

AHPR-Net: Anatomy-aware Hierarchical Prompting Refinement Network for Spine Image Segmentation

Junyong Zhao¹, Kun Wang¹, Dingwei Fan¹, Qi Dou³, Liang Sun^{1,2} (✉), and Daoqiang Zhang^{1,2} (✉)

¹ College of Artificial Intelligence, Nanjing University of Aeronautics and Astronautics, Key Laboratory of Brain-Machine Intelligence Technology, Ministry of Education, Nanjing, 211106, China.

² Shenzhen Research Institute, Nanjing University of Aeronautics and Astronautics, Shenzhen, 518052, China.

³ Department of Computer Science and Engineering, The Chinese University of Hong Kong, Hong Kong, China.
sunl@nuaa.edu.cn, dqzhang@nuaa.edu.cn

Abstract. Accurate spine segmentation is pivotal for surgical planning, yet remains hindered by complex anatomical structures and ambiguous boundaries. Recently, Segment Anything Model (SAM)-based methods have shown impressive generalization. However, single-level prompts often fail to fully capture semantic discriminability and anatomical consistency in spine imaging. To address these challenges, in this paper, we propose an anatomy-aware hierarchical prompting refinement network for spine image segmentation, called AHPR-Net. Specifically, to enhance the semantic richness of prompt information, we propose a hierarchical multi-level prompting strategy that jointly incorporates categorical semantics, anatomical relational priors, and morphological shape information to provide anatomically grounded guidance for spine segmentation. Then, to effectively fuse multi-level prompt information, we design a dynamic prompt tuning module that adaptively selects and enhances the most informative prompts and improves the perception ability of the complex anatomical structures of the spine. Furthermore, richer anatomical prompts alone cannot eliminate boundary ambiguity and the resulting uncertainty. Therefore, we propose an uncertainty-guided refinement strategy that explicitly models boundary uncertainty to improve structural reliability during training. Experimental results on an in-house spine dataset and a public SPIDER dataset demonstrate that our proposed method achieves superior segmentation performance compared to several state-of-the-art methods.

Keywords: Spine Segmentation · Prompt Learning · Uncertainty · Dynamic Tuning.

1 Introduction

Accurate segmentation of the spine is a fundamental prerequisite for safe and personalized surgical planning [14,19]. Surgical interventions critically depend on precise localization and morphological characterization of key anatomical structures, including vertebral bodies and intervertebral discs [7,10]. Consequently, developing anatomically

consistent spine segmentation methods is essential for reliable preoperative planning and surgical decision-making.

In recent years, a growing number of deep learning-based methods have been proposed for spine image segmentation and have achieved impressive performance [12,15,24]. Previous methods have widely designed meticulous features and networks to enhance spine segmentation performance. For example, Qian *et al.* [15] proposed a shape-based attention network for the learning representation of a lumbar spine MR image. Deng *et al.* [5] proposed a complementary spine segmentation network. The method achieves progressive feature fusion by using an attention refinement module and a feature fusion unit. Liu *et al.* [12] proposed a grayscale self-adjusting network for 3D lumbar anatomy segmentation. The method alleviates the imbalance between foreground and background regions in CT images. Despite these advances, most existing methods rely on deterministic predictions and cannot quantify uncertainty during segmentation. In spinal image analysis, uncertainty arises from two main sources: data uncertainty, due to annotation noise and blurred anatomical boundaries, and model uncertainty, caused by limited training data and restricted generalization. Accurately capturing these uncertainties is essential for providing reliable predictions, particularly in cases involving complex anatomy or unclear boundaries.

Recently, segment anything model (SAM)-based segmentation methods have demonstrated superior performance in medical image segmentation tasks [11,9]. This is attributed to a prompt learning strategy, which uses additional guidance information to guide the segmentation process towards the region of interest. For example, the box-based or point-based prompt method can provide a coarse spatial localization of the region of interest [4]. However, these prompts often lack flexibility and provide limited information. To improve the accuracy and flexibility of the prompt, Zhao *et al.* [23] proposed an uncertainty-driven edge prompt generation network for medical image segmentation. The method uses an edge prompt generator to generate edge prompt information and further enhances the structural representation. Similarly, Fan *et al.* [6] proposed a multi-atlas guided pseudo mask prompts network for spine segmentation. This method can learn prior knowledge of anatomical structures by using multi-atlas information. However, current prompts only rely on low-level image prompts, neglecting high-level clinical semantics. By leveraging large language models (LLMs) [20,21] to capture external medical knowledge, multi-level prompt information can be constructed and further improve the segmentation performance for ambiguous boundaries or complex pathologies.

