Surface Vision Mamba: Leveraging Bidirectional State Space Model for Efficient Spherical Manifold Representation

Rongzhao He[1], Weihao Zheng[1✉], Leilei Zhao[2], Ying Wang[1], Dalin Zhu[3], Dan Wu[4✉], and Bin Hu[1✉]

[1] Lanzhou University
[2] Harbin Institute of Technology
[3] Gansu Maternity and Child-care Hospital
[4] Zhejiang University
zhengweihao@lzu.edu.cn, danwu.bme@zju.edu.cn, bh@lzu.edu.cn

Abstract. Attention-based methods have demonstrated exceptional performance in modelling long-range dependencies on spherical cortical surfaces, surpassing traditional Geometric Deep Learning (GDL) models. However, their extensive inference time and high memory demands pose challenges for application to large datasets with limited computing resources. Inspired by the state space model in computer vision, we introduce the attention-free Vision Mamba (Vim) to spherical surfaces, presenting a domain-agnostic architecture for analyzing data on spherical manifolds. Our method achieves surface patching by representing spherical data as a sequence of triangular patches derived from a subdivided icosphere. The proposed Surface Vision Mamba (SiM) is evaluated on multiple neurodevelopmental phenotype regression tasks using cortical surface metrics from neonatal brains. Experimental results demonstrate that SiM outperforms both attention- and GDL-based methods, delivering 4.8 times faster inference and achieving 91.7% lower memory consumption compared to the Surface Vision Transformer (SiT) under the Ico-4 grid partitioning. Sensitivity analysis further underscores the potential of SiM to identify subtle cognitive developmental patterns. The code is available at https://github.com/Rongzhao-He/surface-vision-mamba.

Keywords: Cortical spherical manifold · State space model · Infant magnetic resonance imaging.

1 Introduction

Many methods have been developed for traditional Euclidean space data, such as Convolution Neural Networks (CNNs) and attention-based [2,20] approaches. CNNs calculate the weighted sum at each location, while attention-based methods treat the data as a sequence of patches. However, few models exist for non-Euclidean space data consist of graph, manifold and hyperbolic space data which have more complex geometries and distance metrics.

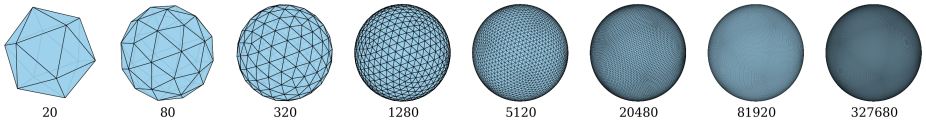

Fig. 1. Representative icosahedron discretized spherical surfaces with sequential subdivisions. The number of faces of each spherical surface is denoted under the surface.

Existing models for processing non-Euclidean data can be broadly classified into attention- and Geometric Deep Learning (GDL)-based [6,16,12,1,3,10,19,13] methods. Attention-based methods are effective in capturing long-range dependencies but are constrained in resource limited situations due to the quadratic complexity of the attention mechanism concerning sequence length, leading to higher memory consumption and slower inference time. Conversely, GDL-based methods are effective in handling complex geometric topology structure and distance metrics. However, they fail to extract global patterns, especially when applied to large-scale and highly intricate data, resulting in diminished performance. Thus, a key challenge for processing non-Euclidean data lies in improving efficiency while maintaining relatively excellent performance.

The emergence of State Space Models (SSMs) [8,11], has revitalized traditional sequence modeling, enabling efficient representation learning. A recent variant Mamba [7], has significantly surpassed traditional SSMs by integrating a selective scan mechanism that adapts parameters based on input and using a hardware-aware algorithm to parallelize scanning, thereby reducing memory I/O for more efficient inference. Motivated by ViT [5] and ViG [9], [21] adapted Mamba to computer vision, introducing a bidirectional SSM structure to address direction-sensitive challenges, termed Vision Mamba (Vim).

