Efficient and Gender-adaptive Graph Vision Mamba for Pediatric Bone Age Assessment

Lingyu Zhou, Zhang Yi, Kai Zhou, and Xiuyuan Xu✉

Machine Intelligence Laboratory, College of Computer Science, Sichuan University, Chengdu, China
xuxiuyuan@scu.edu.cn

Abstract. Bone age assessment (BAA) is crucial for evaluating the skeletal maturity of children in pediatric clinics. The decline in assessment accuracy is attributed to the existence of inter-gender disparity. Current automatic methods bridge this gap by relying on bone regions of interest and gender, resulting in high annotation costs. Meanwhile, the models still grapple with efficiency bottleneck for lightweight deployment. To address these challenges, this study presents **G**ender-adaptive **G**raph **V**ision **Mamba** (GGVMamba) framework with only raw X-ray images. Concretely, a region augmentation process, called directed scan module, is proposed to integrate local context from various directions of bone X-ray images. Then we construct a novel graph Mamba encoder with linear complexity, fostering robust modelling for both within and among region features. Moreover, a gender adaptive strategy is proposed to improve gender consistency by dynamically selecting gender-specific graph structures. Experiments demonstrate that GGVMamba obtains state-of-the-art results with MAE of 3.82, 4.91, and 4.14 on RSNA, RHPE, and DHA, respectively. Notably, GGVMamba shows exceptional gender consistency and optimal efficiency with minimal GPU load. The code is available at https://github.com/SCU-zly/GGVMamba.

Keywords: Bone Age Assessment · X-ray · Vision Mamba · Linear Complexity · Efficient.

1 Introduction

Due to the early developmental changes in pediatric skeletal growth, adolescents routinely visit healthcare facilities for the periodic acquisition of X-ray images aimed at estimating bone age. Pediatric bone age assessment (BAA) is an effective early diagnostic method for detecting growth abnormalities in minors [7]. Conventional methods such as the Greulich-Pyle (GP) approach [9] and the Tanner and Whitehouse method [21], which depend on manual expertise, exhibit significant subjective errors and low efficiency.

Human prior knowledge [21] indicates that models should prioritize the bone's epiphyseal regions of interest (ROI) and causal relationships among highly heterogeneous regions. Early deep-learning methods divide X-ray images into regions and extract features separately. The introduction of Central Positions of

Anatomical ROI (CPAR) enhanced BAA performance through BoNet [5] and SIMBA [8]. PRSNet [13] embeds more effective contextual information in part representations. Doctor Imitator [2] introduces a dual-graph attention module to learn relationships between features. Leveraging attention mechanisms, several studies [18,19] used the Vision Transformer [4] to emphasize image heterogeneity. Recently, a set of one-stage techniques that rely exclusively on image-level annotation data has been developed. MMANet [24] introduces an additional residual spatial attention module to address biases stemming from conventional residual structures and produce more distinct attention maps. Wang et al. [22] employed multi-instance learning to integrate gender-specific details derived from individual images during the prediction of bone age. Nevertheless, the high cost of annotation and computation in two-stage and one-stage methods hinders their clinical utility. The researches [9,21] also indicate significant discrepancies in the areas of interest for BAA between different genders. However, most models [2,5,8,15,24] take gender as input in the highest dimension, rarely allowing the model itself to focus on the inherent differences between genders.

Particularly, PEAR-Net [16] achieves effective performance without relying on gender input. However, it has limitations to implement in terms of computational efficiency and parameter size. Inspired by Mamba [10], which achieves remarkable accuracy in Natural Language Processing (NLP) with linear computational complexity, we introduce the innovative Gender-adaptive Graph Vision Mamba (GGVMamba) based on pure vision Mamba frameworks [17,25]. In addition, GGVMamba addresses gender consistency within the model using graph structures. The outlined contributions are as follows:

1) **We introduce a novel directed scan module**. The directed scan module highlights the heterogeneous characteristics of bone X-ray images both column-wise and row-wise (refer to Fig. 1). This module transforms a non-directed sequence into four directed sequences, enhances region features, and improves the generalization ability across various datasets.
2) **We propose graph Mamba encoder** in achieving two-stage feature extraction capability with one-stage efficiency. We employ bidirectional compression modelling to assist GGVMamba in capturing dense region features. With the relation Mamba learner, the encoder robustly learns the graph structures between regions by linear long-range attention, thereby improving the precision of GGVMamba.
3) **In GGVMamba, a gender adaptive strategy** is formulated on graph regularization constraints. In particular, utilizing the representation of graph nodes in the Mamba latent space as intra-graph consistency, this strategy focuses on balancing intra-graph and inter-graph consistency to enhance gender consistency.