To solve the above challenges, we propose an anatomy-aware hierarchical prompting refinement network for spine image segmentation, called AHPR-Net. Specifically, we propose a hierarchical multi-level prompting strategy to fully leverage the complementary advantages of different levels of prompt information by integrating categorical semantics, anatomical relational priors, and morphological shape information. Then, we design a dynamic prompt tuning module to further enhance the feature representation capability of prompt information. This mechanism utilizes an adaptive selection module to adaptively select prompt information and enhance the perception ability of the complex anatomical structures of the spine. Finally, we propose an uncertainty-guided refinement module to alleviate boundary ambiguities during training and significantly

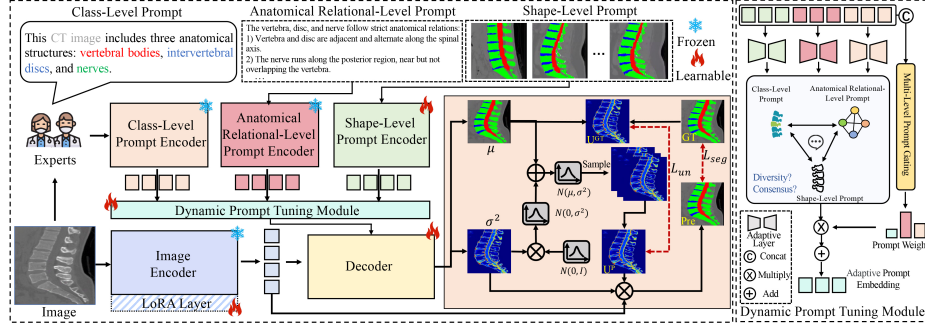


Fig. 1: Illustration of our proposed anatomy-aware hierarchical prompting refinement framework for spine image segmentation.

improve the anatomical reliability of the segmentation results. Experimental results on an in-house spine dataset and a public SPIDER dataset demonstrate that our method achieves better segmentation results than the comparison methods. To summarize, our contributions include:

- To enhance the semantic richness of prompt information, we propose a hierarchical multi-level prompting strategy that jointly incorporates categorical semantics, anatomical relational priors, and morphological shape information to enhance both semantic discriminability and structural consistency.
- To effectively fuse heterogeneous prompt information, we design a dynamic prompt tuning module to adaptively select and improve the perception ability of complex spinal anatomy.
- To address boundary ambiguity and uncertainty, we propose an uncertainty-guided refinement strategy to improve structural reliability during training.

2 Methodology

2.1 Overview

As shown in Fig. 1, we propose a novel spine image segmentation framework that integrates anatomy-aware multi-level prompt learning and uncertainty-guided refinement. Specifically, we first propose a hierarchical multi-level prompting strategy that combines categorical semantics, anatomical relational priors, and morphological shape information to enhance semantic discrimination and structural consistency. Then, we design a dynamic prompt tuning module to adaptively select prompt features and feed them into a SAM-based encoder-decoder network to generate a coarse segmentation mask. Finally, we employ an uncertainty-guided refinement strategy to further improve segmentation performance.

2.2 Hierarchical Multi-level Prompting

Accurate segmentation of spinal structures requires capturing both the global anatomical and fine-grained structural details. Conventional methods relying on a single type of prompt (*i.e.*, points, boxes, or text) often fail to adequately represent the spine complexity, including vertebral class distinctions, inter-vertebral topological relationships, and subtle shape variations. Therefore, we propose a hierarchical multi-level prompting strategy to jointly integrate complementary anatomical information (*i.e.*, categorical, topological, and shape), thereby enhancing the semantic discriminative power of the spine while maintaining structural consistency.