Non-Euclidean data, particularly spherical cortical surface, is characterized by high resolution, rich features, and intricate geometric shapes, as the cortical surface is inherently a high-dimensional manifold. While these data provide valuable insights into neurodevelopment, their effective representation poses a formidable challenge, often requiring a balance between performance and computational efficiency. Inspired by the efficiency of Vim, we extend its application to cerebral cortex analysis—an important yet underexplored area—by proposing Surface Vision Mamba (SiM). To adapt SiM to the unique characteristics of cortical surface data, we adjusted the input sequence length using various surface patching methods, as illustrated in Fig. 1.

The main contributions of this study can be summarized as follows:

1. We introduce SiM, an adaptation of Vim, as a generic backbone network for analyzing data mapped onto genus-zero surfaces.
2. Leveraging the suitability of Mamba for tasks with *long-sequence* and *autoregressive* characteristics [17], we explore the impact of varying input sequence length on surface data in non-Euclidean space. We further implement autoregressive pretraining to validate the effectiveness of this approach.

Table 1. Demographic and clinical information of the subjects.

	dHCP (N=526)	***Replication dataset (N=10)***
Birth age [weeks $^{+\text{days}}$], median (IQR)	38^{+5} (38^{+1} - 40^{+5})	37^{+2} (35^{+4} - 39^{+0})
Scan age [weeks $^{+\text{days}}$], median (IQR)	41^{+0} (39^{+4} - 42^{+2})	39^{+4} (39^{+2} - 40^{+5})
Birth weight, mean (SD)	3.09 (0.64)	2.06 (1.00)
Head circumference at scan, mean (SD)	34.83 (3.72)	-
Radiology score (1/2/3/4/5)	270/171/51/10/24	-
Gender (M/F)	288/138	6/11

3. Extensive experiments on three neurodevelopmental phenotype regression tasks, including the prediction of postmenstrual age (PMA) and long-term language and motor outcomes, demonstrate that our proposed SiM achieves promising performance compared to attention- and GDL-based models and is 4.8× faster than SiT and saves 91.7% GPU memory when performing batch inference under the Ico-4 grid partitioning.

2 Materials and Methods

Image acquisition and Dataset. The imaging data used in this work are from the publicly available Developing Human Connectome Project (dHCP) and a private dataset collected at Gansu Provincial Maternity and Child-care Hospital (GPMCH). We used T1-weighted (T1w) and T2-weighted (T2w) images to calculate morphometric metrics of cerebral cortex.

The dHCP is approved by the United Kingdom Health Research Ethics Authority (reference number: 14/LO/1169). Additionally, we collected T1w and T2w images of 10 infants from GPMCH (2020-GSFY-05). These images were acquired in the resolution of $0.8\times0.8\times1.6$ mm^3 with 0.8 mm overlap, and were reconstructed to 0.5 mm isotropic resolution.

Concerning the data from dHCP, a total of 526 infants covering preterm- and term-born neonates are enrolled. The neurodevelopmental assessments for these infants, conducted at 18 months of age using the Bayley-III Scales of Infant Development, can also be obtained. We used the following exclusion criteria: For PMA prediction, (*i*) we excluded the later scans of participants who were scanned twice; (*ii*) term-born neonates with focal abnormalities (radiology score > 2) were excluded. The remaining infants were then split into two subsets: *Subset 1*: 408 participants who were born and scanned between 34 and 45 PMA; *Subset 2*: 16 preterm infants who were born before 34 PMA and scanned at term-equivalent age ($>$ 37 PMA). For language and motor scores prediction, we retained the scans closest to 40 weeks for neonates scanned twice identified as *Subset 3*: 410 infants born between 23 and 43 gestational weeks (GA), with the scaled language and motor scores are 19.42 (5.26) and 20.55 (3.17), respectively, presented as mean (SD). We further utilized data from the GPMCH as a *Replication dataset*,

