PANet: Probability-Anchored Network for Ultrasound Bone Segmentation

Wenqing Yan[1], Haowei Li[1], Hui Ding[1], Zhe Zhao[2], and Guangzhi Wang[1]*

[1] Tsinghua University, Beijing 100084, China
{ywq22,lhw22}@mails.tsinghua.edu.cn
{dinghui,wgz-dea}@tsinghua.edu.cn
[2] Orthopaedics and Sports Medicine Center, Beijing Tsinghua Changgung Hospital, School of Clinical Medicine, Tsinghua University, Beijing 100084, China
{zhaozhao_02}@163.com

Abstract. Ultrasound bone surface segmentation is a critical prerequisite for orthopedic applications such as quantitative analysis and intraoperative navigation. However, reliable segmentation remains challenging due to ambiguous bone boundaries, strong appearance variability, and interference from bone-like soft tissues, often resulting in unstable predictions in complex clinical scenarios. Existing approaches often rely on network designs tailored to specific ultrasound physical priors, which may limit their adaptability across different imaging conditions. In this work, we propose a Probability-Anchored Network (PANet) for robust ultrasound bone segmentation. Built upon a U-Net backbone, PANet explicitly introduces a probability space as an intermediate representation between image features and segmentation results, enabling controllable activation, discrimination, and refinement of bone boundaries. Specifically, global linear bias, local probability exchange, and global attention mechanisms are jointly employed to enhance bone-related responses while suppressing globally inconsistent false positives. Moreover, a unified segmentation head provides globally consistent supervision across all network stages, enabling probability-aligned learning and segmentation-guided attention. Experiments on freehand-scanned long-bone ultrasound datasets demonstrate that PANet achieves more accurate and stable segmentation performance compared with state-of-the-art methods.

Keywords: Ultrasound bone segmentation · Probabilistic feature modeling · Unified multi-level supervision

1 Introduction

With the increasing availability of ultrasound imaging systems, their application in orthopedic settings has attracted growing attention [4]. Although X-ray, CT, and MRI remain the standard modalities for acquiring clear bone structural

* Corresponding author.

information, ultrasound offers real-time, radiation-free, and portable imaging for continuous bedside and intraoperative assessment [24]. Consequently, ultrasound has been increasingly applied in fracture reduction [29], interventional procedures [1], and joint and tendon assessment [22], all of which require accurate bone surface localization.

Due to the high acoustic impedance of bone tissue, freehand ultrasound imaging often exhibits complex appearances and severe artifacts, making bone interpretation highly dependent on physician experience. This dependence restricts the broader adoption of ultrasound in quantitative analysis, preoperative planning, and intraoperative navigation [9, 28, 20]. Therefore, stable and consistent automated bone surface segmentation has become a critical prerequisite. However, ambiguous bone boundaries and strong interference from surrounding bone-like structures make ultrasound bone segmentation particularly challenging under clinical imaging conditions.

Traditional ultrasound bone segmentation methods rely on manually designed rules based on image features such as intensity, shadow, and local phase [6, 18, 11]. However, they are highly sensitive to imaging conditions and parameter settings, limiting their robustness in complex clinical scenarios. Recent deep learning–based approaches show notable advantages. Existing studies mainly enhance performance by incorporating ultrasound-specific physical priors or designing specialized architectures to strengthen feature representation. Prior-based methods introduce knowledge-driven image enhancement techniques, such as shadow augmentation and local phase filtering, to emphasize bone-related characteristics [26, 27, 2]. Some studies further adopt dual-decoder architectures to jointly segment bone surfaces and shadows, improving the separation of bone structures from bone-like soft tissues [13]. However, such prior-driven designs are sensitive to system-dependent variability and often generalize poorly across devices and imaging settings. Other works design CNN-, GCN-, or attention-based architectures without explicit physical priors [16, 12, 19, 14], yet the lack of explicit prediction confidence regulation makes them vulnerable to false-positive activations from bone-like soft tissues, limiting reliability in clinical navigation.