To improve real-world applicability, divergent from conventional singular dataset validation, we gathered data from three public datasets: Radiological Society of North America (RSNA) challenge [12] for automatic BAA methods, Radiological Hand Pose Estimation [5] (RHPE) dataset and Digital Hand Atlas (DHA) [7]

dataset. The experimental results show that GGVMamba has the lowest MAE compared to state-of-the-art models, with remarkable gender consistency and efficiency.

2 Method

2.1 Problem Formulation

Let $X \in \mathbb{R}^{H \times W \times C}$ represent the input of a hand X-ray image. Following Patch Embedding, the sequence $[x_1, x_2, \cdots, x_n \in X]$ captures patches while preserving 2D positional information, where the ordering from 1 to n corresponds to the top-right. h serves as the latent state within the model space. As shown in Fig. 1, $\mathcal{W}_0$ and $\mathcal{W}_1$ represent the output male-specific graph M and female-specific graph F, respectively.

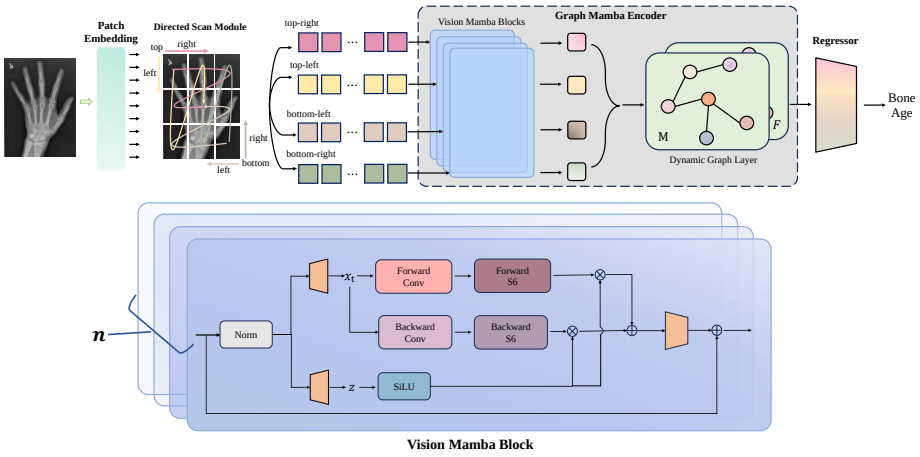


Fig. 1. Overview of GGVMamba framework.

2.2 Directed Scan Module

The concept of "directed" is narrowly defined, indicating the order of semantic relevance information among patches. Natural language possesses an inherent semantic sequence order, whereas visual data lacks this order. From this standpoint, the patches exhibit undirected characteristics after patch embedding.

We characterize semantic relevance information as causal information, which relies on the assumption that sequential dependencies exist within the patch sequence. According to [13], X-ray patches exhibit causal information due to the interrelated structure of bone elements. We introduce a patch augmentation scheme named directed scan module, designed to accentuate local context

without ROI annotation. The objective is to guarantee that each bone element combines causal information by linear projection. Unlike the cross scan [17], we guarantee spatial adjacency for each patch. Unlike the bidirectional scan [25], the directed scan module utilizes four complementary traversal paths to address the limited contextual awareness.

The process involves collecting patches from four specific directions: *top-right, top-left, bottom-right*, and *bottom-left*. Meanwhile, this module integrates with patch embedding through backpropagation, ensuring a smoother augmentation process for region features. Furthermore, data augmentation through linear transformations can enhance the model's generalization capability [23]. Combining transformations applied to the entire X-ray images, such as flipping and translation, scanning based on directed sequences exhibits greater robustness in the unseen bone age domain.