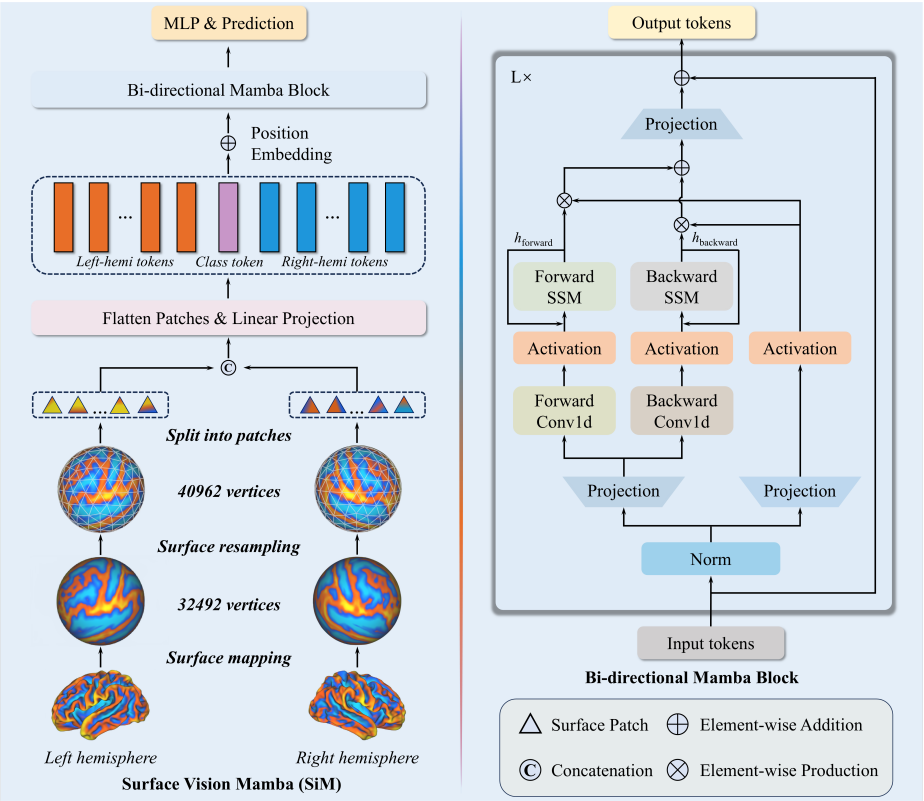

Fig. 2. Overview of the proposed Surface Vision Mamba (SiM) architecture.

consisting of 10 neonates to evaluate the generalization ability of the models. The demographic details are provided in Table 1[1].

Four cortical surface features (curvature, sulcal depth, cortical thickness and T1w/T2w myelination) were adopted. Each feature channel was normalized using Z-score. *Subset 1* and *3* were split into training, validation, and testing datasets in an 8:1:1 ratio. All data were registered to the dHCP 40-week spherical template, representing the cortical surface as an approximated sphere composed of triangles, with 32,492 vertices per hemisphere. We resampled the template sphere to a regular sixth-order icosphere (Ico-6) using barycentric interpolation.

Surface Vision Mamba. We proposed the SiM model, as shown in Fig. 2. Specifically, the input domain is divided into $2N$ patches, represented as $\widetilde{X} = \{\widetilde{L}, \widetilde{R} | \widetilde{L} \in \mathbb{R}^{N \times V \times C}, \widetilde{R} \in \mathbb{R}^{N \times V \times C}\}$, that V is the number of vertices in a patch, and C denotes the number of feature channels. This is then flattened

[1] IQR and SD denote interquartile range and standard deviation, respectively. 20 head circumference data were missed in dHCP. The unit is centimeters.

Table 2. Summary of the parameters for icospheres of different orders.

