UU-Mamba: Uncertainty-aware U-Mamba for Cardiac Image Segmentation

Ting Yu Tsai[1], Li Lin[2], Shu Hu[2], Ming-Ching Chang[1], Hongtu Zhu[3], Xin Wang[1,*]

[1]University at Albany, State University of New York, {ttsai2, mchang2, xwang56}@albany.edu
[2]Purdue University, {lin1785, hu968}@purdue.edu
[3]University of North Carolina at Chapel Hill, htzhu@email.unc.edu

Abstract—Biomedical image segmentation is critical for accurate identification and analysis of anatomical structures in medical imaging, particularly in cardiac MRI. Manual segmentation is labor-intensive, time-consuming, and prone to errors, highlighting the need for automated methods. However, current machine learning approaches face challenges like overfitting and data demands. To tackle these issues, we propose a new UU-Mamba model, integrating the U-Mamba model with the Sharpness-Aware Minimization (SAM) optimizer and an uncertainty-aware loss function. SAM enhances generalization by locating flat minima in the loss landscape, thus reducing overfitting. The uncertainty-aware loss combines region-based, distribution-based, and pixel-based loss designs to improve segmentation accuracy and robustness. Evaluation of our method is performed on the ACDC cardiac dataset, outperforming state-of-the-art models including TransUNet, Swin-Unet, nnUNet, and nnFormer. Our approach achieves Dice Similarity Coefficient (DSC) and Mean Squared Error (MSE) scores, demonstrating its effectiveness in cardiac MRI segmentation. The code can be accessed at https://github.com/tiffany9056/UU-Mamba.

Index Terms—Cardiac image segmentation, Mamba, ACDC dataset, uncertainty-aware loss, sharpness-aware minimization.

I. INTRODUCTION

Biomedical image segmentation is essential in medical imagery analysis for accurately identifying and delineating anatomical structures and anomalies. In cardiac imaging, segmenting the heart and its components from Magnetic Resonance Imaging (MRI) is critical for diagnosing cardiovascular disorders, the development of treatment plans, and assessing therapeutic outcomes [1], [21]. Cardiac MRI provides high-quality images that detail the structure, function, and composition of the heart [27]. However, manual segmentation of the cardiac structures from these images is labor-intensive, time-consuming, and prone to observer variability, highlighting the need for automated segmentation techniques to improve efficiency and consistency [24], [29].

Cardiac MRI segmentation is crucial but challenging due to variations in heart anatomy, pathological changes, and imaging artifacts. Conventional methods like thresholding and edge detection often fail to capture the complexities of cardiac structures. Machine learning techniques, particularly convolutional neural networks (CNNs) and other Deep Learning models, have shown promise by learning intricate patterns from large datasets. However, these models require extensive annotated data, significant computational resources, and may struggle to generalize across different patient groups and imaging settings.
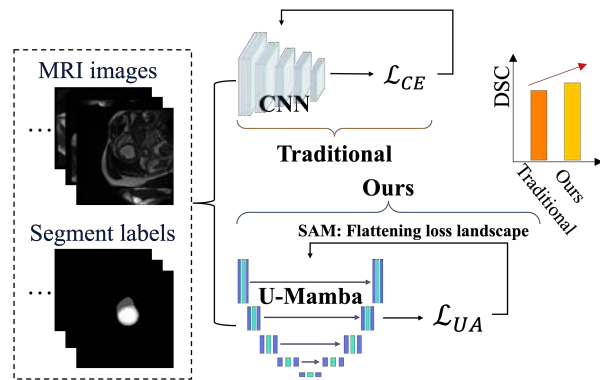

Fig. 1: Comparison between our method and traditional approach. Traditionally, a deep learning model is trained using the cross-entropy loss $\mathcal{L}_{CE}$. Our method enhances U-Mamba by unitizing Uncertainty-aware loss $\mathcal{L}_{UA}$, which is optimized via the SAM optimizer over a flattened loss landscape.

Current research aims to develop more robust, efficient, and generalizable segmentation algorithms to enhance the accuracy and reliability of cardiac MRI analysis [21].

In this paper, we address these challenges by proposing the UU-Mamba model, an extension of the U-Mamba model [23], and improve the training of it by incorporating a **novel uncertainty-aware loss** and a Sharpness-Aware Minimization (SAM) optimizer. Specifically, our newly introduced uncertainty-aware loss function integrates three components:

1) *Region-based loss*: Used for object detection and localization (*e.g.*, Dice coefficient loss).
2) *Distribution-based loss*: Compares predicted and ground truth distributions (*e.g.*, cross-entropy loss).
3) *Pixel-based loss*: Measures differences at the pixel level (*e.g.*, focal loss).