2.3 Graph Mamba Encoder

Mamba Scheme Given input sequence $x(t) \in \mathbb{R}$ and the corresponding output is sequence $y(t) \in \mathbb{R}$, the transition from intermediate state $h(t) \in \mathbb{R}^N$ to output follows a linear Ordinary Differential Equation (ODE): $h'(t) = Ah(t) + Bx(t), y(t) = Ch(t)$. The $h(t)$ signifies the intermediate latent state in the transition from $x(t)$ to $y(t)$. This system is defined by matrices $A \in \mathbb{R}^{N\times N}$, $B \in \mathbb{R}^{N\times 1}$, and $C \in \mathbb{R}^{1\times N}$. In this context, $h'(t)$ denotes the first derivative of $h(t)$. State Space Model(SSM) uses discrete transformations, employing the timescale variable Δ to discretize parameters into $\overline{A}$ and $\overline{B}$. The common method for discretization involves employing a Zero-Order Hold (ZOH), which streamlines the transition from a continuous system to a discrete system, as follows:

$$\begin{cases} \overline{A} = \exp(\Delta A), \\ \overline{B} = (\Delta A)^{-1}(\exp(\Delta A) - I) \cdot \Delta B, \\ h_t = \overline{A}h_{t-1} + \overline{B}x_t, \\ y_t = Ch_t. \end{cases} \quad t = 0, 1, \cdots, n \tag{1}$$

Gu et al. [11] reveals that the earlier Eq.1 is equivalent to the global convolution $\overline{K}{:}\overline{K} = (C\overline{B}, C\overline{AB}, \cdots, C\overline{A}^{n-1}\overline{B}), h = X * \overline{K}$. The crucial distinction between SSM and Mamba is manifested in the expression of multiple parameters Δ, B, C as input-dependent functions. Gu et al. [10] encapsulated the data-dependent mechanism as the Selective Scan Structured State Space model (S6).

Vision Mamba Block. In terms of expressing feature latent spaces within patches, attention-based methods typically necessitate quadratic time complexity calculations. Meanwhile, linear mapping methods do not offer fine-grained interpretation. Consequently, we propose a modeling methodology for the internal representation of region features, referred to as the vision Mamba block (VMB). VMB enhances linear attention inside the highly heterogeneous regions by integrating a bidirectional S6 mechanism. Various observation sequences are

individually mapped to the latent Mamba space through four mutually uncoupled VMBs. VMB leveraged Mamba to achieve a breakthrough in constructing linear-complexity blocks while addressing the limitations of context compression. The details of VMB are illustrated in Fig. 1. The input X undergoes bidirectional transmission through both forward and backward flows. After a sequence convolution to filter out low-specificity features, we obtain the derived sequence X'. For a vector of length n, each x'_t is transformed through a linear mapping, resulting in $\overline{A}_t, \overline{B}_t, C_t, \Delta_t$. Subsequently, we integrate the forward and backward S6 bidirectional computations, focusing linear attention on high-specificity context. The final result is processed through a SiLU(z) activation function via a residual connection gate to accelerate training convergence.

Dynamic Graph Layer. The relationships between the highly heterogeneous regions in bone age X-rays exhibit variations across different gender domains. These relationships are highly correlated with the accuracy of bone age prediction. We offer a gender-adaptive solution of gender-specific feature representation termed the dynamic graph layer (DGL). DGL allows us to dynamically select gender-specific graph, enabling the exploration and learning of intricate relationships between genders. Instead of simply treating the four directed features as graph nodes, DGL reshapes the latent space. This aids in the stereoscopic modelling of directed feature sequences, helping GGVMamba learn the global context within the bones. We introduce DGL with relation Mamba learner to capture graph features with low computation. The learner first initials the adjacency matrix of the gender-specific dynamic graph, denoted as $(g = 0, 1), \overline{W}_g \in \mathbb{R}^{n_d \times n_d}$, where n_d represents the number of nodes. This description addresses the input hidden state $\omega_g \in \mathbb{R}^{L \times n_d}$ with L denoting the number of latent dimensions. The matrix $\overline{W}_g$ is learned based on:

$$\begin{cases} \omega_g^* = \overline{\mathcal{A}}\overline{\mathcal{B}}\omega_g + \overline{\mathcal{B}}\omega_g, \\ \overline{W}_g = \text{softplus}(\overline{\mathcal{C}}\omega_g^*). \end{cases} \tag{2}$$

