Denil, M., Malinowski, M., Razavi, A., Pascanu, R., Hermann, K.M., Battaglia, P., Bapst, V., Raposo, D., Santoro, A., et al.: Hyperbolic attention networks. arXiv preprint arXiv:1805.09786 (2018)
21. Jha, D., Smedsrud, P.H., Riegler, M.A., Halvorsen, P., De Lange, T., Johansen, D., Johansen, H.D.: Kvasir-seg: A segmented polyp dataset. In: Proc. MMM (2019)
22. Ma, J., Li, F., Wang, B.: U-mamba: Enhancing long-range dependency for biomedical image segmentation. arXiv preprint arXiv:2401.04722 (2024)
23. Mei, J., Zhou, T., Huang, K., Zhang, Y., Zhou, Y., Wu, Y., Fu, H.: A survey on deep learning for polyp segmentation: Techniques, challenges and future trends. Visual Intelligence p. 1 (2025)
24. Nickel, M., Kiela, D.: Poincaré embeddings for learning hierarchical representations. In: Proc. NeurIPS (2017)
25. Oktay, O., Schlemper, J., Folgoc, L.L., Lee, M., Heinrich, M., Misawa, K., Mori, K., McDonagh, S., Hammerla, N.Y., Kainz, B., et al.: Attention u-net: Learning where to look for the pancreas. arXiv preprint arXiv:1804.03999 (2018)
26. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: Proc. MICCA (2015)
27. Silva, J., Histace, A., Romain, O., Dray, X., Granado, B.: Toward embedded detection of polyps in wce images for early diagnosis of colorectal cancer. Int. J. Comput. Assist. Radiol. Surg. pp. 283–293 (2014)
28. Ungar, A.A.: Analytic hyperbolic geometry: Mathematical foundations and applications. World Scientific (2005)
29. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Proc. NeurIPS (2017)
30. Wang, W., Chen, C., Ding, M., Yu, H., Zha, S., Li, J.: Transbts: Multimodal brain tumor segmentation using transformer. In: Proc. MICCA (2021)

31. Wei, J., Wang, S., Huang, Q.: F^3net: fusion, feedback and focus for salient object detection. In: Proc. AAAI (2020)
32. Wu, Z., Su, L., Huang, Q.: Cascaded partial decoder for fast and accurate salient object detection. In: Proc. CVPR (2019)
33. Zeng, Y., Li, J., Zhao, Z., Liang, W., Zeng, P., Shen, S., Zhang, K., Shen, C.: Wet-unet: Wavelet integrated efficient transformer networks for nasopharyngeal carcinoma tumor segmentation. Science Progress p. 00368504241232537 (2024)
34. Zhang, H., Duan, Y., Liu, T., Zhang, W., Tang, H.: Kmunet: A novel medical image segmentation model based on kan and mamba. In: Proc. MICCA (2025)
35. Zhang, P., Ouyang, X., Peng, R.: UWT-Net: Mining low-frequency feature information for medical image segmentation. In: Proc. MICCA (2025)
36. Zhou, Z., Rahman Siddiquee, M.M., Tajbakhsh, N., Liang, J.: UNet++: A nested U-Net architecture for medical image segmentation. In: International workshop on Deep Learning in Medical Image Analysis (2018)