In this study, we propose a Probability-Anchored Network (PANet) for ultrasound bone segmentation. PANet introduces a probability space as an intermediate representation between feature learning and segmentation, modeling the binary confidence of bone structures relative to surrounding soft tissues as a probability distribution. CNN features regulate spatial exchange and linear bias within the probability space, while a unified segmentation head enables probability estimation across stages. Encoder attention is guided by probability-aware activation to reinforce high-confidence bone regions and suppress false-positive responses. The decoder performs progressive boundary refinement via local probability-space exchanges, preventing false-positive feature propagation introduced by skip connections. Our contributions are: (1) A probability-anchored architecture introducing an explicit K-dimensional probability space co-updated with CNN features, decomposing segmentation into activation, discrimination, and refinement stages to reduce false-positive predictions; (2) A unified multi-

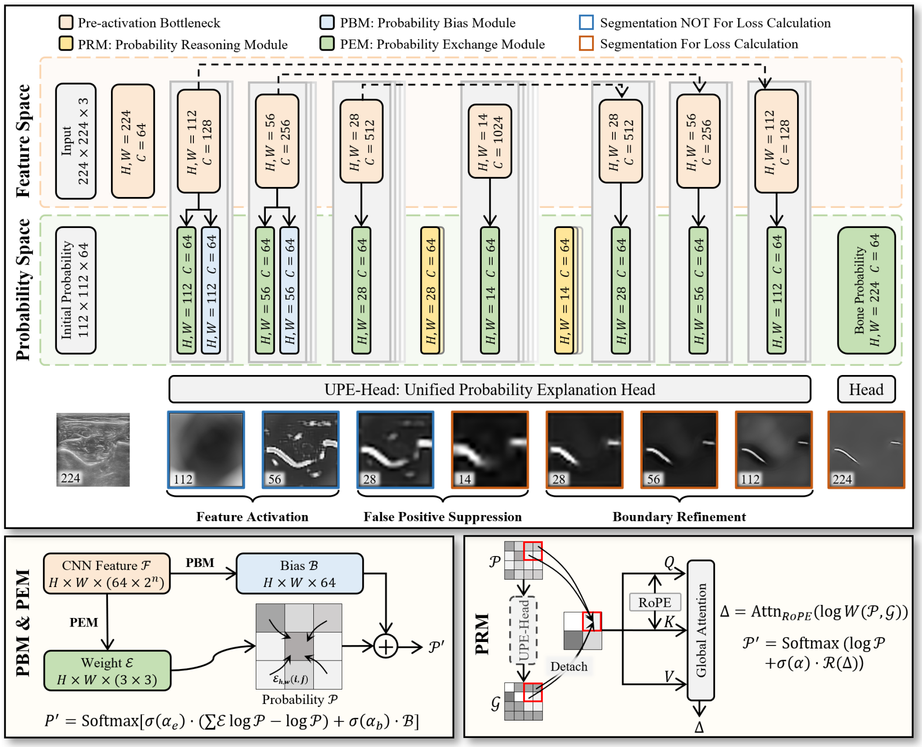

Fig. 1. Overview of the proposed PANet. A probability space is introduced between image features and segmentation outputs. PBM, PEM, and PRM are designed to activate, optimize, and discriminate bone segmentation probabilities within this space, guided by CNN-extracted image features. The UPE-Head decodes multi-scale probability maps into segmentation outputs. The evolution of bone segmentation probabilities across different network stages is visualized.

level supervision strategy enforcing consistent probability alignment throughout the network, enabling more controllable and focused encoder feature evolution.

2 Method

2.1 Probability-Anchored Network Structure

An overview of the proposed PANet architecture is shown in Fig. 1. Ultrasound bone segmentation remains challenging due to weak and ambiguous bone boundaries and the difficulty of distinguishing bone surfaces from visually similar surrounding structures. To address these challenges, PANet introduces an explicit probability space $\mathcal{P} \in \mathbb{R}^{H\times W\times K}$ as an intermediate representation between U-Net [21] style backbone ($\mathcal{F} \in \mathbb{R}^{H\times W\times C}$) and segmentation logits $\mathcal{G} \in \mathbb{R}^{H\times W}$. In this probability space, foreground and background evidence is progressively organized into separable probability clusters as the network evolves. Meanwhile, since probability maps at different stages are supervised under a shared spatially aligned target, the semantic anchors of these clusters remain consistent throughout the network. This design enables structured decomposition into (i) early activation, (ii) attention-based discrimination, and (iii) progressive refinement.