Class-level prompt. This prompt establishes the fundamental semantic baseline by providing explicit categorical descriptions T_{CA} (*e.g.*, "This CT image includes vertebral bodies, intervertebral discs, and nerves"). This prompt primarily addresses semantic confusion among spinal structures. We introduce a class-level prompt into the segmentation process. Specifically, textual descriptions of anatomical classes and imaging modality are encoded using a CLIP-based text encoder E_{CA} to generate category-aware semantic embeddings:

$$\mathbf{P}_{cls} = E_{CA}(T_{CA}). \quad (1)$$

Anatomical relational-level prompt. While class-level semantics provide coarse anatomical guidance, they are insufficient to capture the structured spatial organization of spinal anatomy. Therefore, we introduce anatomical relational prompt I_{rel} information by explicitly modeling anatomical dependencies (*e.g.*, "Vertebrae, discs, and nerves follow ordered anatomical relations along the spinal axis"). These relational descriptions are encoded using the Llama-encoder-1.0B¹ from LLM2Vec to obtain embeddings that capture spatial and sequential dependencies for consistent segmentation.

$$\mathbf{P}_{rel} = E_A(I_{rel}), \quad (2)$$

where E_A represents the anatomical relational level prompt encoder.

Shape-level prompt. While semantic and relational priors guide category recognition and structural organization, they are insufficient to constrain boundary localization. To enforce morphological consistency, we incorporate shape-level priors I_{atlas} derived from pre-registration multi-atlas spine data. A shape encoder E_S extracts prototypical anatomical representations, providing morphological references to refine segmentation:

$$\mathbf{P}_{shp} = E_S(I_{atlas}). \quad (3)$$

2.3 Dynamic Prompt Tuning Module

To effectively fuse the hierarchical multi-level prompt information, it is necessary to address the significant representational heterogeneity between the semantic, topological, and morphological information. Therefore, we propose a dynamic prompt tuning module to enhance the feature representation capability of prompt information. Specifically, let the extracted feature embeddings for categorical semantics, topological relations,

¹ <https://huggingface.co/knowledgator/Llama-encoder-1.0B>

and morphological shape constitute the prompt set $\mathbf{E} = \{\mathbf{P}_{cls}, \mathbf{P}_{rel}, \mathbf{P}_{shp}\}$. To achieve context-aware feature selection, we introduce a multi-level prompt gating that evaluates the relevance of each prompt feature based on the global context of the prompt feature $\mathbf{P}_{in}$. The gating dynamic importance weights are computed as:

$$\mathbf{P}_{in} = [\mathbf{P}_{cls}, \mathbf{P}_{rel}, \mathbf{P}_{shp}], \quad \alpha_{cls}, \alpha_{rel}, \alpha_{shp} = \text{Softmax}(\text{GAP}(\mathbf{P}_{in})), \quad (4)$$

where $\text{GAP}(\cdot)$ denotes global average pooling, $[\cdot]$ represents concat operation. Finally, the dynamic prompt representation $\mathbf{P}_{dynamic}$ is obtained via the weighted aggregation of the prompt feature:

$$\mathbf{P}_{dynamic} = \sum \alpha_k \cdot \mathbf{P}_k \quad k \in \{cls, rel, shp\}. \quad (5)$$

2.4 Uncertainty-guided Refinement Strategy

Although the proposed anatomy-aware multi-level prompt learning framework provides strong semantic and structural guidance, it may not effectively resolve the prediction uncertainty from ambiguous boundaries. To this end, we introduce an uncertainty-guided refinement strategy to explicitly emphasize anatomically ambiguous regions and improve boundary reliability.

Specifically, given the coarse segmentation prediction P_{pre_cro} , we estimate pixel-wise uncertainty by modeling the predictive distribution using a mean map μ and a variance map σ^2 , which are regressed by two lightweight parameterized networks F_{θ_μ} and $F_{\theta_{\sigma^2}}$. This formulation enables uncertainty-aware characterization of structurally ambiguous regions without introducing additional supervision:

$$\mu = F_{\theta_\mu}(P_{pre_cro}), \quad \sigma^2 = F_{\theta_{\sigma^2}}(P_{pre_cro}). \quad (6)$$