In this context, $\overline{\mathcal{A}} \in \mathbb{R}^{n_d \times L}, \overline{\mathcal{B}} \in \mathbb{R}^{L \times L}, \overline{\mathcal{C}} \in \mathbb{R}^{n_d \times L}$ are all data dependent on the hidden state ω_g, and softplus serves as the activation function. Then we add a k-nearest neighbor (KNN) graph $\overline{W}_g^{knn}$ to matrix $\overline{W}_g$, defining each node's k-nearest neighbors based on cosine similarity reinforces the graph's information on node relationships. Lastly, the resulting adjacency matrix W_g is expressed as $W_g = \lambda \overline{W}_g^{knn} + (1 - \lambda)\overline{W}_g$. The hyper-parameter $\lambda \in [0, 1)$ serves to adjust the training direction of the relation Mamba learner. The matrix $\mathcal{W}_g$ is acquired through the parameterized Multi-Layer Perceptron (MLP), learning the internal graph structure between high-dimensional spatial features.

2.4 Gender Adaptive Strategy

A gender adaptive strategy is introduced to strengthen the learning of gender consistency in dynamic selection. Dynamic selection serves as the abstract

expression of multi-graph regularization, which aims to attain desirable properties like smoothness, sparsity, and connectivity [3,20,26]. Notably, minimizing the Dirichlet energy [1] is crucial to prevent over-smoothing of graph nodes. This study accomplishes this by normalizing the graph Laplacian operator $L = D - \mathcal{W}_g$, making the Dirichlet energy independent of node degrees, where D represents the degree matrix of $\mathcal{W}_g$. However, to prevent trivial solutions during regularization, we suggest graph sparsification [14] and ensure strong connectivity. The constraints for the objective function of dynamic selection are detailed in Eq. 3. The $tr(\cdot)$ denotes the trace operation on a matrix, $||\cdot||_F$ represents the Frobenius norm of a matrix, and $\mathbf{1}$ is a column vector consisting of identical elements set to 1, with a length equal to the number of nodes in the graph.

$$\begin{cases} \mathcal{L}_{smooth}(h_g, \mathcal{W}_g) = \frac{1}{N^2} tr(h_g^T L h_g), \\ \mathcal{L}_{sparse}(\mathcal{W}_g) = \frac{1}{N^2} ||\mathcal{W}_g||_F^2, \\ \mathcal{L}_{degree}(\mathcal{W}_g) = -\frac{1}{N} \mathbf{1}^T log(\mathcal{W}\mathbf{1}). \end{cases} \tag{3}$$

In this mechanism, integrating the weighted average of attributes enhances inter-graph consistency. The regularization process in inter-graph consistency representation requires a balance between graph smoothness, sparsity, and connectivity. Concurrently, we apply the Smooth L1 loss $\mathcal{L}_{bone}$ in the bone age output prediction to capture intra-graph contextual similarity. The gender adaptive strategy combines desirable properties to strike a balance between inter-graph and intra-graph consistency: $\mathcal{L}_{BAA} = \frac{1}{2}\sum_{g=0}^{2}(\alpha\mathcal{L}_{smooth}(h_g, \mathcal{W}_g) + \beta\mathcal{L}_{sparse}(\mathcal{W}_g) + \gamma\mathcal{L}_{degree}(\mathcal{W}_g)) + (1-\alpha-\beta-\gamma)\mathcal{L}_{bone}$. Here, $\alpha, \beta, \gamma \in [0, 1]$ are hyper-parameters adjusted during training.

