Naka-SAM: A Cognition-Inspired Framework with Nakagami Prior for Ultrasound Segmentation

Juan Chen (chenjuan1011@xju.edu.cn)[1,2], Jiajie Wu (20251303213@stu.xju.edu.cn)[1], Lei Guo (Leiguo@uestc.edu.cn)[3], Wenping Ge (wenpingge@xju.edu.cn)[1] & Jie Ma(majie@stu.xju.edu.cn)[1]

[1]College of Computer Science and Technology, Xinjiang University,

[2]Joint International Research Laboratory of Silk Road MultilingualCognitive Computing, Xinjiang University

[3]School of Computer Science and Engineering,University of Electronic Science and Technology

Abstract

Radiologists often implicitly rely on tissue backscattering characteristics to interpret noisy ultrasound (US) images under challenging imaging conditions. However, existing deep learning-based segmentation models rarely incorporate such intrinsic physical statistics into representation learning. To address this limitation, we propose Naka-SAM, a physics-informed ultrasound segmentation framework built upon the Segment Anything Model (SAM). Specifically, we design a Physical Prior Generator (PPG) to model ultrasound backscattering statistics through learnable Nakagami distribution estimation, where the shape parameter m characterizes local scattering concentration and the scale parameter Ω reflects backscattered energy variations. These physics-informed priors provide domain-invariant tissue representations associated with underlying microstructural properties. Furthermore, a dual-stage gated fusion strategy is introduced to progressively integrate physical priors with both low-level structural features and high-level semantic representations, thereby improving robustness under noisy and low-contrast imaging conditions. Extensive experiments across seven ultrasound datasets demonstrate that Naka-SAM consistently outperforms both conventional segmentation methods and recent SAM-adapted frameworks, while exhibiting strong cross-domain generalization capability across heterogeneous ultrasound imaging scenarios.

Keywords: Ultrasound segmentation; Nakagami Distribution; prior features; foundation model

Introduction

Medical image segmentation is a core component of computer-aided diagnosis, fundamentally aiming to build the representation and segmentation capabilities for medical images. It facilitates rapid lesion localization and quantitative tissue characterization, thereby potentially improving diagnostic efficiency and clinical decision-making (X. Zhang et al., 2025).However, the inherent imaging uncertainty in ultrasound imaging presents a significant challenge for robust feature learning. With the rapid development of deep learning technologies and the emergence of the Segment Anything Model (SAM) (Kirillov et al., 2023) and its variant, Medical SAM adapter(Wu et al., 2025), the field of medical image segmentation has undergone a paradigm shift. These foundational models demonstrate strong zero-shot generalization capabilities in natural image or general medical imaging scenarios. However, directly applying these general-purpose visual models to ultrasound (US) imaging still faces fundamental challenges. Unlike natural images, which have clear object boundaries, ultrasound images are full of imaging uncertainty, such as severe speckle noise, acoustic shadow artifacts, and low-contrast soft tissue boundaries. Existing deep learning methods, whether based on CNN-based U-Net (Ronneberger et al., 2015)or Transformer-based architectures(Cao et al., 2022) , largely rely on local texture and edge information for pixel-level classification. This bottom-up, pixel-by-pixel or patch-wise feature learning mechanism often lacks robustness when faced with the low signal-to-noise ratio and blurred boundaries inherent in ultrasound images.

The persistent performance bottleneck in automated segmentation is arguably an *epistemological* one: machine models typically optimize for local pixel-level coherence, whereas experienced radiologists often rely on global contextual reasoning and prior anatomical knowledge. From the perspective of cognitive science, radiological interpretation is characterized by a holistic "gist " extraction that precedes focal analysis. This top-down influence allows experts to maintain *structural consistency* (Gilbert & Li, 2013) despite the significant speckle noise and acoustic artifacts inherent in ultrasound imaging.We posit that this intuition is effectively an internal calibration against the statistical regularities in tissue backscatter.