By employing **auto-learnable weights** [16] rather than fixed weights, our model can dynamically adjust the contribution of each loss component based on the uncertainty associated with each prediction. This approach allows the model to prioritize confident predictions and reduce the impact of ambiguous or noisy data. The auto-learnable weights enable the model to balance different aspects of the segmentation task adaptively, improving overall performance and robustness, and addressing issues like class imbalance [5].

To improve the generalization of our model, we utilize the

* Coressponding author.

1

Sharpness-Aware Minimization (SAM) optimizer, which finds parameter values that result in flat minima in the loss landscape. This technique improves the model's ability to generalize [9], [17]–[20]. and mitigates overfitting, a common issue in deep learning applications for medical imaging [26].

Fig. 1 compares our method against other existing methods. In summary, we make the following key contributions:

- We introduce an uncertainty-aware loss combining region-based loss (*e.g.*, Dice coefficient loss), distribution-based loss (*e.g.*, cross-entropy loss), and pixel-based loss (*e.g.*, focal loss). This design allows the model to prioritize confident predictions and reduce the influence of noisy data.
- We integrate Sharpness-Aware Minimization (SAM) optimization into our model. SAM optimization find parameter values that lead to flat minima in the loss landscape, which mitigates overfitting and reduces sensitivity to small perturbations in the input data.
- We evaluate and compare our method against state-of-the-art models (TransUNet [4], Swin-Unet [2], nnUNet [14], U-Mamba [23]), demonstrating superior performance, achieving the highest Dice Similarity Coefficient (DSC) and lowest Mean Squared Error (MSE) by promoting sharper and more precise segmentation boundaries improving segmentation accuracy and robustness on the ACDC dataset.

II. RELATED WORK

A. Cardiac MRI Segmentation

Cardiac MRI segmentation has advanced significantly with deep learning methods, particularly Convolutional Neural Networks (CNNs) [33]. While cardiac segmentation datasets such as ACDC have played a crucial role in this progress, existing methods still face challenges [1]. Bernard *et al.* [1] showed that CNNs, especially U-Net and its variants, achieve high accuracy in cardiac MRI segmentation. However, these methods often struggle with patient-specific variances and suffer from overfitting and class imbalance. Isensee *et al.* [14] improved segmentation accuracy by combining U-Net and V-Net architectures, relying heavily on extensive annotations. Despite improvements, their approach still struggles to generalize to new data due to its dependence on large annotated datasets. Transfer learning has been explored to mitigate the scarcity of labeled data. Chen *et al.* [3] pre-trained models on the ACDC dataset and fine-tuned them on smaller datasets. This approach improves performance but can suffer from domain shift issues when applied to different data distributions. Traditional methods often use standard loss functions like Cross-Entropy (CE) loss, which may not handle class imbalance or the fine details required for accurate cardiac segmentation. These methods also tend to neglect optimizing for flatter minima in the loss landscape, which can enhance generalization.

B. Mamba for Medical Image Segmentation

The Mamba architecture [7] represents a significant advancement in medical image segmentation, combining the strength of CNNs and Vision Transformers (ViTs). It excels in managing long sequences and integrating global contextual information, which is crucial for precision in medical imaging tasks [35]. Building on this foundation, U-Mamba [23] enhances the traditional U-Net framework by incorporating attention mechanisms and multi-scale processing. This significantly improves segmentation accuracy and robustness by focusing on pertinent details within complex anatomical structures and effectively processing a range of scales, from broad contextual information to precise low-level details. U-Mamba also integrates deep supervision [38] to speed up training and improve convergence, making it efficient and reliable for clinical use, where fast and accurate image processing is crucial. The adaptability of the Mamba architecture is further demonstrated in specialized variants like Weak-Mamba-UNet [34], which excels in scribble-based segmentation tasks, handling complex scenarios with enhanced performance. In summary, models based on U-Mamba facilitates superior segmentation across various medical applications, such as cardiac MRI and histopathological imaging.

III. METHOD

Fig. 2 shows the overview of our proposed UU-Mamba architecture with training improvements. § III-A describes the Mamba block and the U-Mamba network, detailing the integration of state space models and their role in capturing long-range dependencies. § III-B presents our proposed uncertainty-aware loss, explaining how it combines multiple loss functions to enhance model performance and robustness. § III-C describes the Sharpness-Aware Minimization Optimization, highlighting its benefits in achieving flat minima in the loss landscape to improve generalization and reduce overfitting.

A. Mamba Block and U-Mamba Network

The U-Mamba network integrates the strengths of the Mamba block [7] and U-Net [30] to enhance global context understanding and improve the accuracy of medical image segmentation. The Mamba block, designed for Selective Structured State Space Sequence Models (S6), efficiently handles long-range dependencies and sequential information, making it highly suitable for medical image segmentation tasks.