3 Experiments

3.1 Materials

The experimental procedures applied to the RSNA and RHPE datasets are based on their origin settings. The DHA dataset consists of 1,400 digitized left hand radiographs, divided into training, validation, and testing sets in a ratio of 7:2:1. All X-ray inputs are normalized and resized to 512×512 pixels. The software environment includes a PyTorch 2.1.1 base framework with Python 3.10 running on an NVIDIA GeForce RTX 4090 GPU. We allocate 12 VMBs for every directed sequence. The optimal performance is achieved when α, β, γ is set to 0.2 and λ to 0.8. The Adam optimizer is used with ϵ set to 1×10^{-8}. The learning rate scheduler starts at 0.01 and follows a cosine decay. Overall, training is done with a batch size of 32.

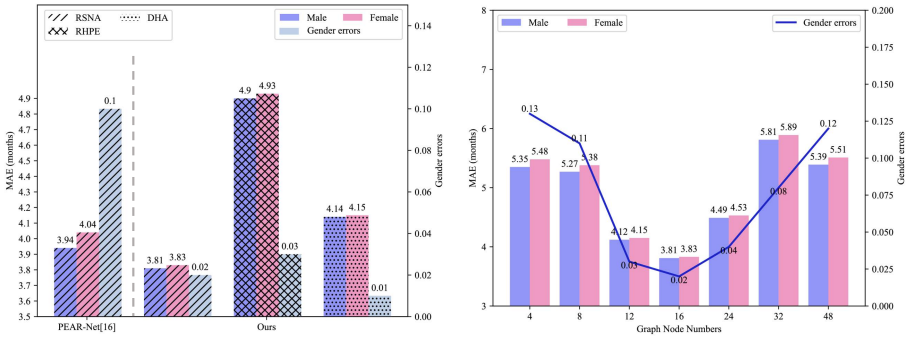
We present all results from the test set in terms of Mean Absolute Error (MAE). Gender errors denote the absolute errors between MAE values within the subset of gender. We also use metrics like floating point operations per second (FLOPs) and model size to evaluate model efficiency in the experiments.

Table 1. Quantitative comparison of different studies on public datasets.

Method	Without ROI annotation	Without gender input	FLOPs↓ ($\times 10^9$)	Model size↓	MAE(months)↓		
					RSNA	RHPE	DHA
Bonet [5]	×	×	123.2	17.8M	4.14	7.60	-
PEAR-Net [16]	✓	✓	180.0	42.0M	3.99	-	-
DI [2]	×	×	13.1	9.8M	4.30	8.15	-
SIMBA [8]	×	×	138.2	38.7M	-	5.47	-
Fahmida et al. [6]	✓	×	73.5	16.3M	-	-	4.21
MMANet [24]	✓	×	45.8	30.8M	3.88	-	-
Ours	✓	✓	**11.3**	**8.5M**	**3.82**	**4.91**	**4.14**

3.2 Compare with State-of-the-Arts

Table 1 compares GGVMamba with the state-of-the-art methods. GGVMamba achieves the lowest MAE of 3.82 months, 4.91 months, and 4.14 months on three public datasets, respectively. Generally, GGVMamba significantly reduces the challenges of high computation and model complexity compared to previous studies [2,5,6,8,16,24]. Specifically, Our model shows a 13.7% efficiency improvement and a 13.2% reduction in computational complexity when compared to the state-of-the-art DI [2]. This progress is attributed to a reduction of 1.3 million parameters and a decrease in FLOPs from 13.1×10^9 to 11.3×10^9.



(a) Comparative experiments of inter-gender disparity on public datasets

(b) Ablation study on different Graph Node Numbers

Fig. 2. Comparative experiments and ablation study of inter-gender disparity.

Fig. 2(a) illustrates the MAE and gender errors on gender subsets RSNA, RHPE, and DHA. Compared to the PEAR-Net [16] validated on RSNA, our approach reduces inter-gender disparity from 0.1 to 0.02. Furthermore, GGVMamba achieves a further reduction in MAE. In addition, GGVMamba demon-

strates remarkable gender consistency in RHPE and DHA, affirming the model's generalization.

3.3 Ablation Studies

Table 2. Ablation study on RSNA dataset.