In ultrasound physics, these regularities are rigorously described by the Nakagami distribution (Nakagami, 1960). The shape parameter m serves as a critical diagnostic index, quantifying microstructural heterogeneity ranging from pre-Rayleigh scattering ($m < 1$) in malignant tissues to post-Rayleigh modes ($m > 1$) in resolved media (Shankar, 2000; Tsui et al., 2008). While traditional deep learning approaches have typically treated these physical invariants as extraneous noise, recent physics-informed frameworks have begun to acknowledge their utility (Banerjee et al., 2025). Naka-SAM, however, moves beyond treating physics as a mere regularization term; it reformulates physical laws into a learnable representation. By integrating the Nakagami m-parameter as a dynamic prior, our model mimics the radiologist's ability to selectively filter sensory data through a physics-informed understanding of tissue properties. To operationalize this synthesis, we propose **Naka-SAM**: a Segment Anything Model (SAM) framework driven by Nakagami priors. Inspired by the theory of *probabilistic feature matching* (Knill & Pouget, 2004), our approach shifts toward a generative architecture via a Physical Prior Generator (PPG). This module explicitly estimates Nakagami distribution parameters—dynamically ad-

In G. Boleda, I. Dautriche, M. Maldonado, C. Saldana and T. Torrent (Eds.), *Proceedings of the 48th Annual Conference of the Cognitive Science Society*

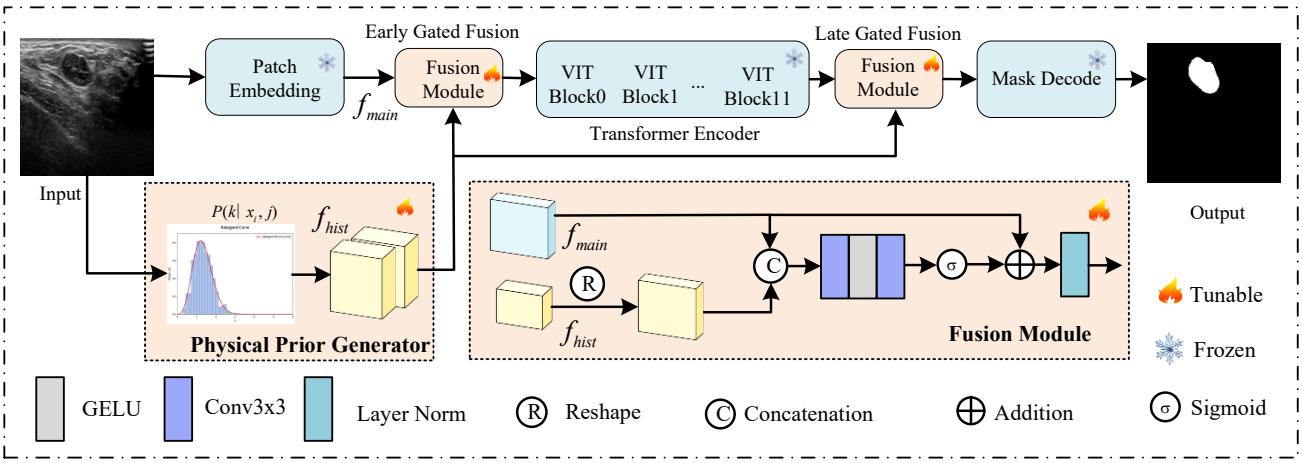


Figure 1: Overview of the Naka-SAM architecture. The framework employs a Physical Prior Generator (PPG) to extract Nakagami-based "physical fingerprints" (f_{hist}), which are adaptively integrated with SAM's semantic features (f_{main}) via dual-stage Fusion Modules. Fire and snowflake icons denote tunable and frozen parameters, respectively.

justing the shape parameter m and the scaling parameter Ω to capture the acoustic fingerprints of tissue. Inspired by adaptive evidence integration mechanisms, we introduce a dual-stage gated fusion strategy: an early-stage gate suppresses non-physical noise, while a late-stage gate acts as a regulator, dynamically amplifying physical priors when visual uncertainty is high. Our primary contributions are summarized as follows:

- **Physical Prior Generator (PPG):** A module that transforms statistical acoustic properties into a differentiable representation, enabling the network to perceive intrinsic tissue characteristics.
- **Dual-stage Gated Fusion Strategy:** A mechanism simulating adaptive evidence integration by bridging global physical hypotheses with local visual verification, enhancing robustness in low-SNR environments.
- **Cross-domain Validation:** Extensive experiments across multiple datasets confirming that embedding invariant physical laws facilitates robust perception, particularly when handling low contrast and speckle noise in ultrasound images.

Methodology

Human radiologists, when interpreting ultrasound images, employ a global-to-local cognitive strategy using the statistical characteristics of tissue echoes, To simulate this process,we propose Naka-SAM for ultrasound image segmentation. Traditional deep learning methods typically start with a "tabula rasa" state and learn pixel-by-pixel or patch-by-patch representations. In contrast, As shown in Figure 1,our framework explicitly incorporates the intrinsic physical tissue characteristics as learnable physics-informed priors.