To capture both model uncertainty and annotation ambiguity, we employ two complementary strategies. First, we approximate the predictive distribution via T times Monte Carlo sampling and quantify model uncertainty as the average entropy of sampled predictions U^p . A variable ε_t is sampled from the standard Gaussian distribution $\mathcal{N}(0, I)$. Then, we compute a prediction discrepancy map U^{GT} relative to the ground truth Y_{GT} to highlight annotation ambiguity near boundaries:

$$\begin{aligned} p_{i,t} &= \text{Sigmoid}(\mu_i + \varepsilon_t \sigma_i), \quad \varepsilon_t \sim \mathcal{N}(0, I), \\ U^p &= -\frac{1}{T} \sum_{t=1}^T [p_{i,t} \log p_{i,t} + (1 - p_{i,t}) \log(1 - p_{i,t})], \\ U^{GT} &= Y_{GT} \odot (1 - \hat{\mu}) + (1 - Y_{GT}) \odot \hat{\mu}, \end{aligned} \quad (7)$$

where $\hat{\mu} = \text{Sigmoid}(\mu)$, $p_{i,t}$ represents the t -th sample probability score for the i -th pixel. $\odot$ denotes the Hadamard product. Then, these two sources of uncertainty are jointly utilized to construct an uncertainty-aware weighting map,

$$E_w = P_{pre_cro} \odot (\alpha \sigma^2 + \beta U^p), \quad (8)$$

where α and β are balancing coefficients. Finally, the uncertainty-weighted features are concatenated with the original segmentation features E and fed into a lightweight refinement head to generate the refined prediction:

$$P_{pre_ref} = \Phi(\text{Concat}(E, E_w)), \quad (9)$$

where $\Phi(\cdot)$ consists of two 1×1 convolution layers followed by softmax activation layers. Thus, the proposed strategy effectively improves boundary precision.

2.5 Loss Function

In this paper, we employ the segmentation loss L_{seg} and the uncertainty consistency loss L_{un} to optimize our segmentation task. Specifically, we employ the Dice loss L_{Dice} and the binary cross-entropy loss L_{BCE} as the segmentation loss. To encourage reliable uncertainty estimation, we further introduce an uncertainty consistency loss L_{un} , which is defined as,

$$L_{un} = \frac{1}{N} \sum_{i=1}^N \frac{1}{2\sigma_i^2} \|U_i^{GT} - U_i^p\|^2 + \frac{1}{2} \log \sigma_i^2, \quad (10)$$

where $\|\cdot\|$ denotes the L2 normalization, N is the total number of pixels in the predicted result. Finally, we train our framework with a total loss function $L_{Total} = L_{seg} + L_{un}$.

3 Experiments and Results

3.1 Datasets and Evaluation Metrics

We validate our method on two spine segmentation datasets: **1) Spine Dataset:** The in-house dataset includes 120 CT scans from the hospital and annotated via a two-stage expert verification process using ITK-SNAP [22]. The dataset is partitioned into 84 training, 12 validation, and 24 testing cases. Data will be made available on request. **2) SPIDER Dataset:** A public spine segmentation dataset ² containing 196 T1-weighted scans with heterogeneous resolutions, split into 137 training, 20 validation, and 39 testing cases. For both datasets, we target four semantic classes: vertebrae, intervertebral discs, nerves, and background. All slices were resized to 512×512 and strictly normalized. In addition, we employ the Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and 95% Hausdorff Distance (95HD) to evaluate the average segmentation performance of our proposed method across the three classes.

3.2 Implementation Details

In this experiment, we implemented our method using PyTorch and trained on four NVIDIA GeForce RTX 3090 GPUs. We employed the original SAM2 framework [16] for the image encoder (**ViT-B**) and decoder, with the encoder further fine-tuned via the

² <https://spider.grand-challenge.org/>

Table 1: The segmentation performance of our proposed method and other competing methods on the spine dataset and the SPIDER dataset. The symbol "*" indicates a significant improvement ($p < 0.05$) by our method (Mean $\pm$ Std).