The State Space Model (SSM) [8] describes a system in terms of its internal state and observations over time, effectively modeling sequences through underlying states. Given the input state vector $\mathbf{x}_t$, control input $\mathbf{u}_t$, process noise $\mathbf{w}_t$, state transition matrix $\mathbf{A}$, and control input matrix $\mathbf{B}$, the basic form is represented by:

$$\mathbf{x}_{t+1} = \mathbf{A}\mathbf{x}_t + \mathbf{B}\mathbf{u}_t + \mathbf{w}_t. \tag{1}$$

For the observation matrix $\mathbf{C}$, feedthrough matrix $\mathbf{D}$, and observation noise $\mathbf{v}_t$, the observation $\mathbf{y}_t$ is then given by:

$$\mathbf{y}_t = \mathbf{C}\mathbf{x}_t + \mathbf{D}\mathbf{u}_t + \mathbf{v}_t. \tag{2}$$

The S6 architecture enhances traditional state space models with selective attention mechanisms and structured parameterization. Specifically, the selective attention mechanism can be expressed as:

$$\mathbf{a}_t = \text{softmax}(\mathbf{Q}\mathbf{K}^T/\sqrt{d_k})\mathbf{V}, \tag{3}$$
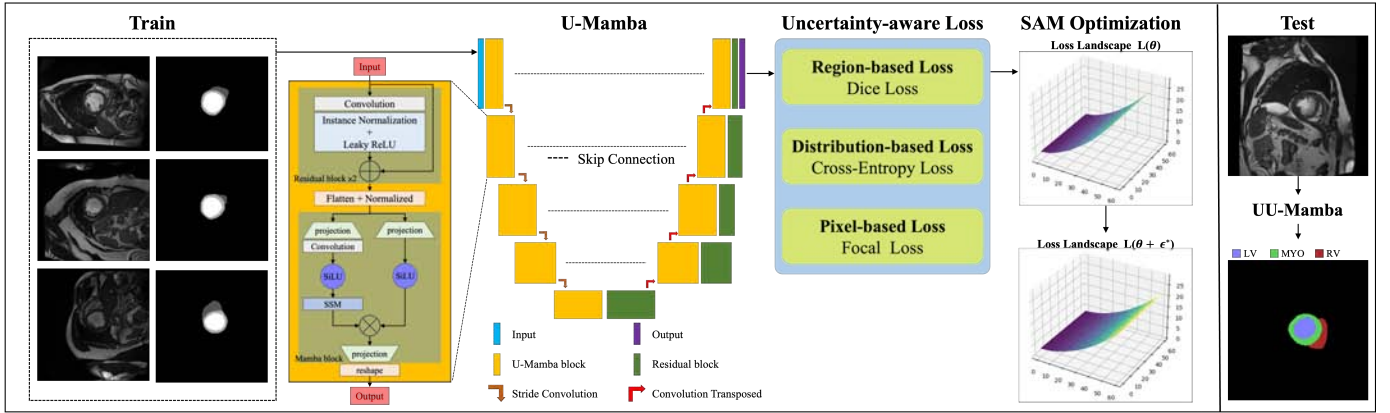
Fig. 2: Overview of our proposed **UU-Mamba model**: Leveraging the U-Mamba architecture, we encode input images and incorporate a novel uncertainty-aware loss function. Optimization is performed using the Sharpness-Aware Minimization (SAM) optimizer, which operates within a flattened loss landscape. Experiments on the ACDC dataset perform 3D heart segmentation on cardiac MRI images, delineating the right ventricle (RV), myocardium (Myo), and left ventricle (LV).

where $\mathbf{Q}$, $\mathbf{K}$, and $\mathbf{V}$ are the query, key, and value matrices derived from the state vector $\mathbf{x}_t$; d_k is the dimension of the key vectors. This selective attention mechanism focuses on relevant input sequence parts and captures complex dependencies.

Integrating S6 within the Mamba block is crucial for sequential medical image processing, such as cardiac MRI segmentation, where capturing temporal dynamics and structure is essential. Our approach focuses exclusively on per-image segmentation tasks, treating each image as an independent entity, which means the state transition and observation matrices ($\mathbf{A}$, $\mathbf{C}$, *etc.*) are applied to individual images.

U-Mamba leverages Mamba's linear scaling advantage to improve CNNs' long-range dependency modeling without the high computational burden of self-attention mechanisms used in Transformers like ViT [31] and SwinTransformer [22]. Fig. 2 shows the U-Mamba block, which consists of two sequential residual blocks followed by a Mamba block. Each block includes convolutional layers, Instance Normalization, and Leaky ReLU activation. Image features are flattened, transposed, normalized, and processed through Mamba blocks with two parallel branches: one with an SSM layer and one without. These features are merged via the Hadamard product, projected back, and transposed to their original shape. The complete U-Mamba network architecture features an encoder with these blocks to capture local features and long-range dependencies, and a decoder composed of residual blocks and transposed convolutions for detailed local information and resolution recovery. Skip connections link hierarchical features from the encoder to the decoder, and the final decoder feature is passed through a $1 \times 1 \times 1$ convolutional layer and a Softmax layer to produce the final segmentation probability map.