Modeling Ultrasound Physics as Cognitive Priors

In ultrasound imaging, from a physics perspective, the envelope of the backscattered signals from biological tissues follows the Nakagami distribution. We hypothesize that expert radiologists, when distinguishing tissue types, are implicitly estimating the parameters of this distribution.

Formally, for a random variable x (representing the envelope amplitude of the feature map), the probability density function (PDF) of the Nakagami distribution is defined as:

$$f(x \mid m, \Omega) = \frac{(2m^m)}{\Gamma(m)\Omega^m} x^{2m-1} \exp\left(-\frac{m}{\Omega}x^2\right)$$

Where:

- $x \geq 0$ represents pixel intensity or feature amplitude.
- $m \geq 0.5$ is the shape parameter, describing the scattering conditions (ranging from pre-Rayleigh to Rice distribution). In our framework, this parameter quantifies microstructural heterogeneity, reflecting the spatial distribution and density of scattering centers.
- $\Omega > 0$ is the scale parameter, representing the time-averaged power of the backscattered signal ($\Omega = \mathbb{E}[x^2]$), corresponding to signal energy.
- $\Gamma(\cdot)$ is the Gamma function.

Physical Prior Generator (PPG)

To endow the network with explicit physical reasoning, we propose the Physical Prior Generator (PPG). Beyond simple feature extraction, this module serves as a differentiable *cognitive schema* that internalizes the physics of tissue backscattering, particularly the Nakagami statistical model, into the network parameters.

Dynamic Internalization of Physical Parameters Unlike traditional methods that rely on the static fitting of physical parameters m and Ω, the PPG treats them as evolving internal states of the network, denoted as $\theta = \{m, \Omega\}$. This allows the model to adaptively seek the optimal physical manifold during end-to-end training via gradient descent.

To constrain the generated prior distributions within a physically valid space, specifically requiring $m \geq 0.5$ and $\Omega > 0$, we implement a Physically-Constrained Reparameterization mechanism employing the Softplus function:

$$m = \text{Softplus}(w_m) + 0.5, \quad \Omega = \text{Softplus}(w_\Omega) + \epsilon \tag{1}$$

where w_m and w_Ω represent the latent unconstrained weights of the network, and $\epsilon = 10^{-6}$ is a numerical stability term. The Softplus function, defined as $\text{Softplus}(x) = \ln(1 + \exp(x))$, is specifically selected for its smooth and second-order differentiable properties. Unlike the standard ReLU function, Softplus provides a continuous gradient profile that facilitates stable optimization during backpropagation, allowing the PPG to effectively internalize the physics of tissue backscattering within the neural architecture. By incorporating an offset of 0.5 for the shape parameter m, this mechanism effectively mitigates the risk of the model generating distributions that lack physical validity. Consequently, the PPG functions as a differentiable physics-aware representation that goes beyond generic feature extraction, ensuring that the learned representations remain grounded in the physical laws of ultrasound propagation.

Probabilistic Soft-Matching and Differentiable Density Estimation To achieve matching of discrete physical modalities within a continuous feature space, we abandon hard binning operations in favor of a differentiable probabilistic response inspired by Kernel Density Estimation (KDE).

The PPG calculates the posterior probability that a feature $x_{i,j}$ belongs to the k-th physical mode (defined by m_k, Ω_k). This process simulates the probabilistic feature matching mechanism found in cognitive science:

$$P(k \mid x_{i,j}) = \frac{\exp\left(\Phi(x_{i,j}, m_k, \Omega_k)/\tau\right)}{\sum_{l=1}^{K} \exp\left(\Phi(x_{i,j}, m_l, \Omega_l)/\tau\right)} \tag{2}$$

Here, $\Phi(\cdot)$ represents the physical energy function, while τ serves as a pre-defined temperature coefficient. This hyperparameter, empirically fixed at 0.07 to enhance distributional discriminability, dynamically regulates the sensitivity of the model to statistical features. Functionally, this mechanism is functionally analogous to contrast sensitivity in the human visual system, determining the precision with which the network matches latent physical priors to visual evidence.

Log-Domain Physical Response Considering that the exponential nature of the Nakagami distribution may induce gradient vanishing during backpropagation, we map the physical consistency calculation to the logarithmic domain. The response map $H \in \mathbb{R}^{B\times K\times H\times W}$ for the k-th physical channel is formalized as a log-likelihood estimation:

$$\ln(H_k) \propto \underbrace{\ln(m) - \ln(\Gamma(m)) - m\ln(\Omega)}_{\text{Intrinsic Physical Entropy}} + \underbrace{(2m-1)\ln(x) - \frac{m}{\Omega}x^2}_{\text{Data Likelihood}} \tag{3}$$

By adopting this formulation, the PPG achieves significantly improved numerical stability compared to the original exponential formulation.