Method	Spine Dataset			SPIDER Dataset		
	DSC ($\uparrow$)	IoU ($\uparrow$)	95HD ($\downarrow$)	DSC ($\uparrow$)	IoU ($\uparrow$)	95HD ($\downarrow$)
U-Net [17]	0.8739 $\pm$ 0.02*	0.7860 $\pm$ 0.04*	14.7401 $\pm$ 0.25*	0.8322 $\pm$ 0.03*	0.7145 $\pm$ 0.02*	13.464 $\pm$ 0.14*
nn-UNet [8]	0.8897 $\pm$ 0.03*	0.8026 $\pm$ 0.02*	13.3027 $\pm$ 0.31*	0.8525 $\pm$ 0.02*	0.7447 $\pm$ 0.01*	4.7458 $\pm$ 0.23*
TransUNet [2]	0.9105 $\pm$ 0.05*	0.8374 $\pm$ 0.04*	4.4353 $\pm$ 0.14*	0.8748 $\pm$ 0.01*	0.7799 $\pm$ 0.12*	1.6348 $\pm$ 0.11*
Swin-UNet [11]	0.8937 $\pm$ 0.02*	0.8133 $\pm$ 0.03*	6.0845 $\pm$ 0.42*	0.8644 $\pm$ 0.11*	0.7633 $\pm$ 0.10*	1.9300 $\pm$ 0.13*
U-Mamba [13]	0.8823 $\pm$ 0.04*	0.7746 $\pm$ 0.03*	14.2315 $\pm$ 0.14*	0.8602 $\pm$ 0.12*	0.7523 $\pm$ 0.11*	3.8214 $\pm$ 0.26*
VM-UNet [18]	0.8914 $\pm$ 0.20*	0.8032 $\pm$ 0.22*	7.3571 $\pm$ 0.24*	0.8711 $\pm$ 0.15*	0.7689 $\pm$ 0.13*	1.7125 $\pm$ 0.18*
SAM_Med2D [4]	0.9160 $\pm$ 0.04*	0.8416 $\pm$ 0.06*	3.5752 $\pm$ 0.22*	0.8789 $\pm$ 0.07*	0.7860 $\pm$ 0.10*	1.5620 $\pm$ 0.14*
SAM2 [16]	0.9024 $\pm$ 0.03*	0.8152 $\pm$ 0.02*	5.8289 $\pm$ 0.16*	0.8825 $\pm$ 0.03*	0.7902 $\pm$ 0.01*	1.4232 $\pm$ 0.21*
MA-SAM [6]	0.9126 $\pm$ 0.09*	0.8391 $\pm$ 0.10*	4.3203 $\pm$ 0.32*	0.8964 $\pm$ 0.04*	0.8133 $\pm$ 0.05*	1.3538 $\pm$ 0.24*
SAM3-Adapter [3]	0.9032 $\pm$ 0.13*	0.8182 $\pm$ 0.17*	5.5619 $\pm$ 0.16*	0.8913 $\pm$ 0.03*	0.8012 $\pm$ 0.05*	1.3725 $\pm$ 0.21*
AHPR-Net(Ours)	0.9242$\pm$0.02	0.8438$\pm$0.04	2.9832$\pm$0.12	0.9138$\pm$0.03	0.8353$\pm$0.02	1.1346$\pm$0.26

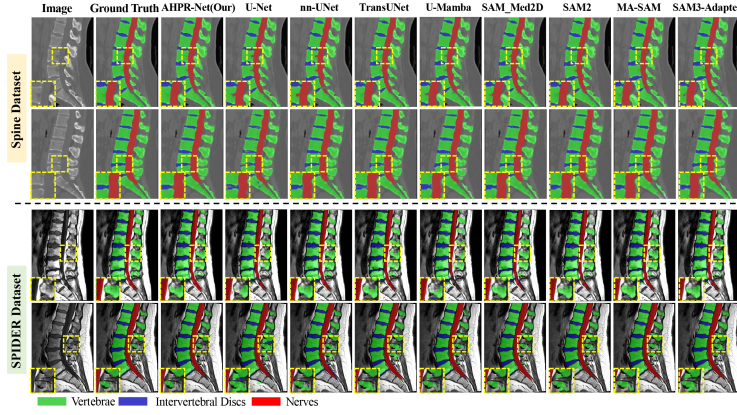


Fig. 2: Visual illustration of the segmentation results of our proposed AHPR-Net and different competing methods on the spine dataset and the SPIDER dataset.