B. Uncertainty-aware Loss

Integrating uncertainty into loss functions involves weighting different elements of a loss function according to the estimated uncertainty for each data point [36], [37]. This approach allows the model to prioritize learning from reliable instances and minimize the impact of potentially erroneous or ambiguous data. Kendall and Gal introduced the use of homoscedastic and heteroscedastic uncertainty to adjust loss functions [15]. Homoscedastic uncertainty is constant across all data points, while heteroscedastic uncertainty varies between instances. This technique enables the model to adapt its learning process based on these uncertainties, enhancing resilience and precision by focusing on confident predictions and reducing the influence of ambiguous ones. As a result, it optimizes the training process across various datasets and improves overall performance [10], [28], [32].

To enhance segmentation accuracy, we implement an uncertainty-aware loss function that combines the following region-based, distribution-based, and pixel-based losses, leveraging their complementary strengths.

1) **Dice Coefficient (DC) loss**: This region-based metric emphasizes the overlap between predicted and ground truth areas, preserving the accuracy of the shape and boundaries of segmented regions.
2) **Cross-Entropy (CE) loss**: This distribution-based loss ensures accurate categorization of individual pixels, improving classification precision.
3) **Focal loss**: This pixel-level loss addresses class imbalance by assigning greater importance to difficult-to-classify instances, enhancing the model's capability to handle complex scenarios [11]–[13].

Given the predicted segmentation and the respective ground-truth mask, let p_i denote the predicted probability and g_i denote the ground truth label. The Dice Similarity Coefficient (DSC) measures the overlap between the predicted segmentation and the ground truth. It is defined as:

$$DSC = \frac{2 \sum_i p_i g_i}{\sum_i p_i + \sum_i g_i} \tag{4}$$

The DSC ranges from 0 to 1, where a DSC of 1 indicates perfect overlap between the prediction and the ground truth, and a DSC of 0 indicates no overlap.

To use this measure in a loss function for training segmentation models, we define the Dice Coefficient (DC) loss as:

$$\mathcal{L}_{DC} = 1 - DSC = 1 - \frac{2\sum_i p_i g_i}{\sum_i p_i + \sum_i g_i} \qquad (5)$$

The DC loss aims to minimize the difference between the predicted segmentation and the ground truth by maximizing the DSC. By minimizing the DC loss, the model is trained to produce segmentations that have a higher overlap with the ground truth, thus improving segmentation accuracy.

The Cross-Entropy (CE) loss is defined following the standard entropy calculation:

$$\mathcal{L}_{CE} = -\sum_i g_i \log(p_i) \qquad (6)$$

The Focal loss is designed to address class imbalance by focusing more on hard-to-classify samples:

$$\mathcal{L}_{focal} = -\sum_i (1 - p_i)^\gamma g_i \log(p_i) \qquad (7)$$

where γ is a focusing parameter default to 2.

We define the uncertainty-aware loss $\mathcal{L}_{UA}$ by combining these losses into an uncertainty-aware framework:

$$\mathcal{L}_{UA} = \sum_{i=1}^{M} \left(\frac{1}{2\sigma_m^2} \mathcal{L}_m + \log(1 + \sigma_m^2) \right) \qquad (8)$$

where M is the number of individual loss components, $\mathcal{L}_m$ represents each individual loss component (such as DC, CE, and focal loss), and σ_m are learnable parameters that adjust the contribution of each loss component based on uncertainty. These learnable parameters are optimized during the training process to minimize the overall loss.

Introducing uncertainty into the loss computation allows the model to dynamically assign weights to each loss component. This approach effectively balances global and local accuracy while mitigating the impact of class imbalance. Consequently, the model becomes more resilient to noisy or ambiguous data, leading to overall improved segmentation performance.

C. Sharpness-Aware Minimization Optimization

Our approach involves utilizing Sharpness-Aware Minimization (SAM) optimization to improve the performance of the U-Mamba model in segmenting cardiac MRI images such as those on the ACDC dataset [1]. SAM optimization flattens the loss landscape, thereby improving the model's generalization potential. Conventional optimization techniques aim to find the lowest points in the loss landscape, but these points can be steep and lead to poor generalization on new data. In contrast, SAM identifiers smoother minima, which are regions in the parameter space where the model's performance is more consistent and less susceptible to disturbances.