Adaptive Gated Fusion for Evidence Integration

Cognitive science suggests that perception is an adaptive evidence integration process. When visual input is ambiguous, the brain relies more on global statistical priors; conversely, when edges are clear, local visual cues dominate. To simulate this mechanism, we employ an Adaptive Gated Fusion module. Let $F_{\text{main}} \in \mathbb{R}^{C\times H\times W}$ be the semantic features extracted by the Transformer backbone, and $F_{\text{hist}} \in \mathbb{R}^{C\times H\times W}$ represent the microstructural features projected from the Nakagami layer.As shown in the fusion module in Figure 1, the fused representation F_{out} is calculated through a nonlinear gating mechanism:

$$G = \sigma(W_g * [F_{\text{main}} \parallel F_{\text{hist}}] + b_g) \tag{4}$$

$$F_{\text{out}} = \text{LayerNorm}(F_{\text{main}} + G \odot \phi(F_{\text{hist}})) \tag{5}$$

Where:

- $\parallel$ denotes concatenation along the channel dimension.
- $\sigma(\cdot)$ is the Sigmoid activation function, generating the gating map $G \in [0, 1]^{C\times H\times W}$ as the adaptive gating map, which dynamically determines the reliance on physical priors.
- W_g and b_g are the weights and biases of the convolutional gating network.
- $\phi(\cdot)$ is the linear projection used to align the feature spaces.
- $\odot$ denotes the Hadamard product (element-wise multiplication).

Dual-Stage Prior-Guided Feature Fusion

Finally, to enforce structural consistency at different scales, we integrate this fusion mechanism into two key cognitive stages:

- **Early-Sensory Stage**: Immediately after Patch Embedding, original statistical descriptors are fused to guide the encoder's attention to regions with physical relevance.
- **Late-Decision Stage**: Before the final segmentation head, global microstructural statistics are reintroduced to ensure the predicted mask maintains the structural consistency of the underlying tissue.

Table 1: Quantitative comparison of Naka-SAM with state-of-the-art methods and SAM variants across five ultrasound datasets. ↑ indicates higher is better; ↓ indicates lower is better. Bold font denotes the best results in each category.

Method	CAMUS (LA)		TN3K		BUSI		CAMUS (LV)		CAMUS (MYO)	
	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓	Dice↑	HD↓
U-Net	91.00	12.91	79.01	34.12	78.11	33.60	93.56	9.90	86.86	16.87
AAU-Net	91.33	12.12	82.28	30.53	80.81	30.39	93.32	9.97	86.98	16.49
SwinUNet	89.80	14.74	70.08	44.13	67.23	47.02	91.72	12.80	84.46	20.25
MISSformer	91.18	11.82	79.42	32.85	78.43	30.10	93.25	9.94	86.57	16.50
TransUNet	91.37	12.46	81.44	30.98	82.22	27.54	**93.60**	**9.60**	87.20	17.25
H2Former	90.98	11.92	82.48	30.58	81.48	27.84	93.44	**9.60**	**87.31**	16.60
SAM	17.28	193.70	29.59	134.87	54.01	82.39	28.18	196.64	29.42	184.10
MedSAM	88.06	15.70	71.09	42.91	77.75	34.26	87.52	15.28	76.07	25.72
SAMed	90.92	12.60	80.40	31.29	74.82	34.60	87.67	13.24	82.60	19.48
MSA	91.80	11.59	82.67	29.15	81.66	28.87	90.95	11.29	82.47	19.28
GSAM	90.65	10.88	86.68	20.48	81.65	28.50	92.39	11.59	86.09	**10.88**
Naka-SAM	**92.65**	**10.81**	**87.91**	**19.58**	**84.65**	**25.33**	93.12	10.71	86.41	11.78

Experiments

Experimental Setup and Datasets

We evaluated Naka-SAM on seven challenging ultrasound datasets: TN3K (Gong et al., 2023), TG3K (Wunderling et al., 2017), BUSI (Al-Dhabyani et al., 2020), CAMUS (Leclerc et al., 2019), DDTI (Pedraza et al., 2015), UDIAT (Yap et al., 2020), and HMC-QU (Kiranyaz et al., 2020). Following the CC-SAM(Gowda & Clifton, 2024) protocol, we partitioned the BUSI dataset into a 7:1:2 ratio for training, validation, and testing. For the TN3K and TG3K datasets, we followed the TRFE strategy , using TG3K exclusively for training. To assess zero-shot generalization, models were trained on seen datasets and evaluated on unseen subsets (DDTI, UDIAT, HMC-QU) without further fine-tuning.