Low-Rank Adaptation method (LoRA). The network was optimized using AdamW for 300 epochs, with an initial learning rate of $1e^{-3}$, weight decay of $1e^{-3}$, and momentum of 0.9. The batch size and atlas number were set to 16 and 20, respectively. For fair comparison, all comparison methods were trained under the same experimental settings. The bounding box prompts were used for SAM-based methods. The code of this work is available at Github.

3.3 Results

Comparison with State-of-the-Art Methods: We compare AHPR-Net with CNN-based (U-Net [17], nnU-Net [8]), Transformer-based (TransUNet [2], Swin-UNet [1]), Mamba-based (U-Mamba [13], VM-UNet [18]), and SAM-based methods (SAM_Med2D [4], SAM2 [16], MA-SAM [6], SAM3-Adapter [3]) on the spine dataset and the SPIDER dataset. As shown in Table 1, our proposed AHPR-Net method achieves the best segmentation performance on both datasets. Specifically, our method achieves 0.9242

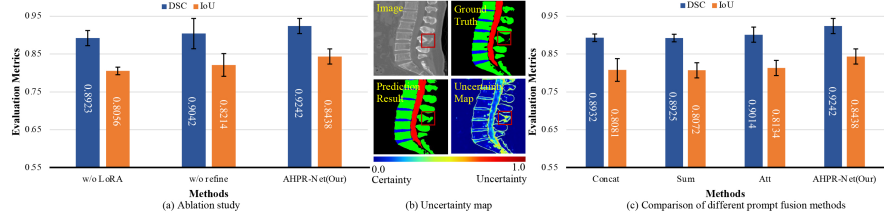


Fig. 3: Illustration of segmentation performance and uncertainty map on the spine dataset.

Table 2: Segmentation performance with multi-level prompt method on the spine dataset. 'CLS', 'REL', and 'SHA' represent the class-level, anatomical relational-level, and shape-level prompt information, respectively (Mean $\pm$ Std).

CLS	REL	SHA	DSC ($\uparrow$)	IoU ($\uparrow$)	95HD ($\downarrow$)
✓			0.8725 $\pm$ 0.02	0.7852 $\pm$ 0.01	14.8271 $\pm$ 0.21
	✓		0.8832 $\pm$ 0.02	0.7750 $\pm$ 0.04	14.2274 $\pm$ 0.13
		✓	0.8862 $\pm$ 0.01	0.7814 $\pm$ 0.04	14.2132 $\pm$ 0.14
✓	✓		0.8861 $\pm$ 0.12	0.7812 $\pm$ 0.10	14.2165 $\pm$ 0.11
✓	✓	✓	0.8924 $\pm$ 0.05	0.8060 $\pm$ 0.08	7.1420 $\pm$ 0.21
	✓	✓	0.9016 $\pm$ 0.02	0.8143 $\pm$ 0.05	5.9264 $\pm$ 0.18
✓	✓	✓	0.9162 $\pm$ 0.02	0.8419 $\pm$ 0.04	3.5741 $\pm$ 0.15
✓	✓	✓	0.9242$\pm$0.02	0.8438$\pm$0.04	2.9832$\pm$0.12

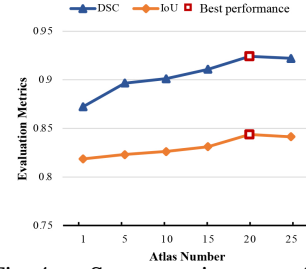


Fig. 4: Segmentation performance using different atlas numbers on the spine dataset.

and 0.9138 on the spine dataset and the SPIDER dataset, respectively, in terms of DSC. In addition, our method always obtains the best segmentation results from Fig. 2. These results demonstrate that by integrating class-level, anatomical relational-level, and shape-level prompts within a dynamic prompt tuning module, our method provides richer and more anatomically grounded guidance to the segmentation network. In addition, the uncertainty-guided refinement stage enables the segmentation network to pay more attention to and preserve fine structural details. Therefore, our method can further enhance segmentation performance than other methods.