The motivation for utilizing SAM stems from its ability to mitigate overfitting, a common challenge in medical image segmentation. Traditional optimization techniques often result in narrow valleys in the loss landscape, leading to suboptimal performance on new data. SAM, however, targets flatter minima, which are associated with better generalization performance. This is particularly beneficial when working with complex and diverse datasets like ACDC, where the variability in cardiac MRI images can lead to overfitting if not properly managed.

The SAM optimization involves an iterative two-step process. First, for every mini-batch, the parameters are adjusted in the direction that maximizes the loss. Next, the model parameters are updated to minimize the highest altered loss. The purpose of this perturbation is to find model parameters that reside in flatter regions of the loss landscape, which typically correspond to better generalization and robustness to small changes in the input data. Specifically, let θ represent the model parameters, $\mathcal{L}$ denote the loss function, and $\mathcal{D}$ denote the training dataset. Let ϵ denote the perturbation, and ρ control the size of the neighborhood around θ. The perturbation ϵ is added to the parameters θ to explore the loss landscape in a local neighborhood defined by the norm constraint $\|\epsilon\|_2 \leq \rho$.

The SAM optimization is mathematically formulated as:

$$\theta^* = \arg\min_\theta \max_{\epsilon:\|\epsilon\|_2 \leq \rho} \mathcal{L}(\theta + \epsilon; \mathcal{D}). \qquad (9)$$

The first step involves finding the perturbation ϵ within the $\|\epsilon\|_2 \leq \rho$ that maximizes the loss. This step identifies the worst-case direction in the local neighborhood of θ, ensuring that the model parameters are adjusted towards regions of the loss landscape that are not steep. In the second step, the model parameters θ are updated to minimize the loss at this worst-case perturbed location, making the parameters robust to perturbations. By iteratively applying these two steps, SAM ensures the optimizer seeks parameter configurations robust to perturbations, improving generalization and performance. This approach helps the model to find flatter minima in the loss landscape with better generalization [6].

IV. EXPERIMENTS

A. Experimental Settings

1) Datasets: The Automated Cardiac Diagnosis Challenge (ACDC) dataset [1] is a widely used benchmark in the field of medical image analysis, particularly for cardiac MRI segmentation tasks. The dataset comprises MRI scans of 150 patients, each with multiple slices, aggregating to a total of 300 images and 2,978 slices. Patients in the dataset are distributed evenly across five groups, deliberated by distinct cardiac pathologies: *normal subjects*, *myocardial infarction*, *dilated cardiomyopathy*, *hypertrophic cardiomyopathy*, and *abnormal right ventricle*. The dataset features short-axis cardiac MRI images, which provide comprehensive heart coverage. Ground truth labels for the left ventricle (LV), right ventricle (RV), and myocardium (MYO) are provided for each image, facilitating the development and evaluation of segmentation algorithms. Variability is present in both image spacing and size across different dimensions within the ACDC dataset.
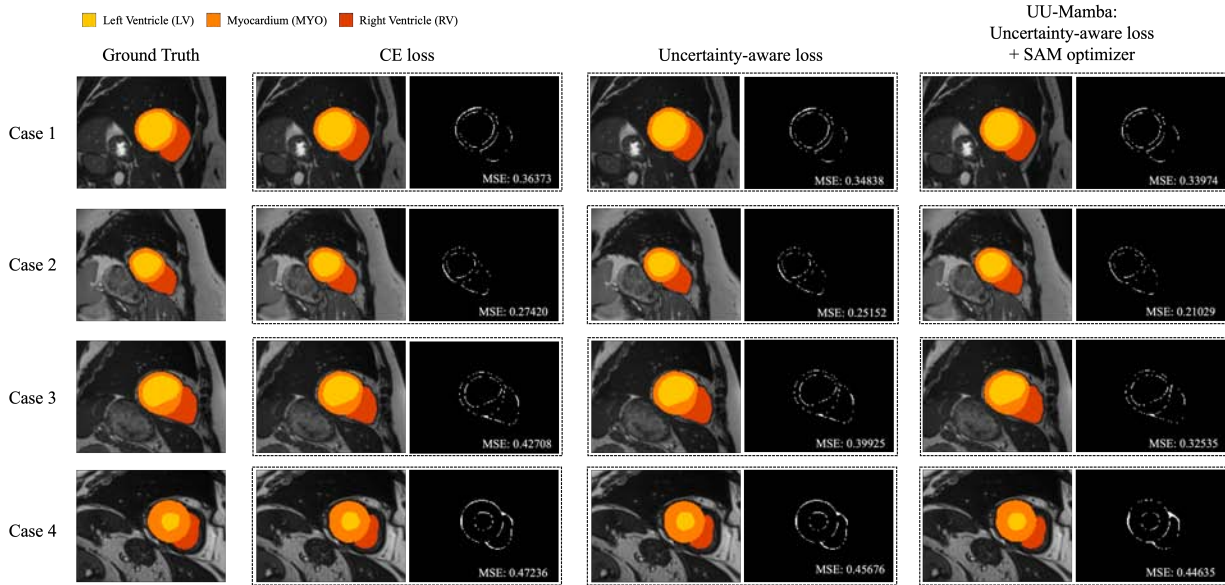
Fig. 3: Segmentation results for various methods on sample images from the ACDC dataset. The Mean Squared Error (MSE) between the output segmentation and the ground truth is shown for each method.