Implementation Details

The backbone is based on the ViT-B architecture of SAM. To preserve general visual representations while enabling physics-aware adaptation, we fine-tuned only the Physical Prior Generator (PPG) and the Gated Fusion modules, resulting in a highly parameter-efficient architecture with only 39.65M trainable parameters. The PPG module utilizes a histogram estimation layer combined with a probabilistic feature matching function, with its temperature parameter τ set to 0.07 to enhance distributional discriminability. Naka-SAM was implemented in PyTorch and trained on a single NVIDIA RTX 3090 GPU for 200 epochs. Taking the BUSI dataset (containing 647 images) as a representative benchmark, the entire training process for 200 epochs takes approximately four hours.We optimized the model using AdamW ($\beta_1 = 0.9, \beta_2 = 0.999$, weight decay = 0.0001) with an initial learning rate of 1×10^{-4}. Standard augmentations, including elastic deformation and rotation, were applied to 256×256 input images. Performance was quantified via Dice Similarity Coefficient (Dice) and Hausdorff Distance (HD).

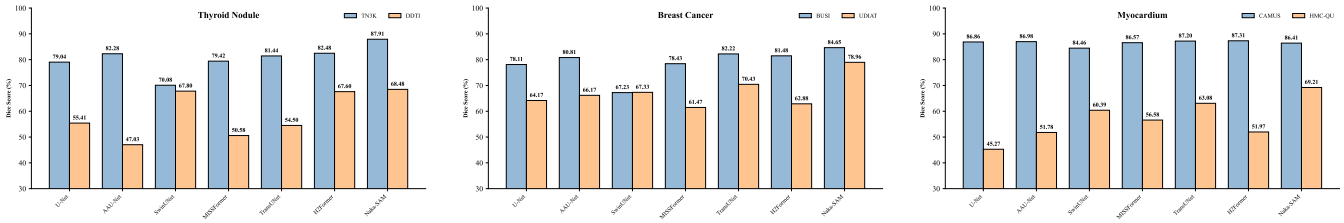
Quantitative Results and Analysis

Comparison with Established Methods: As shown in Table 1, Naka-SAM achieves superior performance on several challenging ultrasound segmentation tasks, particularly under low-contrast and poor-quality imaging conditions. On TN3K (Dice: 87.91%) and BUSI (Dice: 84.65%), Naka-SAM significantly outperforms advanced architectures such as TransUNet and H2Former. This advantage is especially evident in terms of Hausdorff Distance (HD). For example, on TN3K, the HD is reduced to 19.58, indicating that the incorporation of physics-informed priors effectively alleviates boundary leakage and suppresses artifact-induced segmentation errors.

Compared with other datasets, CAMUS contains a substantially larger number of samples (19,232 images), resulting in relatively stable and consistent performance across different segmentation models.

Comparison with SAM Variants: As illustrated in Table 1, we further compare Naka-SAM with representative SAM-based architectures, including the original SAM (Kirillov et al., 2023), SAMed (K. Zhang & Liu, 2023), MSA (Wu et al., 2025), MedSAM (Ma et al., 2024), and GSAM (Kato et al., 2024). Although these methods provide competitive baselines, Naka-SAM consistently achieves the best overall performance across multiple benchmarks.

Vanilla SAM exhibits substantially inferior performance, with Dice scores ranging from 17.28% to 54.01% and extremely large HD values. This performance degradation mainly arises from three limitations: (1) *Domain mismatch*: SAM is pre-trained on natural image datasets (e.g., SA-1B) and lacks adaptation to ultrasound-specific artifacts such as speckle noise and acoustic shadowing; (2) *Feature irrelevance*: its feature extraction mechanism prioritizes generic visual patterns rather than the statistical characteristics of ultrasound tissue backscatter; and (3) *Lack of physics-informed guidance*: without structured physical priors, the model fre-



(a) Thyroid Nodule (b) Breast Tumor (c) Myocardium

Figure 2: Experimental results of the proposed method on three different datasets: (a) Thyroid Nodule, (b) Breast Tumor, and (c) Myocardium. We evaluate the cross-domain robustness of Naka-SAM by comparing its performance on seen distributions (blue bars) versus unseen, zero-shot distributions (orange bars).