Icosphere Order	First	Second	Third	Fourth	Fifth
The number of patches (N)	80	320	1280	5120	20480
The number of vertices (V)	561	153	45	15	6
Input Dimension (VC)	2244	612	180	60	24
Sequence Length ($2N$)	160	640	2560	10240	40960

Table 3. Hyperparameters for all training strategies in PMA prediction. Specifically, scratch means that training from scratch, fine-tuning refers to using ImageNet pretraining weights in Vim. For self-supervised pretraining, AR means autoregressive. T, S, B represent tiny-size, small-size, base-size, respectively.

	Scratch			*Fine-tuning*			*AR Pretraining*			*AR Fine-tuning*		
	T	S	B	T	S	B	T	S	B	T	S	B
Epochs	1000			600			4000	3000	3000	600		
Batch size	32			32			32			32		
Optimizer	AdamW			AdamW			AdamW			AdamW		
Adam ϵ	1e-8			1e-8			1e-8			1e-8		
Adam (β_1, β_2)	(0.9, 0.999)			(0.9, 0.999)			(0.9, 0.999)			(0.9, 0.999)		
LR	5e-5			5e-5	5e-5	3e-5	1.5e-4			1.5e-4	1e-4	8e-5
LR decay	Linear			Linear			Cosine			Cosine		
Step size	500			200			-			-		
Gamma	0.5			0.5			-			-		
Warmup epochs	None			None			10			10		
Weight decay	1e-8			1e-8			0.5			1e-6		

to $X = \{L, R | L \in \mathbb{R}^{N\times(VC)}, R \in \mathbb{R}^{N\times(VC)}\}$. Next, we projected X into D-dimensional vectors using a trainable fully connected layer. Following the design of ViT and BERT [4], a learnable class token X_{cls} is concatenated between the left and right hemispheres. To retain positional information, standard 1D position embeddings E_{pos} are added to the patch features.

$$S_0 = [X_L^1 W; \cdots X_L^N; X_{cls}; X_R^1 W; \cdots; X_R^N W] + E_{pos} \tag{1}$$

where $W \in \mathbb{R}^{(VC)\times D}$, $E_{pos} \in \mathbb{R}^{(2N+1)\times D}$, $S_0 \in \mathbb{R}^{(2N+1)\times D}$ is the initial input of SiM, X_L^1 and X_R^1 represent the first patches of the left and right hemispheres, respectively. Specifically, layer l processes the input S_{l-1} as follows:

$$S_l = SiM(S_{l-1}) + S_{l-1}, T = LayerNorm(S_l^N), \widehat{p} = MLP(T) \tag{2}$$

Surface Patching Methods. We extend the sequence length by progressively subdividing the icosahedron into finer discrete levels, including first- to third-order icosphere, as summarized in Table 2. The icosphere subdivision is detailed in [18]. Different surface patching methods are visually represented in Fig. 1.

Training Methods. In this study, we explore three training strategies: (*i*) training models from scratch; (*ii*) fine-tuning pretrained weights from ImageNet (as released in Vision Mamba); and (*iii*) autoregressive pretraining due to the suitability of the Mamba for autoregressive modeling [14].

Table 4. Performance comparison on dHCP.