2) Evaluation Metrics: Following the evaluation protocol outlined in [25], we utilize the Dice Similarity Coefficient (DSC) as our primary metric for segmentation evaluation. The DSC quantifies the overlap between the predicted segmentation and the ground truth mask, offering a robust assessment of the model's accuracy in delineating cardiac structures. The DSC is defined in Eq. (4). Additionally, we use the Mean Squared Error (MSE) to measure the average squared difference between the predicted probabilities and the ground truth labels. Let N is the number of testing images, H and W are the height and width of the images, p_{ij}^n is the predicted probability at pixel (i, j) for the n-th image, and g_{ij}^n is the ground truth label at pixel (i, j) for the n-th image. The MSE is calculated as follows:

$$\text{MSE} = \frac{1}{N} \sum_{n=1}^{N} \frac{1}{HW} \sum_{i=1}^{H} \sum_{j=1}^{W} (p_{ij}^n - g_{ij}^n)^2 \qquad (10)$$

MSE provides a complementary evaluation that helps in understanding performance in terms of pixel-wise precision.

3) Implementation Details: The experiments were conducted using the PyTorch framework and two NVIDIA A100 Tensor Core GPUs for training. During training, a patch size of [20, 256, 224] and a batch size of 4 were employed. The network configuration comprises 6 stages, with the number of pooling operations per axis set to [2, 5, 5]. An initial learning rate of 5×10^{-3} was utilized, and the training proceeded for 500 epochs. In the SAM optimization, the hyperparameter ρ controlling the perturbation in Eq. (9) was set to 0.05. The focusing parameter γ in the focal loss in Eq. (7) was set to 2. The parameter M of the uncertainty-aware loss in Eq. (8) was set to 3, incorporating DC loss, CE loss, and focal loss.

B. Experimental Results

We compare UU-Mamba against five leading segmentation models: TransUNet [4], Swin-Unet [2], nnUNet [14], nnFormer [39], and U-Mamba [23]. TransUNet and Swin-Unet are Transformer-based networks, while nnUNet and nnFormer incorporate CNN-based architectures. U-Mamba is a hybrid architecture that integrates elements from both Transformer-based networks and CNN-based architectures.

We conducted a quantitative evaluation of our UU-Mamba against the five other 3D heart area segmentation models on the ACDC dataset, using DSC as the evaluation metric. Fig. 3 shows the segmentation results for the methods in comparison on a few images from the ACDC dataset. Table II reports the DSC scores for each model at three heart regions: the right ventricle (RV), myocardium (Myo), and left ventricle (LV), as well as the average DSC scores of all regions. The best performing scores are marked in bold.

Our UU-Mamba model demonstrates the highest overall performance, achieving an average DSC of 92.79%. The scores for each individual region are as follows: 92.41% for RV, 90.90% for Myo, and 95.04% for LV. Our method surpasses all other models in accurately segmenting different heart components in the right ventricle and myocardium regions, showcasing its flexibility and efficacy. In Table II, the slight decrease in DSC for the left ventricle, compared to other models, is mitigated by the higher scores exhibited in other regions and the highest average DSC.

In comparison, TransUNet achieves an average DSC of 89.71%. While it performs well overall, it shows a lower DSC for Myo compared to other regions. Swin-Unet achieves an average DSC of 90.00%, exhibiting the highest DSC for the LV but a lower DSC for RV, compared to other regions. The nnUNet model achieves an average DSC of 91.61%, demonstrating significant improvement in DSC for Myo and

Method	Avg. DSC % ↑	RV DSC % ↑	Myo DSC % ↑	LV DSC % ↑	Avg. MSE % ↓
CE loss	92.263	91.81	90.31	94.67	0.3843
Uncertainty-aware loss (CE, DC, Focal)	92.602	92.36	90.51	94.94	0.3640
Uncertainty-aware loss + SAM (Ours)	**92.787**	**92.41**	**90.90**	**95.04**	**0.3304**

TABLE I: Ablation study on various configurations with U-Manmba backbone: (1) using only the traditional Cross-Entropy (CE) loss, (2) using the uncertainty-aware Loss without the SAM optimizer, and (3) the proposed UU-Mamba model. The evaluation metric is DSC (%) and MSE.