Backbone and co-update in probability space. Pre-activation bottleneck blocks [8] are adopted throughout the CNN backbone. The encoder reduces the spatial resolution from 224×224 to 14×14, followed by a decoder upsampling features to 112×112. After each convolutional update, features are projected into the probability space via two lightweight branches that generate a local exchange kernel and a global bias term (Section 2.2), which jointly drive the probability update. This feature-probability co-update is repeated 2, 3, 4, and 5 times across encoding stages and twice per decoding stage. The probability space is initialized with Gaussian anchor priors as:

$$\mathcal{P}_k^{(0)}(x, y) = \mathrm{Softmax}_k\Big(-\left\|(x, y) - \boldsymbol{\mu}_k\right\|^2 / 2\sigma^2\Big), k = 1, \ldots, K \tag{1}$$

where (x, y) denotes the spatial coordinate, and $\boldsymbol{\mu}_k$ represents the center of the Gaussian anchor in the k-th channel. The probability space is uniformly partitioned into $\sqrt{K} \times \sqrt{K}$ grid regions, with $\boldsymbol{\mu}_k$ sequentially assigned to the center of each grid cell. This Gaussian-anchor initialization provides smooth spatial gradients and stabilizes the early-stage training of the probability space.

Late-stage global reasoning and supervision. At late encoding stages (28×28 and 14×14 resolutions), PANet applies global self-attention [25] in the probability space with rotary positional embedding (RoPE) [23] to enhance discrimination under complex backgrounds (Section 2.3). A convolutional Unified Probability Explanation Head (UPE-Head) with two 3×3 and one 1×1 convolutions produces intermediate logits at resolutions ranging from 14×14 to 112×112, while an additional head (one 3×3 and one 1×1 convolution) outputs the final full-resolution prediction. All outputs are jointly supervised by a multi-level loss:

$$\mathcal{L} = \frac{1}{n} \sum_{i=1}^{n} \left(\mathrm{BCE}(pred_i, gt_i) + \mathrm{Dice}(pred_i, gt_i)\right) \tag{2}$$

where n denotes the number of supervised resolutions, and $(pred_i, gt_i)$ denote the prediction and ground truth at the i-th resolution. This unified head enforces consistent probability-space alignment across stages and provides encoder-level segmentation cues to guide subsequent attention reasoning.

2.2 Probability Exchange and Bias Modules.

The Probability Exchange Module (PEM) and Probability Bias Module (PBM) transform CNN features into update signals for the probability representation

$\mathcal{P} \in \mathbb{R}^{H\times W\times K}$. Given a feature map $\mathcal{F}$, a convolution followed by a softmax operation generates a spatially varying 3×3 local exchange kernel $\mathcal{E} \in \mathbb{R}^{H\times W\times 9}$. Guided by $\mathcal{E}$, PEM redistributes probabilities via weighted aggregation of neighboring log-probabilities within the 3×3 neighborhood. In parallel, PBM projects features to a global bias term $\mathcal{B} \in \mathbb{R}^{H\times W\times K}$ in the same probability space. Their combined effect yields the probability update:

$$\mathcal{P}'_{h,w} = \mathrm{Softmax}\left[\sigma(\alpha_e)\left(\sum_{(i,j)\in\Omega} \mathcal{E}_{h,w}(i,j)\cdot\hat{\mathcal{P}}_{h+i,w+j} - \hat{\mathcal{P}}_{h,w}\right) + \sigma(\alpha_b)\mathcal{B}_{h,w}\right] \tag{3}$$

where $\mathcal{P}_{a,b} \in \mathbb{R}^K$ denotes the probability vector at (a,b), $\hat{\mathcal{P}} = \log\mathcal{P}$, and Ω denotes the 3×3 neighborhood centered at (h,w). The sigmoid function $\sigma(\cdot)$ constrains the contributions of both PEM and PBM, while α_e and α_b are learnable scaling parameters initialized to -5 to encourage conservative early updates.