Ablation Study: We performed a variety of ablation studies to thoroughly evaluate the proposed AHPR-Net and validate the performance under different settings. First, we investigate the effectiveness of LoRA-based fine-tuning and the uncertainty-guided refinement stage. As illustrated in Fig. 3 (a) and Fig. 3 (b), both components consistently improve segmentation performance, validating the practicality of prompt fine-tuning and uncertainty-aware refinement for spine segmentation. We further examine different prompt fusion strategies, including concatenation, summation, and attention-based fusion. The results in Fig. 3 (c) demonstrate that more effective prompt integration leads to improved segmentation accuracy, thereby highlighting the advantage of the proposed dynamic prompt tuning module over conventional static fusion schemes.

To assess the contribution of hierarchical prompts, we evaluate different prompts and atlas numbers while keeping all other experimental settings unchanged. As shown in Table 2 and Fig. 4, our AHPR-Net achieves the best segmentation performance by

combining multi-level prompts and using 20 atlases. These findings indicate that hierarchical multi-level prompting effectively captures complementary category semantics, anatomical relationships, and structural morphology, ultimately leading to more accurate spine segmentation.

4 Conclusion

In this paper, we propose a novel anatomy-aware hierarchical prompting refinement network for spine image segmentation, called AHPR-Net. Specifically, we propose a hierarchical multi-level prompting strategy to provide comprehensive guidance. Then, we propose a dynamic prompt tuning module to improve the perception ability of the complex anatomical structures. Finally, we propose an uncertainty-guided refinement strategy to improve boundary precision. Experimental results on an in-house dataset and a public dataset demonstrate that our proposed method achieves superior segmentation performance compared to several state-of-the-art methods.

Acknowledgments. This work was supported in part by the National Natural Science Foundation of China under Grants 62136004, 62276130, 62376123, 62402219, in part by the National Key R&D Program of China (Grant No. 2023YFF1204803), in part by Guangdong Basic and Applied Basic Research Foundation under Grant 2024A1515011925, in part by Shenzhen Longhua Science and Technology Innovation Special Funding Project (Industrial Sci-Tech Innovation Center of Low-Altitude Intelligent Networking), in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region, China (Project No. T45-401/22-N).

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

References

1. 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: European Conference on Computer Vision. pp. 205–218. Springer (2022)
2. 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)
3. Chen, T., Cao, R., Yu, X., Zhu, L., Ding, C., Ji, D., Chen, C., Zhu, Q., Xu, C., Mao, P., et al.: SAM3-Adapter: Efficient Adaptation of Segment Anything 3 for Camouflage Object Segmentation, Shadow Detection, and Medical Image Segmentation. arXiv preprint arXiv:2511.19425 (2025)
4. Cheng, J., Ye, J., Deng, Z., Chen, J., Li, T., Wang, H., Su, Y., Huang, Z., Chen, J., Jiang, L., et al.: SAM-Med2D. arXiv preprint arXiv:2308.16184 (2023)
5. Deng, Y., Gu, F., Zeng, D., Lu, J., Liu, H., Hou, Y., Zhang, Q.: An effective U-Net and BiSeNet complementary network for spine segmentation. *Biomedical Signal Processing and Control* **89**, 105682 (2024)
6. Fan, D., Zhao, J., Li, C., Wang, X., Zhang, R., Zhu, Q., Wang, M., Si, H., Zhang, D., Sun, L.: MA-SAM: A multi-atlas guided sam using pseudo mask prompts without manual annotation for spine image segmentation. *IEEE Transactions on Medical Imaging* **44**(5), 2157–2169 (2025)