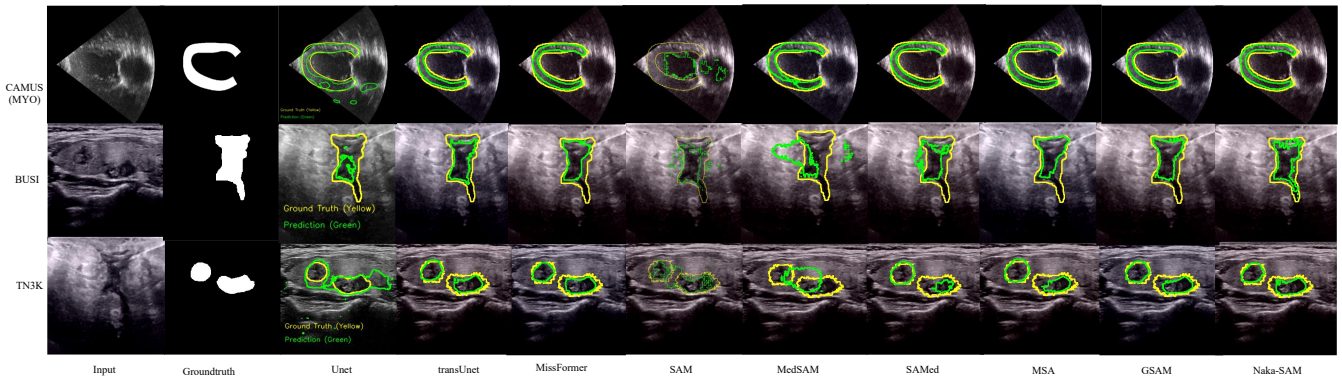


Figure 3: Qualitative comparison of segmentation results on CAMUS (MYO), BUSI, and TN3K datasets. Yellow contours represent the ground truth, and green contours represent the model's predictions. Naka-SAM achieves superior boundary alignment and robustness against noise.

quently misclassifies noise and artifacts as anatomical structures.

In contrast, Naka-SAM achieves performance improvements exceeding 5% on datasets such as CAMUS (MYO) and BUSI. These results suggest that merely increasing medical training data is insufficient for robust ultrasound segmentation, whereas incorporating physics-informed priors provides more effective structural guidance under challenging imaging conditions.

Quantitative Evaluation of Zero-shot Generalization

As illustrated in Figure 2,we evaluate the cross-domain robustness of Naka-SAM by comparing its performance on seen distributions (blue bars) versus unseen, zero-shot distributions (orange bars).

- **In-distribution Performance (Blue):** For datasets included in the training phase, such as TN3K, BUSI, and CAMUS-LA, Naka-SAM consistently achieves state-of-the-art Dice scores, effectively setting a high ceiling for ultrasound segmentation.
- **Zero-shot Generalization (Orange):** The model's true strength is demonstrated in the orange categories, which represent datasets that were never seen during training. Despite the significant domain shift caused by different imaging equipment and patient demographics, Naka-SAM maintains stable performance with minimal decay compared to the training benchmarks.

Specifically, As illustrated in Figure 2(b), the model was trained exclusively on the BUSI dataset, while the UDIAT dataset was utilized solely for zero-shot validation to assess cross-domain generalization. The performance gap between the seen BUSI dataset and the unseen UDIAT dataset is significantly narrower than that of competing SAM variants. This resilience suggests that our Physical Prior Generator (PPG) primarily extracts domain-invariant physical fingerprints rather than merely overfitting to dataset-specific noise patterns.

Qualitative Results: Simulating Expert structural consistency

Visualizations (Figure 3) intuitively demonstrate Naka-SAM's superiority. In BUSI, where speckle noise obscures tumor edges, models like MedSAM produce fragmented and drifting boundaries. In contrast, Naka-SAM's boundaries (green) closely align with the gold standard (yellow). In CAMUS myocardium segmentation, Naka-SAM maintains the topological consistency of the annular structure, whereas competitors often show "breaks" in low-contrast regions. This phenomenon proves the model's *predictive coding* capability—predicting

Table 2: Ablation Study Results on BUSI and TN3K

Components	BUSI (Dice)↑	BUSI (HD)↓	TN3K (Dice)↑	TN3K (HD)↓
Baseline (GSAM)	81.65	28.5	86.68	20.48
w/ PPG	83.66	26.21	86.92	20.27
Full Model (PPG + fusion module)	84.65	25.33	87.26	19.89