Methods	*Supervised*		*Fine-tuning*		*Autoregressive*		*P*	*M*
	MAE	**MSE**	**MAE**	**MSE**	**MAE**	**MSE**	**(M)**	**(G)**
MoNet	0.64±0.54	0.70±1.37	-	-	-	-	-	-
S2CNN	0.69±0.45	0.69±0.73	-	-	-	-	-	-
ChebNet	0.71±0.59	0.85±1.14	-	-	-	-	-	-
GConvNet	0.86±0.73	1.27±1.91	-	-	-	-	-	-
PointNet++	0.67±0.07	0.76±0.10	-	-	-	-	-	-
Spherical UNet	0.72±0.58	0.85±1.31	-	-	-	-	-	-
HRINet/1	0.75±0.67	1.05±0.05	-	-	-	-	10	-
SiT-Tiny/1	0.79±0.62	1.04±0.42	0.81±0.59	1.37±0.46	-	-	6	0.9
SiT-Small/1	0.81±0.57	1.06±0.36	0.87±0.63	0.93±0.28	-	-	22	3.6
SiT-Base/1	0.82±0.56	0.98±0.33	0.86±0.74	1.13±0.47	-	-	87	14.0
SiM-Tiny/1	0.85±0.64	0.91±0.31	0.76±0.66	1.21±0.26	1.03±0.70	1.87±0.42	7	1.9
SiM-Small/1	0.87±0.74	1.60±0.38	0.76±0.69	1.55±0.64	1.26±0.83	2.43±0.20	24	4.2
SiM-Base/1	0.86±0.69	1.20±0.04	0.84±0.63	1.24±0.19	1.06±0.67	1.34±0.30	92	15.4
HRINet/2	0.62±0.44	**0.39±0.25**	-	-	-	-	10	-
SiT-Tiny/2	0.69±0.52	0.47±0.36	0.66±0.58	0.78±0.01	-	-	6	3.5
SiT-Small/2	0.67±0.50	0.62±0.11	0.65±0.46	0.65±0.03	-	-	22	13.9
SiT-Base/2	0.64±0.57	0.57±0.55	0.72±0.47	0.58±0.20	-	-	86	55.0
SiM-Tiny/2	1.09±0.80	2.00±0.22	0.74±0.59	0.98±0.10	1.04±0.77	2.19±0.65	6	4.7
SiM-Small/2	0.98±0.80	1.89±0.38	0.60±0.49	0.52±0.10	1.12±0.92	3.50±1.79	24	16.5
SiM-Base/2	0.88±0.66	1.27±0.09	0.74±0.68	0.89±0.15	1.20±0.83	3.17±1.33	91	61.5
HRINet/3	*OOM*	*OOM*	-	-	-	-	10	-
SiT-Tiny/3	0.60±0.48	0.47±0.16	0.62±0.50	0.53±0.13	-	-	6	4.7
SiT-Small/3	0.60±0.51	0.54±0.41	0.60±0.43	0.42±0.16	-	-	24	16.5
SiT-Base/3	*OOM*	*OOM*	*OOM*	*OOM*	-	-	87	220.9
SiM-Tiny/3	1.09±0.76	2.79±1.30	0.60±0.46	0.85±0.35	0.91±0.81	3.62±2.73	7	18.9
SiM-Small/3	1.09±0.84	1.78±0.14	**0.56±0.50**	0.59±0.04	0.87±0.65	1.89±0.91	24	66.1
SiM-Base/3	1.03±0.81	2.30±0.73	0.62±0.45	0.65±0.07	1.09±0.71	2.27±0.76	93	245.5

3 Results and Discussion

Model Variants. The proposed SiM configurations are built upon three variants of Vim: Vim-Tiny, Vim-Small, and Vim-Base. We adopt concise notations for model size. For instance, SiM-B/3 refers to the "Base" variant with an input size of 2560×180, using an Ico-3 grid on the sphere.

Implementation Details. All the experiments are implemented with Python 3.10.13 and PyTorch library and conducted on 4 NVIDIA A100 GPUs. The Vim-Tiny†, Vim-Small† and Vim-Base weights are adapted that Vim-Tiny† and Vim-Small† are fine-tuned under long sequence. Training details are in Table 3.

Infant Brain Age Prediction. As shown in Table 4[2], the comparison results of SiM against benchmark GDL methods [6,16] and attention-based models [2,20] on PMA prediction of *Subset 1* on the three training strategies. Notably, when

[2] Case marked as ('-') represents unspecified values. ('*OOM*') means out of memory. Bold intricates the best performance in MAE and MSE, respectively. P denotes Parameters and M refers to MACs.