PEM operates throughout the network for continuous local refinement, while PBM is applied only at 112×112 and 56×56 resolutions to promote early bone activation, leaving hard global discrimination to the subsequent attention reasoning module.

2.3 Probability Reasoning Module

To better distinguish true bone structures from visually similar soft tissues, the Probability Reasoning Module (PRM) is introduced at late encoding stages where sufficient semantic context is available. PRM performs global reasoning directly in the probability space, complementing local exchange by suppressing globally inconsistent activations.

Given the probability representation $\mathcal{P} \in \mathbb{R}^{H\times W\times K}$ and the confidence map $\mathcal{G} \in \mathbb{R}^{H\times W}$ produced by the UPE-Head (after sigmoid), PRM aggregates probabilities within $b\times b$ blocks to form compact tokens. For each block $\Omega_{p,q}$, the block-level probability $\mathcal{P}^{\mathrm{blk}}$ is computed by confidence-weighted aggregation:

$$\mathcal{P}^{\mathrm{blk}}_{k,p,q} = \sum_{(i,j)\in\Omega_{p,q}} w^{p,q}_{i,j}\,\mathcal{P}_{k,i,j}, \qquad w^{p,q}_{i,j} = \frac{\mathcal{G}_{i,j}}{\sum_{(u,v)\in\Omega_{p,q}} \mathcal{G}_{u,v} + \epsilon} \tag{4}$$

Block-level probabilities are embedded via log mapping and linear projection as $\mathbf{t}_n = \phi\big(\log(\mathcal{P}^{\mathrm{blk}}_n + \epsilon)\big),\ n = 1,\ldots,H_bW_b$. Here, $\phi(\cdot)$ consists of layer normalization and linear projection. Global token reasoning is performed using multi-head self-attention with RoPE applied to $\mathbf{Q}$ and $\mathbf{K}$:

$$\mathbf{Y} = \mathrm{Softmax}\Big(\hat{\mathbf{Q}}\hat{\mathbf{K}}^\top/\sqrt{d_h}\Big)\mathbf{V}, \quad \hat{\mathbf{Q}} = \mathrm{RoPE}(\mathbf{Q}), \quad \hat{\mathbf{K}} = \mathrm{RoPE}(\mathbf{K}) \tag{5}$$

where $\mathbf{Q},\mathbf{K},\mathbf{V}$ are linear projections of $\mathbf{t}$ and d_h is the per-head dimension.

The attention output is projected to a block-level residual update Δ^{blk} and broadcast to full resolution via $\mathcal{R}(\cdot)$. To preserve a valid probability distribution, PRM performs a bounded residual update in log-probability space:

$$\mathcal{P}' = \mathrm{Softmax}\big[\log\mathcal{P} + \sigma(\alpha)\cdot\mathcal{R}\big(\Delta^{\mathrm{blk}}\big)\big] \tag{6}$$

where $\sigma(\alpha)$ controls the update magnitude at early stage.

3 Experiments and Results

3.1 Dataset and Implementations

The dataset was collected from 10 healthy volunteers (5 males and 5 females, aged 24–42 years) using a Mindray Resona 6 system equipped with a linear probe (L14-5WU) at a depth of 5 cm. Freehand upper-limb long-bone scans were acquired for all subjects, with additional lower-limb scans obtained from two of the subjects. The study was approved by the Beijing Tsinghua Changgung Hospital Medical Ethics Committee (Approval No.: 24204-4-02), and written informed consent was obtained from all participants. Annotations were generated by an experienced researcher using a SAM2-based semi-automatic tool integrated into 3D Slicer and were reviewed by an orthopedic surgeon with slice-by-slice confirmation. The dataset contains 246,471 images (one or two long bones per image) and was split at limb level (48 limbs) to prevent leakage.