Method	Average	RV ↑	Myo ↑	LV ↑
TransUNet [4]	89.71	88.86	84.53	95.73
Swin-Unet [2]	90.00	88.55	85.62	**95.83**
nnUNet [14]	91.61	90.24	89.24	95.36
nnFormer [39]	92.06	90.94	89.58	95.65
U-Mamba [23]	92.22	91.83	90.22	94.54
UU-Mamba (Ours)	**92.79**	**92.41**	**90.90**	95.04

TABLE II: Performance comparison of our UU-Mamba with leading medical image segmentation methods on the ACDC dataset. The evaluation metric is DSC (%).

overall performance. The nnFormer model achieves an average DSC of 92.06%, offering the best performance among the four comparing models across all regions, particularly the Myo. The U-Mamba method further improves upon these results with an average DSC of 92.22%, showing notable performance gains in the RV and Myo regions.

The quantitative evaluation highlights the superior performance of our UU-Mamba compared to existing models. With the highest average DSC and exceptional segmentation of the right ventricle and myocardium, our method demonstrates its potential for advancing the accuracy of 3D heart segmentation in medical imaging. The improvements over other models like U-Mamba, nnFormer and nnUNet highlight the effectiveness of our approach in leveraging both global and local features through innovative loss functions and optimization techniques.

C. Ablation Study

We perform ablation study of our UU-Mamba model, by taking out the SAM optimization, and incorporating only the cross-entropy (CE) loss or the uncertainty-aware loss, and compare the respective segmentation performance on the ACDC dataset. Table I summarizes these results. The results demonstrate that incorporating uncertainty-aware loss and the SAM optimization indeed improves segmentation accuracy.

Training using only the CE loss yields 92.263% DSC. In comparison, training using the uncertainty-aware loss (which incorporates the DC loss, CE loss, and focused loss), enhances DSC to 92.602%. This result shows the capability of the uncertainty-aware loss to increase the robustness and accuracy of models by prioritizing confident predictions and minimizing the impact of uncertain ones. It effectively addresses the limitations of using a single loss function by balancing various aspects of the segmentation task.

Incorporating the SAM optimization into the uncertainty-aware loss yields 92.787% DSC, the highest among the examined approaches. SAM improves segmentation accuracy by generating clearer and more accurate borders, guiding the optimization process toward flat minima in the loss landscape. This leads to better generalization across diverse data samples.

The increase in DSC from 92.263% to 92.787% highlights the effectiveness of SAM in enhancing segmentation results and more precisely identifying heart components.

D. Robustness Analysis

We perform experiments to evaluate each method on the Mean Squared Error (MSE) of the DSC scores to assess their robustness quantitatively. The MSE is calculated as shown in Eq. (10). Results are shown in the Table I. These results show that the uncertainty-aware loss reduces the MSE compared to the standard CE loss, reflecting its ability to better address the variability and uncertainty in the data. The inclusion of SAM optimization significantly decreases the MSE, achieving the lowest error value. This reduction in MSE highlights SAM's role in minimizing errors and producing more accurate segmentation maps. Fig. 3 shows the MSE between the output segmentation and the ground truth for each method, providing a visual comparison of the segmentation quality.

V. CONCLUSION

We introduce a new model, UU-Mamba, designed for 3D heart segmentation in cardiac MRI images. This model integrates the U-Mamba architecture with an uncertainty-aware loss function and the SAM optimizer. Our method improves biological image segmentation by improving generalization and boundary accuracy. The uncertainty-aware loss combines region-based, distribution-based, and pixel-based losses to improve segmentation performance by harmonizing tasks and prioritizing confident predictions. Meanwhile, the SAM optimizer guides the model toward flat minima in the loss landscape, enhancing resilience and reducing overfitting, ultimately resulting in more accurate segmentation. Comparative analysis against five leading models demonstrates the superiority of UU-Mamba, achieving a DSC of 92.787%, signifying both accuracy and robustness in segmentation performance.

Future work will encompass the exploration of additional data augmentation techniques, the investigation of alternative uncertainty modeling methods, and the validation of the model on larger and more diverse datasets. Our overarching aim is to advance automated medical imaging by continuously refining and broadening the UU-Mamba approach.

REFERENCES

[1] Olivier Bernard, Antoine Lalande, Caian Zotti, Filip Cervenansky, Xin Yang, Pheng-Ann Heng, Irem Cetin, Karim Lekadir, Oscar Camara, Miguel Angel Gonzalez Ballester, et al. Deep learning techniques for automatic MRI cardiac multi-structures segmentation and diagnosis:

Is the problem solved? *IEEE Transactions on Medical Imaging*, 37(11):2514–2525, 2018.

[2] Hu Cao, Yueyi Wang, Joy Chen, Dong Jiang, Xiaopeng Zhang, Qi Tian, and Manning Wang. Swin-Unet: Unet-like pure transformer for medical image segmentation. *arXiv preprint arXiv:2105.05537*, 2021.