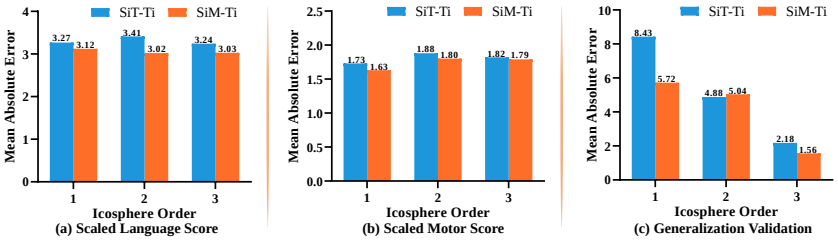

Fig. 3. Prediction performance comparison between SiT and SiM.

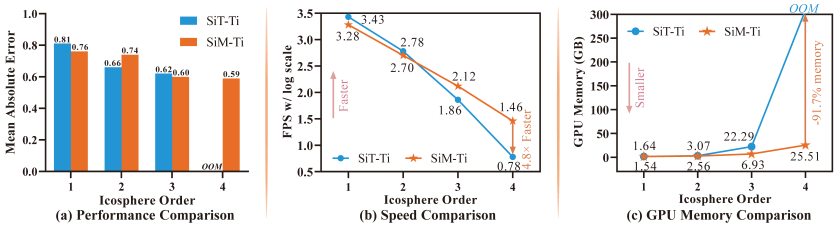

Fig. 4. PMA prediction performance and efficiency comparison between SiT and SiM.

fine-tuning with ImageNet pretraining weights, three variants of SiM outperform all the GDL methods using an Ico-3 grid. With comparable model parameters and MACs, SiM-S/3 achieves a performance of 0.56±0.50, surpassing SiT-S/3 (0.60±0.43). However, when training from scratch, the performance of all SiM variants decreased obviously, likely reflecting a tendency to overfit on small datasets due to the absence of strong prior weights. Although self-supervised pretraining has been shown to effectively strengthen model performance in previous studies, this benefit is less evident in our results. Actually, only the SiM-T/3 and SiM-S/3 exhibit improvements compared with training from scratch. This may be attributed to overfitting, which hampers generalization, or to the limited sample size, restricting the ability of model to capture sufficient features. In the *Subset 2*, we find that predicted brain age in preterm infants was significantly lower than chronological age, suggesting that preterm birth may delay brain development at term-equivalent age. Besides, we conducted further experiments:

1. Ablation studies on the decoder design in autoregressive pretraining, the best performance is achieved when the decoder depth is 1 and width is 256.
2. Language and motor scores prediction, which were shown in Fig. 3a and Fig. 3b, respectively. Generalization validation was displayed in Fig. 3c. The proposed SiM achieved the best performance compared to other methods.

Long Sequence and Efficiency Analysis. Fig. 4 illustrates the performance and efficiency of the tiny-sized SiM model across different surface patching methods. In terms of MAE, SiM slightly outperforms SiT across all other patching

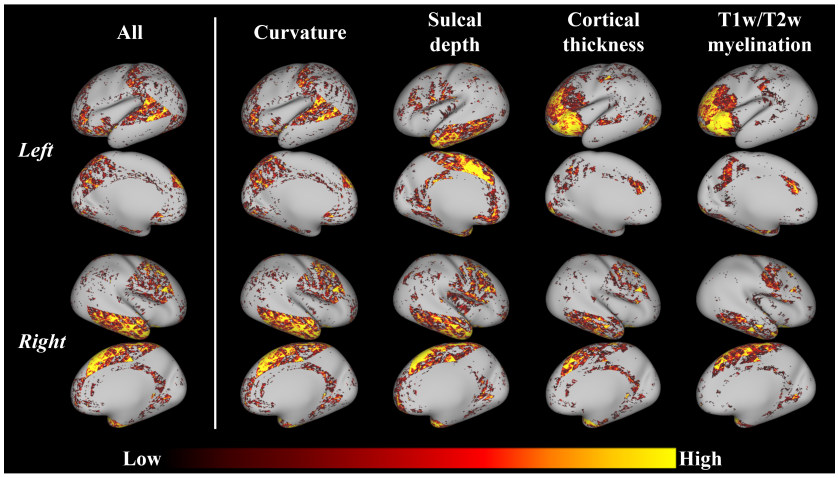

Fig. 5. Spatial distribution of informative vertices for PMA prediction.