We compared PANet with U-Net [21], FCN [15], ResNet with a task-specific head [7], SegNet [3], Att-Unet [17], ResUNet [30], ResUNet++ [10], TransUNet [5], and BCA-Net [13], covering encoder–decoder, residual, attention-based, and task-specific architectures. Due to limited publicly available implementations and datasets for ultrasound bone segmentation, direct comparison is constrained. Evaluation metrics included IoU, Dice, HD95, recall, precision, and inference speed (fps). Five-fold cross-validation was used, with metrics computed per image and averaged across folds.

Inputs were resized to 224×224 with rotation, flipping, and grayscale augmentation. All models were trained on an NVIDIA RTX 4090 GPU using SGD with momentum 0.9 and weight decay 10^{-4}. The initial learning rate was set to 0.0167 and linearly decayed over 50 epochs with a batch size of 32. Baselines used publicly available implementations under identical data splits and training settings, while BCA-Net followed the original configuration. PANet used encoder channels of 64, 128, 256, 512, and 1024 with probability dimension $K = 64$. In PRM, block size was set to $b = 2$ at 28×28 and 14×14 resolutions, and self-attention employed 4 heads with a per-head dimension of 16. In total, PANet contains 14.8M trainable parameters. The code and trained models will be publicly released, and the dataset will be made available upon reasonable request.

3.2 Performance Comparison

As shown in Table 1, PANet achieves the best overall performance, attaining an IoU of 81.39%, a Dice score of 89.43%, and an HD95 of 0.97mm, with balanced recall (90.20%) and precision (90.00%). In contrast, competing methods exhibit noticeable recall–precision imbalance. BCA-Net achieves the highest recall (91.10%) but with reduced precision (85.67%, corresponding to a 5.43-point gap), while ResUNet shows the opposite trend (83.43% vs 90.68%). PANet reduces this gap to 0.20 points, achieving a balanced trade-off between sensitivity and specificity. Fig. 2 further shows that PANet produces fewer false positives and missed bone surfaces under severe interference from multiple bone-like soft-tissue structures. Despite the additional feature–probability co-updates, PANet achieves 194 fps on an RTX 4090 using `torch.compile`, exceeding the 60 fps clinical real-time requirement.

The current dataset was collected from a single institution with one ultrasound system and mainly includes healthy long-bone images. Future work will extend the evaluation to multi-center and cross-device datasets, clinically challenging cases, and diverse bone structures. Additional boundary metrics, stronger comparison baselines, and quantitative interpretability analyses will also be in-

Table 1. Quantitative comparison of different methods.

Method	IoU (%) ↑	Dice (%) ↑	HD95 (mm) ↓	Recall (%) ↑	Precision (%) ↑	FPS ↑
U-Net	77.85 ± 0.29	86.99 ± 0.23	1.47 ± 0.15	87.23 ± 2.24	88.90 ± 2.07	444
FCN	69.33 ± 0.85	81.45 ± 0.60	1.59 ± 0.41	85.65 ± 1.27	79.43 ± 2.00	408
ResNet	70.99 ± 0.67	82.66 ± 0.45	1.28 ± 0.08	86.51 ± 1.58	80.80 ± 2.19	321
SegNet	77.22 ± 0.54	86.69 ± 0.34	1.37 ± 0.12	86.75 ± 1.80	88.63 ± 1.98	**510**
Att-Unet	78.57 ± 0.59	87.47 ± 0.37	1.41 ± 0.16	87.50 ± 1.74	89.47 ± 1.98	375
ResUNet	75.99 ± 0.65	85.69 ± 0.50	1.70 ± 0.12	83.43 ± 1.84	**90.68 ± 1.63**	369
ResUNet++	77.24 ± 0.94	86.61 ± 0.62	1.53 ± 0.17	86.77 ± 2.93	88.70 ± 2.57	330
TransUNet	79.32 ± 0.68	88.08 ± 0.44	1.35 ± 0.20	89.73 ± 1.82	88.21 ± 2.02	202
BCA-Net	78.44 ± 1.08	87.54 ± 0.70	1.23 ± 0.13	**91.10 ± 1.39**	85.67 ± 1.71	111
PANet	**81.39 ± 0.92**	**89.43 ± 0.58**	**0.97 ± 0.10**	90.20 ± 1.04	90.00 ± 1.14	194