7. Hallinan, J.T.P.D., Zhu, L., Yang, K., Makmur, A., Algazwi, D.A.R., Thian, Y.L., Lau, S., Choo, Y.S., Eide, S.E., Yap, Q.V., et al.: Deep learning model for automated detection and classification of central canal, lateral recess, and neural foraminal stenosis at lumbar spine MRI. *Radiology* **300**(1), 130–138 (2021)
8. Isensee, F., Jaeger, P.F., Kohl, S.A., Petersen, J., Maier-Hein, K.H.: nnU-Net: A self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods* **18**(2), 203–211 (2021)
9. Jiang, C., Ding, T., Song, C., Tu, J., Yan, Z., Shao, Y., Wang, Z., Shang, Y., Han, T., Tian, Y.: Medical SAM3: A Foundation Model for Universal Prompt-Driven Medical Image Segmentation. *arXiv preprint arXiv:2601.10880* (2026)
10. Kowlagi, N., Kemppainen, A., McSweeney, T., Saarakkala, S., Noailly, J., Williams, F.M., Cheung, J.P.Y., Karppinen, J., Tiulpin, A.: Open lumbar spine image analysis: A 3d slicer extension for segmentation, grading, and intervertebral disc height index with multi-data set validation. *Spine* **51**(2), E35–E44 (2026)
11. Li, J., Li, Y., Wu, J., Ji, C., Chen, Z., Lei, W., Wang, R.: Semi-MedSAM: Adapting SAM-assisted Semi-supervised Multi-modality Learning for Medical Endoscopic Image Segmentation. *Pattern Recognition* p. 113206 (2026)
12. Liu, J., Cui, Z., Desrosiers, C., Lu, S., Zhou, Y.: Grayscale self-adjusting network with weak feature enhancement for 3D lumbar anatomy segmentation. *Medical Image Analysis* **81**, 102567 (2022)
13. Ma, J., Li, F., Wang, B.: U-Mamba: Enhancing Long-range Dependency for Biomedical Image Segmentation. *arXiv preprint arXiv:2401.04722* (2024)
14. Möller, H., Graf, R., Schmitt, J., Keinert, B., Schön, H., Atad, M., Sekuboyina, A., Streckenbach, F., Kofler, F., Kroencke, T., et al.: SPINEPS—automatic whole spine segmentation of T2-weighted MR images using a two-phase approach to multi-class semantic and instance segmentation. *European Radiology* **35**(3), 1178–1189 (2025)
15. Qian, L., Chen, J., Ma, L., Urakov, T., Gu, W., Liang, L.: Attention-based Shape-Deformation networks for Artifact-Free geometry reconstruction of lumbar spine from MR images. *IEEE Transactions on Medical Imaging* **44**(12), 5258–5277 (2025)
16. Ravi, N., Gabeur, V., Hu, Y.T., Hu, R., Ryali, C., Ma, T., Khedr, H., Rädle, R., Rolland, C., Gustafson, L., et al.: SAM 2: Segment Anything in Images and Videos. *arXiv preprint arXiv:2408.00714* (2024)
17. Ronneberger, O., Fischer, P., Brox, T.: U-Net: Convolutional networks for biomedical image segmentation. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 234–241. Springer (2015)
18. Ruan, J., Li, J., Xiang, S.: VM-UNet: Vision Mamba UNet for Medical Image Segmentation. *ACM Transactions on Multimedia Computing, Communications and Applications* (2024)
19. Simons, S.J., Papież, B.W.: SpineFM: Leveraging foundation models for automatic spine X-ray segmentation. In: *2025 IEEE 22nd International Symposium on Biomedical Imaging (ISBI)*. pp. 1–4. IEEE (2025)
20. Touvron, H., Lavril, T., Izacard, G., Martinet, X., Lachaux, M.A., Lacroix, T., Rozière, B., Goyal, N., Hambro, E., Azhar, F., et al.: Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023)
21. Xu, J., He, K., Tang, Y., Li, W., Lan, M., Dong, X., Ke, Y., Feng, M.: BrainPrompt: Multi-level Brain Prompt Enhancement for Neurological Condition Identification. In: *International Conference on Medical Image Computing and Computer-Assisted Intervention*. pp. 172–182. Springer (2025)
22. Yushkevich, P.A., Piven, J., Cody Hazlett, H., Gimpel Smith, R., Ho, S., Gee, J.C., Gerig, G.: User-Guided 3D Active Contour Segmentation of Anatomical Structures: Significantly Improved Efficiency and Reliability. *Neuroimage* **31**(3), 1116–1128 (2006)

23. Zhao, J., Sun, L., Fan, D., Wang, K., Si, H., Fu, H., Zhang, D.: Uncertainty-driven edge prompt generation network for medical image segmentation. *IEEE Transactions on Medical Imaging* **44**(10), 3950–3961 (2025)
24. Zhou, Z., Wang, P., Ji, X., Xu, W., Han, Z., Wei, B.: From contrast-driven segmentation to central lumbar spinal stenosis grading: a comprehensive multi-view spinal MRI image analysis. *IEEE Transactions on Medical Imaging* (2026)