Exp.	Directed scan module	Graph Mamba encoder	Gender adaptive strategy	MAE(month)↓	Gender errors↓
1				5.35	0.55
2	✓			4.86	0.53
3	✓	✓		4.21	0.39
4	✓		✓	4.67	0.10
5	✓	✓	✓	**3.82**	**0.02**

Table 2 presents ablation experiments on the effectiveness of the proposed components for GGVMamba. In each experiment (Exp.), inactive modules are substituted with the ViT-B/16 structure [4]. As evidenced by Exp. 1 and Exp. 2, the directed scan Module improves the model's precision by incorporating local attention. A comparison between Exp.2 and Exp.3 shows a significant improvement in the model's MAE, providing strong evidence for the graph Mamba encoder's role in enhancing output accuracy. Contrasting Exp.2 and Exp.4 reveals a noticeable decrease in errors within the gender subset, indicating that the gender adaptive strategy ensures stable gender consistency and increases precision enhancement. Exp.5, when compared to Exp.3 & Exp.4, illustrates that combining the three components leads to optimal outcomes.

Fig. 2(b) depicts a diminishing trend in both MAE and gender errors as the number of graph nodes increases. The optimal values for MAE and gender errors are reached when the graph nodes reach 16. Beyond this point, as the graph nodes increase further, MAE displays some fluctuations, while gender errors show an upward trajectory.

4 Conclusion

Leveraging our expertise, we introduce the pioneering Graph Vision Mamba network for BAA, achieving robust high accuracy with a one-stage, low-annotation, and computationally efficient approach. GGVMamba effectively integrates the highly heterogeneous epiphyseal regions, and addresses gender consistency in bone age X-ray images. Furthermore, GGVMamba enhances the method's generalization across three benchmark datasets by employing patch-level data augmentation. Our approach demonstrates superior accuracy, generalization, and

gender consistency, effectively solving prevalent clinical challenges of low precision, inefficiency, and frequent domain transfer issues. Future work will further enhance the model's zero-shot learning on unseen datasets and explore additional medical applications.

Acknowledgement. This work was supported in part by the National Natural Science Foundation of China under Grant 62106163, in part by the Natural Science Foundation Project of Sichuan Province under Grant 2023YFG0283, in part by the Key Research Support Project of Chengdu City under Grant 2019-YF09-00228-SN, and in part by the Key Research Project of Sichuan province under Grant 2022YFS0190.

Disclosure of Interests. We have thoroughly examined and hereby declare that the authors of this paper have no potential conflicts of interest to disclose.

References

1. Cai, C., Wang, Y.: A note on over-smoothing for graph neural networks. arXiv preprint arXiv:2006.13318 (2020)
2. Chen, J., Yu, B., Lei, B., Feng, R., Chen, D.Z., Wu, J.: Doctor imitator: A graph-based bone age assessment framework using hand radiographs. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. pp. 764–774. Springer (2020)
3. Chen, Y., Wu, L., Zaki, M.: Iterative deep graph learning for graph neural networks: Better and robust node embeddings. Advances in neural information processing systems **33**, 19314–19326 (2020)
4. Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929 (2020)
5. Escobar, M., González, C., Torres, F., Daza, L., Triana, G., Arbeláez, P.: Hand pose estimation for pediatric bone age assessment. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. pp. 531–539. Springer (2019)
6. Fahmida, M., Khaliluzzaman, M., Hossain, S.M.M., Deb, K.: Automated bone age assessment using deep learning with attention module. In: International Conference on Intelligent Computing & Optimization. pp. 217–226. Springer (2023)
7. Gertych, A., Zhang, A., Sayre, J., Pospiech-Kurkowska, S., Huang, H.: Bone age assessment of children using a digital hand atlas. Computerized medical imaging and graphics **31**(4-5), 322–331 (2007)
8. González, C., Escobar, M., Daza, L., Torres, F., Triana, G., Arbeláez, P.: Simba: Specific identity markers for bone age assessment. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2020: 23rd International Conference, Lima, Peru, October 4–8, 2020, Proceedings, Part VI 23. pp. 753–763. Springer (2020)