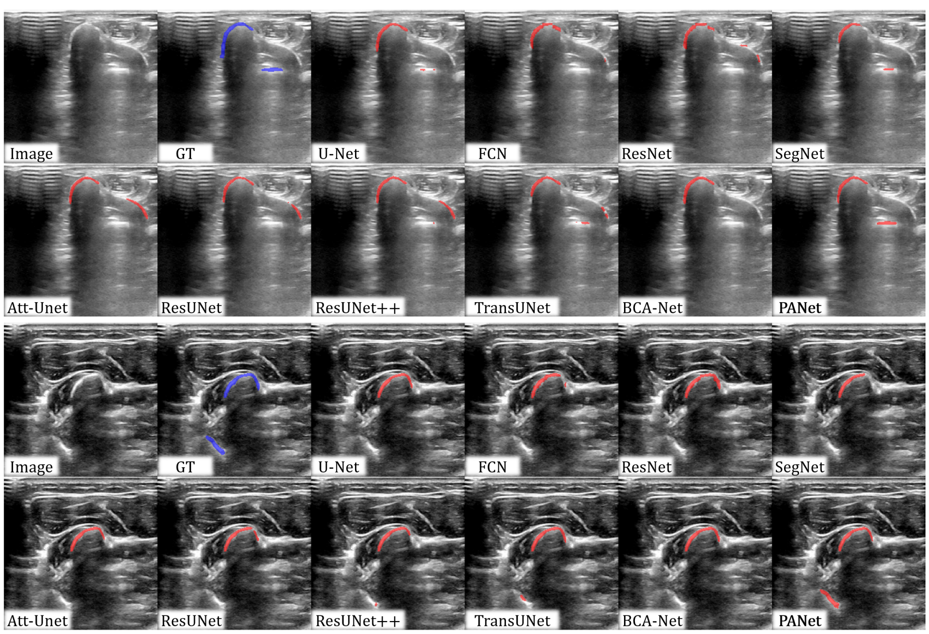

Fig. 2. Qualitative comparison on two challenging cases. Ground-truth bone surfaces are shown in blue, while predicted results are shown in red.

corporated to further assess the robustness, applicability, and overall performance of PANet.

3.3 Ablation Study

To evaluate the contribution of key components, ablation experiments were conducted by removing individual modules, with results summarized in Table 2.

Removing the probability space (w/o $\mathcal{P}$) reduces the model to a backbone with one single convolutional supervision, resulting in large IoU drop (-1.88 points) and a 47% increase in HD95. This confirms that the probability representation provides modeling capacity beyond deep supervision alone. Removing PEM leads to the most severe degradation, falling below the U-Net baseline. As PEM operates at every co-update step in both encoder and decoder, its removal eliminates local probability redistribution, disabling boundary-aware optimization and leading to poor boundary alignment. Removing PRM causes a moderate performance drop, with recall decreasing from 90.20% to 89.13% while precision remains comparable, indicating that PRM improves global discrimination, facilitating bone activation in ambiguous regions.

Removing PBM yields performance within one standard deviation of the full model, suggesting that PEM-driven local exchange partially compensates for the absence of feature-to-probability bias. Building upon w/o PBM, further removing Gaussian initialization (w/o PBM+Uni. $P_k(0)$) causes an additional degradation of 0.47 IoU points and 0.30 Dice points, leading to unstable convergence. Removing PBM delays structure activation to later encoder stages, wasting early capacity, while the twisting-based injection is less effective than linear bias for complex structure perception. Although removing the UPE-Head produces comparable results, it is retained to provide intermediate supervision for PRM attention, enable probability-space monitoring across stages, and improve robustness under challenging imaging conditions. The influence of hyperparameters, including the probability dimension K, PRM block size, and loss weights, will be further investigated in future work.

Table 2. Results of ablation study. “w/o” denotes removing the corresponding module.