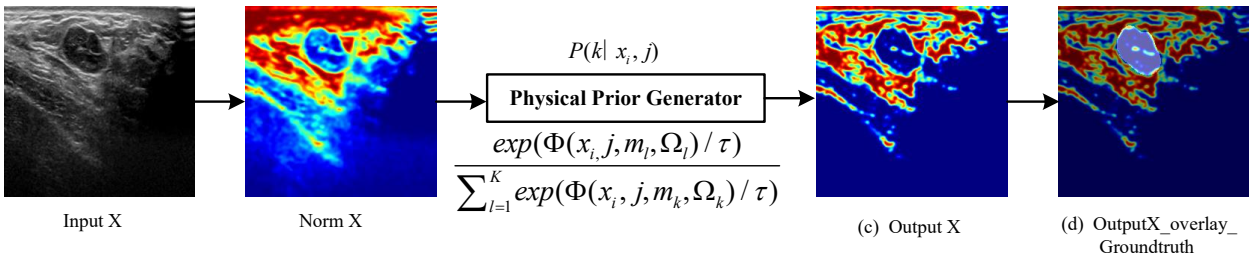


Figure 4: Visualization of the PPG workflow. It illustrates the transition from input X to normalized features, the estimation of Nakagami parameters (m, Ω) via probabilistic feature matching, and the resulting physical fingerprint output overlayed with the ground truth.

true boundaries via internal physical models to actively filter out signals that contradict physical logic.

Ablation Study

To evaluate the contributions of different components in the proposed model, we conducted an ablation study on two datasets: BUSI and TN3K. We compared four configurations by systematically removing key components of our model. As shown in the table 2, the addition of the PPG and fusion module improved the performance across both Dice and HD scores, particularly on the BUSI dataset, where the Dice score increased from 81.65 to 83.68 and the HD decreased from 28.5 to 26.28. These improvements demonstrate the effectiveness of incorporating PPG features and the fusion module into our model.

Discussion: Explanatory Power of Physical Fingerprints

The PPG output (Figure 4) reveals the internal decision mechanism. The generated maps clearly capture Nakagami distribution differences between microstructural heterogeneity. This structured physical prior serving as a cross-domain invariant feature, is key to Naka-SAM's exceptional zero-shot generalization.

General Discussion

Physical Fingerprints as Structured Priors

Traditional deep learning models often exhibit limited robustness under heterogeneous ultrasound imaging conditions due to their strong reliance on local appearance features. In contrast, the Nakagami parameters captured by the proposed PPG module encode domain-invariant statistical characteristics of tissue backscatter and microstructural heterogeneity. By integrating these physics-informed priors into the segmentation process, the proposed framework provides complementary structural constraints that help reduce boundary fragmentation and anatomically inconsistent predictions, particularly in low-contrast and speckle-corrupted regions.

Uncertainty-Aware Feature Integration

Our dual-stage gating mechanism performs adaptive integration of semantic features and physics-informed priors under challenging ultrasound imaging conditions. The early-stage gate suppresses noise-sensitive responses during low-level feature extraction, while the late-stage gate dynamically enhances the contribution of physical priors in ambiguous or low-contrast regions. Experimental results suggest that the proposed mechanism improves structural consistency and robustness against speckle noise and boundary uncertainty.

Conclusion

This study introduces Naka-SAM, a physics-informed framework for ultrasound image segmentation that integrates learnable Nakagami-based priors with adaptive gated feature fusion. By explicitly modeling the statistical characteristics of ultrasound backscatter, the proposed framework improves structural consistency and robustness under noisy and low-contrast imaging conditions. Comprehensive evaluations across seven datasets demonstrate that Naka-SAM not only outperforms existing segmentation methods but also exhibits strong zero-shot generalization capability across heterogeneous ultrasound domains. Future work will investigate the extension of the proposed physics-informed prior modeling strategy to other medical imaging modalities.

Acknowledgments

This work was supported by Xinjiang University Training Program of lnnovation and Entrepreneurship for Undergraduates(No. 202510755064).

References

Al-Dhabyani, W., Gomaa, M., Khaled, H., & Fahmy, A. (2020). Dataset of breast ultrasound images. *Data in Brief*, *28*, 104863. https://doi.org/10.1016/j.dib.2019.104863

Banerjee, C., Nguyen, K., Salvado, O., Tran, T., & Fookes, C. (2025). Physics-informed machine learning for medical image analysis. *ACM Computing Surveys*, *58*(4), 1–35.

Cao, H., Wang, Y., Chen, J., Jiang, D., Zhang, X., Tian, Q., & Wang, M. (2022). Swin-unet: Unet-like pure transformer for medical image segmentation. *Proceedings of the European Conference on Computer Vision (ECCV)*, 205–218. https://doi.org/10.1007/978-3-031-19784-0_12

Gilbert, C. D., & Li, W. (2013). Top-down influences on visual processing. *Nature reviews neuroscience*, *14*(5), 350–363.