methods except when using an Ico-2 grid, and both models exhibit decreasing MAE as the icosphere order increases (Fig. 4a). Regarding FPS (Frames Per Second), SiM is slightly slower than SiT when the icophere order is below 3 but surpasses it as the order increases (Fig. 4b). For GPU memory usage, SiM exhibits better efficiency with SiT when icosphere orders raises. Notably, when using an Ico-4 as grid, SiM is 4.8 times faster and consumes 91.7% less GPU memory compared to SiT (Fig. 4c). All experiments on efficiency analysis are conducted on a 40G A100 GPU. These results highlight the suitability of SiM for finer-grained tasks and its potential for practical clinical applications.

Cortical Regions with Significant Contributions to Age Prediction. We perform sensitive analysis [15] on the test dataset of *Subset 1* to evaluate the contribution of each vertex on cortical surface to brain age prediction, as illustrated in Fig. 5. For each vertex, we assessed four morphometric features. "All" nullifies all features, while others nullify one feature at a time per vertex.

The mean performance change for each vertex/feature was computed, normalized using Z-score. The intensity of the hot color signifies the influences of each vertex/feature. In the right hemisphere, key regions include the temporal lobe, precentral gyrus, and prefrontal and paracentral cortices. The left hemisphere shows a similar focus on the prefrontal cortex, sensory cortex, language areas, and parietal cortex when all features or only curvature are masked. Sulcal depth complements these findings by emphasizing the temporal lobe, central sulcus regions, and superior frontal and parietal areas. Cortical thickness and myelination highlight frontal regions, particularly the anterior insula, rostral middle frontal gyrus, and orbitofrontal cortex.

4 Conclusion

In this study, we introduced Surface Vision Mamba (SiM), a novel vision backbone with sub-quadratic time complexity, tailored for genus-zero surfaces. We validated SiM as a more robust and efficient alternative to SiT in the challenging task of neurodevelopmental phenotype prediction from cortical surface data. Leveraging the strengths of Mamba in handling *long-sequence* and *autoregressive modeling*, we extended sequence lengths using various surface patching methods and conducted autoregressive pretraining. While SiM demonstrated sensitivity to sequence length, the benefits of autoregressive pretraining were limited, likely due to constraints of small samples. The use of longer sequences facilitated finer-grained partitioning, enhancing the ability to identify potential pathological features critical in clinical applications. Furthermore, SiM offers faster inference speeds and lower GPU memory consumption, making it both efficient and practical. Sensitivity analysis also emphasized the interpretability of SiM, highlighting its potential utility in medical research and applications.

Acknowledgments. This work was supported by the STI2030-Major Projects (2021ZD0202002), the National Natural Science Foundation of China (62202212).

Disclosure of Interests. There are no conflicts of interest to declare.