Method	IoU (%) ↑	Dice (%) ↑	HD95 (pixel) ↓	Recall (%) ↑	Precision (%) ↑
w/o $\mathcal{P}$	79.51 ± 0.43	88.22 ± 0.26	1.43 ± 0.22	88.32 ± 1.39	89.78 ± 1.52
w/o PEM	75.07 ± 1.26	85.41 ± 0.85	1.23 ± 0.09	85.90 ± 1.10	86.40 ± 1.21
w/o PRM	80.30 ± 1.03	88.75 ± 0.67	1.02 ± 0.08	89.13 ± 1.40	89.79 ± 1.31
w/o PBM+Uni. $\mathcal{P}_k(0)$	81.06 ± 0.99	89.22 ± 0.64	0.98 ± 0.09	89.55 ± 0.80	90.26 ± 0.89
w/o PBM	81.53 ± 1.07	89.52 ± 0.69	0.95 ± 0.11	90.32 ± 1.25	90.04 ± 1.15
w/o UPE	81.47 ± 1.05	89.47 ± 0.67	0.98 ± 0.08	90.47 ± 1.27	89.84 ± 1.25
PANet	81.39 ± 0.92	89.43 ± 0.58	0.97 ± 0.10	90.20 ± 1.04	90.00 ± 1.14

4 Conclusion

We propose a Probability-Anchored Network (PANet) for robust and stable ultrasound bone segmentation. By modeling an intermediate probability space with unified multi-level supervision, segmentation is decomposed into activation, discrimination, and refinement stages. Experiments on a large-scale freehand long-bone ultrasound dataset demonstrate consistent improvements in segmentation accuracy, boundary precision, and reliability under challenging imaging conditions. Moreover, the introduced probability space provides a quantifiable intermediate representation that enables stage-wise visualization of the network's decision process, offering a potential foundation for future research on controllable network design and clinical interpretability.

Acknowledgments. This work was supported by the NSFC (U22A20355), National Key R&D Program of China (2022YFC2405304), Beijing Natural Science Foundation (7242059, L232055) and Medivis. Inc.

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

References

1. Ahmed, A.S., Ramakrishnan, R., Ramachandran, V., Ramachandran, S.S., Phan, K., Antonsen, E.L.: Ultrasound diagnosis and therapeutic intervention in the spine. Journal of spine surgery **4**(2), 423 (2018)
2. Alsinan, A.Z., Patel, V.M., Hacihaliloglu, I.: Automatic segmentation of bone surfaces from ultrasound using a filter-layer-guided cnn. International journal of computer assisted radiology and surgery **14**(5), 775–783 (2019)
3. Badrinarayanan, V., Kendall, A., Cipolla, R.: Segnet: A deep convolutional encoder-decoder architecture for image segmentation. IEEE transactions on pattern analysis and machine intelligence **39**(12), 2481–2495 (2017)
4. Blankstein, A.: Ultrasound in the diagnosis of clinical orthopedics: The orthopedic stethoscope. World journal of orthopedics **2**(2), 13 (2011)
5. 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)
6. Foroughi, P., Boctor, E., Swartz, M.J., Taylor, R.H., Fichtinger, G.: P6d-2 ultrasound bone segmentation using dynamic programming. In: 2007 IEEE Ultrasonics Symposium Proceedings. pp. 2523–2526. IEEE (2007)
7. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
8. He, K., Zhang, X., Ren, S., Sun, J.: Identity mappings in deep residual networks. In: European conference on computer vision. pp. 630–645. Springer (2016)
9. Hohlmann, B., Glanz, J., Radermacher, K.: Segmentation of the distal femur in ultrasound images. In: Current Directions in Biomedical Engineering. vol. 6, p. 20200034. De Gruyter (2020)