Gong, H., Chen, J., Chen, G., Li, H., Wang, Y., Zhang, J., Yu, J., Li, G., & Lin, L. (2023). Thyroid region prior guided attention for ultrasound segmentation of thyroid nodules. *Computers in Biology and Medicine*, *155*, 106389. https://doi.org/10.1016/j.compbiomed.2022.106389

Gowda, S. N., & Clifton, D. A. (2024). Cc-sam: Sam with cross-feature attention and context for ultrasound image segmentation. *European Conference on Computer Vision*, 108–124.

Kato, S., Mitsuoka, H., & Hotta, K. (2024). Generalized sam: Efficient fine-tuning of sam for variable input image sizes. *European Conference on Computer Vision (ECCV)*, 167–182. https://doi.org/10.1007/978-3-031-72684-2_9

Kiranyaz, S., Uzunsoy, E., Ince, T., Iosifidis, A., Gabbouj, M., & Gabbouj, M. (2020). Left ventricular wall motion estimation by active polynomials for acute myocardial infarction detection. *IEEE Access*, *8*, 210301–210317. https://doi.org/10.1109/ACCESS.2020.3038743

Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, C., Gustafson, L., Xiao, T., Whitehead, S., Berg, A. C., Lo, W.-Y., et al. (2023). Segment anything. *Proceedings of the IEEE/CVF international conference on computer vision*, 4015–4026.

Knill, D. C., & Pouget, A. (2004). The bayesian brain: The role of uncertainty in neural coding and computation. *TRENDS in Neurosciences*, *27*(12), 712–719.

Leclerc, S., Smistad, E., Pedrosa, J., Ostvik, A., Cervenansky, F., Espinosa, T., Espeland, T., Berg, E. A. R., Julliard, P.-M., Feuillet, P., et al. (2019). Deep learning for segmentation using an open large-scale dataset in 2D echocardiography. *IEEE Transactions on Medical Imaging*, *38*(9), 2198–2210. https://doi.org/10.1109/TMI.2019.2900516

Ma, J., He, Y., Li, F., Zhang, X., & Wang, Y. (2024). Segment anything in medical images. *Nature Communications*, *15*(1), 689. https://doi.org/10.1038/s41467-024-46898-9

Nakagami, M. (1960). The m-distribution—a general formula of intensity distribution of rapid fading. In *Statistical methods in radio wave propagation* (pp. 3–36). Elsevier.

Pedraza, L., Vargas, C., Narváez, F., Durán, O., Muñoz, E., & Romero, E. (2015). An open access thyroid ultrasound image database. *10th International symposium on medical information processing and analysis*, *9287*, 188–193.

Ronneberger, O., Fischer, P., & Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. *Proceedings of the 18th International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, 234–241. https://doi.org/10.1007/978-3-319-24574-4_28

Shankar, P. M. (2000). A general statistical model for ultrasonic backscattering from tissues. *IEEE transactions on ultrasonics, ferroelectrics, and frequency control*, *47*(3), 727–736.

Tsui, P.-H., Yeh, C.-K., Chang, C.-C., & Chen, W.-S. (2008). Performance evaluation of ultrasonic nakagami image in tissue characterization. *Ultrasonic imaging*, *30*(2), 78–94.

Wu, J., Wang, Z., Hong, M., Ji, W., Fu, H., Xu, Y., Xu, M., & Jin, Y. (2025). Medical SAM adapter: Adapting segment anything model for medical image segmentation. *Medical Image Analysis*, *102*, 103547. https://doi.org/10.1016/j.media.2025.103547

Wunderling, T., Golla, B., Poudel, P., Arens, C., Friebe, M., & Hansen, C. (2017). Comparison of thyroid segmentation techniques for 3d ultrasound. *Medical Imaging 2017: Image Processing*, *10133*, 346–352. https://doi.org/10.1117/12.2254374

Yap, M. H., Goyal, M., Osman, M., Martí, R., Denton, E., Juette, A., & Zwiggelaar, R. (2020). Breast ultrasound region of interest detection and lesion localisation. *Artificial Intelligence in Medicine*, *107*, 101880. https://doi.org/10.1016/j.artmed.2020.101880

Zhang, K., & Liu, D. (2023). Customized segment anything model for medical image segmentation. *arXiv preprint arXiv:2304.13785*. https://arxiv.org/abs/2304.13785

Zhang, X., Chen, E. Z., Zhao, L., Chen, X., Liu, Y., Maihe, B., Duncan, J. S., Chen, T., & Sun, S. (2025). Adapting vision foundation models for real-time ultrasound image segmentation. *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 24–34.