9. Greulich, W.W., Pyle, S.I.: Radiographic atlas of skeletal development of the hand and wrist. The American Journal of the Medical Sciences **238**(3), 393 (1959)
10. Gu, A., Dao, T.: Mamba: Linear-time sequence modeling with selective state spaces. arXiv preprint arXiv:2312.00752 (2023)
11. Gu, A., Dao, T., Ermon, S., Rudra, A., Ré, C.: Hippo: Recurrent memory with optimal polynomial projections. Advances in neural information processing systems **33**, 1474–1487 (2020)
12. Halabi, S.S., Prevedello, L.M., Kalpathy-Cramer, J., Mamonov, A.B., Bilbily, A., Cicero, M., Pan, I., Pereira, L.A., Sousa, R.T., Abdala, N., et al.: The rsna pediatric bone age machine learning challenge. Radiology **290**(2), 498–503 (2019)
13. Ji, Y., Chen, H., Lin, D., Wu, X., Lin, D.: Prsnet: part relation and selection network for bone age assessment. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. pp. 413–421. Springer (2019)
14. Kalofolias, V.: How to learn a graph from smooth signals. In: Artificial intelligence and statistics. pp. 920–929. PMLR (2016)
15. Liu, C., Xie, H., Liu, Y., Zha, Z., Lin, F., Zhang, Y.: Extract bone parts without human prior: End-to-end convolutional neural network for pediatric bone age assessment. In: Medical Image Computing and Computer Assisted Intervention–MICCAI 2019: 22nd International Conference, Shenzhen, China, October 13–17, 2019, Proceedings, Part VI 22. pp. 667–675. Springer (2019)
16. Liu, C., Xie, H., Zhang, Y.: Self-supervised attention mechanism for pediatric bone age assessment with efficient weak annotation. IEEE Transactions on Medical Imaging **40**(10), 2685–2697 (2020)
17. Liu, Y., Tian, Y., Zhao, Y., Yu, H., Xie, L., Wang, Y., Ye, Q., Liu, Y.: Vmamba: Visual state space model. arXiv preprint arXiv:2401.10166 (2024)
18. Payer, C., Štern, D., Bischof, H., Urschler, M.: Integrating spatial configuration into heatmap regression based cnns for landmark localization. Medical image analysis **54**, 207–219 (2019)
19. Ren, X., Li, T., Yang, X., Wang, S., Ahmad, S., Xiang, L., Stone, S.R., Li, L., Zhan, Y., Shen, D., et al.: Regression convolutional neural network for automated pediatric bone age assessment from hand radiograph. IEEE journal of biomedical and health informatics **23**(5), 2030–2038 (2018)
20. Tang, S., Dunnmon, J.A., Liangqiong, Q., Saab, K.K., Baykaner, T., Lee-Messer, C., Rubin, D.L.: Modeling multivariate biosignals with graph neural networks and structured state space models. In: Conference on Health, Inference, and Learning. pp. 50–71. PMLR (2023)
21. Tanner, J., Whitehouse, R., Takaishi, M.: Standards from birth to maturity for height, weight, height velocity, and weight velocity: British children, 1965. ii. Archives of disease in childhood **41**(220), 613 (1966)
22. Wang, C., Wu, Y., Wang, C., Zhou, X., Niu, Y., Zhu, Y., Gao, X., Wang, C., Yu, Y.: Attention-based multiple-instance learning for pediatric bone age assessment with efficient and interpretable. Biomedical Signal Processing and Control **79**, 104028 (2023)
23. Wu, S., Zhang, H., Valiant, G., Ré, C.: On the generalization effects of linear transformations in data augmentation. In: International Conference on Machine Learning. pp. 10410–10420. PMLR (2020)
24. Yang, Z., Cong, C., Pagnucco, M., Song, Y.: Multi-scale multi-reception attention network for bone age assessment in x-ray images. Neural Networks **158**, 249–257 (2023)

25. Zhu, L., Liao, B., Zhang, Q., Wang, X., Liu, W., Wang, X.: Vision mamba: Efficient visual representation learning with bidirectional state space model. arXiv preprint arXiv:2401.09417 (2024)
26. Zhu, Y., Xu, W., Zhang, J., Du, Y., Zhang, J., Liu, Q., Yang, C., Wu, S.: A survey on graph structure learning: Progress and opportunities. arXiv preprint arXiv:2103.03036 (2021)