10. Jha, D., Smedsrud, P.H., Riegler, M.A., Johansen, D., De Lange, T., Halvorsen, P., Johansen, H.D.: Resunet++: An advanced architecture for medical image segmentation. In: 2019 IEEE international symposium on multimedia (ISM). pp. 225–2255. IEEE (2019)
11. Jia, R., Mellon, S.J., Hansjee, S., Monk, A.P., Murray, D.W., Noble, J.A.: Automatic bone segmentation in ultrasound images using local phase features and dynamic programming. In: 2016 IEEE 13th international symposium on biomedical imaging (ISBI). pp. 1005–1008. IEEE (2016)
12. Li, P., Zhao, H., Liu, P., Cao, F.: Automated measurement network for accurate segmentation and parameter modification in fetal head ultrasound images. Medical & Biological Engineering & Computing **58**(11), 2879–2892 (2020)
13. Liu, C., Wang, W., Sun, R., Wang, T., Shen, X., Sun, T.: A dual-decoder banded convolutional attention network for bone segmentation in ultrasound images. Medical Physics **52**(3), 1556–1572 (2025)
14. Liu, R., Liu, M., Sheng, B., Li, H., Li, P., Song, H., Zhang, P., Jiang, L., Shen, D.: Nhbs-net: A feature fusion attention network for ultrasound neonatal hip bone segmentation. IEEE transactions on medical imaging **40**(12), 3446–3458 (2021)
15. Long, J., Shelhamer, E., Darrell, T.: Fully convolutional networks for semantic segmentation. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 3431–3440 (2015)
16. Luan, K., Li, Z., Li, J.: An efficient end-to-end cnn for segmentation of bone surfaces from ultrasound. Computerized Medical Imaging and Graphics **84**, 101766 (2020)
17. 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)
18. Quader, N., Hodgson, A., Abugharbieh, R.: Confidence weighted local phase features for robust bone surface segmentation in ultrasound. In: Workshop on Clinical Image-Based Procedures. pp. 76–83. Springer (2014)
19. Rahman, A., Bandara, W.G.C., Valanarasu, J.M.J., Hacihaliloglu, I., Patel, V.M.: Orientation-guided graph convolutional network for bone surface segmentation. In: International conference on medical image computing and computer-assisted intervention. pp. 412–421. Springer (2022)
20. Rahman, A., Valanarasu, J.M.J., Hacihaliloglu, I., Patel, V.M.: Simultaneous bone and shadow segmentation network using task correspondence consistency. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 330–339. Springer (2022)
21. 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)
22. Soliman, S.B., Davis, J.J., Muh, S.J., Vohra, S.T., Patel, A., van Holsbeeck, M.T.: Ultrasound evaluations and guided procedures of the painful joint arthroplasty. Skeletal radiology **51**(11), 2105–2120 (2022)
23. Su, J., Ahmed, M., Lu, Y., Pan, S., Bo, W., Liu, Y.: Roformer: Enhanced transformer with rotary position embedding. Neurocomputing **568**, 127063 (2024)
24. Tzatzairis, T., Skarentzos, K., Grammatikos, C., Karamalis, C., Korakianitis, K., Kourempeles, R., Drosos, G.: Ultrasound applications in pediatric orthopedics. Archives of Bone and Joint Surgery **12**(7), 457 (2024)
25. Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A.N., Kaiser, Ł., Polosukhin, I.: Attention is all you need. Advances in neural information processing systems **30** (2017)

26. Villa, M., Dardenne, G., Nasan, M., Letissier, H., Hamitouche, C., Stindel, E.: Fcn-based approach for the automatic segmentation of bone surfaces in ultrasound images. International journal of computer assisted radiology and surgery **13**(11), 1707–1716 (2018)
27. Wang, P., Patel, V.M., Hacihaliloglu, I.: Simultaneous segmentation and classification of bone surfaces from ultrasound using a multi-feature guided cnn. In: International conference on medical image computing and computer-assisted intervention. pp. 134–142. Springer (2018)
28. Wang, P., Vives, M., Patel, V.M., Hacihaliloglu, I.: Robust real-time bone surfaces segmentation from ultrasound using a local phase tensor-guided cnn. International Journal of Computer Assisted Radiology and Surgery **15**(7), 1127–1135 (2020)
29. Yan, W., Li, H., Huang, S., Qian, L., Ding, H., Wang, G., Zhao, Z.: In-situ ultrasound guided closed long bone shaft fracture treatments. In: 2025 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). pp. 1302–1303. IEEE (2025)
30. Zhang, Z., Liu, Q., Wang, Y.: Road extraction by deep residual u-net. IEEE Geoscience and Remote Sensing Letters **15**(5), 749–753 (2018)