References

1. Cohen, T.S., Geiger, M., Köhler, J., Welling, M.: Spherical cnns. arXiv preprint arXiv:1801.10130 (2018)
2. Dahan, S., Fawaz, A., Williams, L.Z., Yang, C., Coalson, T.S., Glasser, M.F., Edwards, A.D., Rueckert, D., Robinson, E.C.: Surface vision transformers: Attention-based modelling applied to cortical analysis. In: International Conference on Medical Imaging with Deep Learning. pp. 282–303. PMLR (2022)
3. Defferrard, M., Bresson, X., Vandergheynst, P.: Convolutional neural networks on graphs with fast localized spectral filtering. Advances in neural information processing systems **29** (2016)
4. Devlin, J.: Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805 (2018)
5. Dosovitskiy, A.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
6. Fawaz, A., Williams, L.Z., Alansary, A., Bass, C., Gopinath, K., da Silva, M., Dahan, S., Adamson, C., Alexander, B., Thompson, D., et al.: Benchmarking geometric deep learning for cortical segmentation and neurodevelopmental phenotype prediction. bioRxiv pp. 2021–12 (2021)
7. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
8. Gu, A., Goel, K., Ré, C.: Efficiently modeling long sequences with structured state spaces. arXiv preprint arXiv:2111.00396 (2021)
9. Han, K., Wang, Y., Guo, J., Tang, Y., Wu, E.: Vision gnn: An image is worth graph of nodes. Advances in neural information processing systems **35**, 8291–8303 (2022)

10. Kipf, T.N., Welling, M.: Semi-supervised classification with graph convolutional networks. arXiv preprint arXiv:1609.02907 (2016)
11. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Jiao, J., Liu, Y.: Vmamba: Visual state space model (2024), https://arxiv.org/abs/2401.10166
12. Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J., Bronstein, M.M.: Geometric deep learning on graphs and manifolds using mixture model cnns. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 5115–5124 (2017)
13. Qi, C.R., Yi, L., Su, H., Guibas, L.J.: Pointnet++: Deep hierarchical feature learning on point sets in a metric space. Advances in neural information processing systems **30** (2017)
14. Ren, S., Li, X., Tu, H., Wang, F., Shu, F., Zhang, L., Mei, J., Yang, L., Wang, P., Wang, H., et al.: Autoregressive pretraining with mamba in vision. arXiv preprint arXiv:2406.07537 (2024)
15. Saltelli, A.: Sensitivity analysis for importance assessment. Risk analysis **22**(3), 579–590 (2002)
16. Vosylius, V., Wang, A., Waters, C., Zakharov, A., Ward, F., Le Folgoc, L., Cupitt, J., Makropoulos, A., Schuh, A., Rueckert, D., et al.: Geometric deep learning for post-menstrual age prediction based on the neonatal white matter cortical surface. In: Uncertainty for Safe Utilization of Machine Learning in Medical Imaging, and Graphs in Biomedical Image Analysis: Second International Workshop, UNSURE 2020, and Third International Workshop, GRAIL 2020, Held in Conjunction with MICCAI 2020, Lima, Peru, October 8, 2020, Proceedings 2. pp. 174–186. Springer (2020)
17. Yu, W., Wang, X.: Mambaout: Do we really need mamba for vision? arXiv preprint arXiv:2405.07992 (2024)
18. Zhao, F., Wu, Z., Wang, L., Lin, W., Gilmore, J.H., ren Xia, S., Shen, D., Li, G.: Spherical deformable u-net: Application to cortical surface parcellation and development prediction. IEEE Transactions on Medical Imaging **40**, 1217–1228 (2021), https://api.semanticscholar.org/CorpusID:231301848
19. Zhao, F., Xia, S., Wu, Z., Duan, D., Wang, L., Lin, W., Gilmore, J.H., Shen, D., Li, G.: Spherical u-net on cortical surfaces: methods and applications. In: Information Processing in Medical Imaging: 26th International Conference, IPMI 2019, Hong Kong, China, June 2–7, 2019, Proceedings 26. pp. 855–866. Springer (2019)
20. Zhao, L., Zhu, D., Wang, X., Liu, X., Li, T., Wang, B., Yao, Z., Zheng, W., Hu, B.: An attention-based hemispheric relation inference network for perinatal brain age prediction. IEEE Journal of Biomedical and Health Informatics (2024)
21. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. In: Forty-first International Conference on Machine Learning