[3] Cheng Chen, Chen Qin, Hongwei Qiu, Giacomo Tarroni, Jin Duan, Wenjia Bai, and Daniel Rueckert. Deep learning for cardiac image segmentation: A review. *Frontiers in Cardiovascular Medicine*, 7:25, 2020.

[4] Jieneng Chen, Yongyi Lu, Qihang Yu, Xiangde Luo, Ehsan Adeli, Yun Wang, Le Lu, Yuyin Zhou, Yefeng Wang, and Alan Yuille. TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2102.04306*, 2021.

[5] Min-Su Cho and Thanh Nguyen. Improving segmentation with advanced loss functions. *Journal of Computational Diagnostics*, 2022.

[6] Pierre Foret, Ariel Kleiner, Hossein Mobahi, and Behnam Neyshabur. Sharpness-aware minimization for efficiently improving generalization. In *Proceedings of the 8th International Conference on Learning Representations (ICLR)*, 2021.

[7] Albert Gu and Tri Dao. Mamba: Linear-time sequence modeling with selective state spaces. *arXiv preprint arXiv:2312.00752*, 2023.

[8] Albert Gu, Isys Johnson, Karan Goel, Khaled Saab, Tri Dao, Atri Rudra, and Christopher Ré. Combining recurrent, convolutional, and continuous-time models with linear state-space layers. *Advances in Neural Information Processing Systems*, 34, 2021.

[9] James Harper and Mark Stone. SAM optimizer: Advancing model generalization. *Journal of Machine Learning Research*, 2022.

[10] Jing Hu, Qinrui Fan, Shu Hu, Siwei Lyu, Xi Wu, and Xin Wang. UMedNeRF: Uncertainty-aware single view volumetric rendering for medical neural radiance fields. *IEEE International Symposium on Biomedical Imaging*, 2024.

[11] Shu Hu, Xin Wang, and Siwei Lyu. Rank-based decomposable losses in machine learning: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 2023.

[12] Shu Hu, Yiming Ying, Siwei Lyu, et al. Learning by minimizing the sum of ranked range. *Advances in Neural Information Processing Systems*, 33:21013–21023, 2020.

[13] Shu Hu, Yiming Ying, Xin Wang, and Siwei Lyu. Sum of ranked range loss for supervised learning. *Journal of Machine Learning Research*, 23(112):1–44, 2022.

[14] Fabian Isensee, Paul F. Jaeger, Simon A. A. Kohl, Jens Petersen, and Klaus H. Maier-Hein. nnU-Net: a self-configuring method for deep learning-based biomedical image segmentation. *Nature Methods*, 18(2):203–211, 2021.

[15] Alex Kendall and Yarin Gal. What uncertainties do we need in bayesian deep learning for computer vision? In *Proceedings of the 31st International Conference on Neural Information Processing Systems*, 2017.

[16] Jong Kim and Sun-Hee Park. Multiloss strategy for medical image segmentation. *Computational Medicine*, 2023.

[17] Li Lin, Irene Amerini, Xin Wang, Shu Hu, et al. Robust CLIP-based detector for exposing diffusion model-generated images. *arXiv preprint arXiv:2404.12908*, 2024.

[18] Li Lin, Xinan He, Yan Ju, Xin Wang, Feng Ding, and Shu Hu. Preserving fairness generalization in deepfake detection. *CVPR*, 2024.

[19] Li Lin, Yamini Sri Krubha, Zhenhuan Yang, Cheng Ren, Xin Wang, and Shu Hu. Robust COVID-19 detection in CT images with CLIP. *MIPR*, 2024.

[20] Li Lin, Sarah Papabathini, Xin Wang, and Shu Hu. Robust light-weight facial affective behavior recognition with CLIP. *MIPR*, 2024.

[21] Geert Litjens, Thijs Kooi, Babak Ehteshami Bejnordi, Arnaud Arindra Adiyoso Setio, Francesco Ciompi, Mohsen Ghafoorian, Jeroen AWM van der Laak, Bram van Ginneken, and Clara I Sánchez. A survey on deep learning in medical image analysis. *Medical image analysis*, 42:60–88, 2017.

[22] Ze Liu, Yutong Lin, Yue Cao, Han Hu, Yixuan Wei, Zheng Zhang, Stephen Lin, and Baining Guo. Swin transformer: Hierarchical vision transformer using shifted windows. *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 10012–10022, 2021.

[23] Jun Ma, Feifei Li, and Bo Wang. U-Mamba: Enhancing long-range dependency for biomedical image segmentation. *arXiv preprint arXiv:2401.04722*, 2024.

[24] Andreas K Maier, Christoph Syben, Tobias Lasser, and Christian Riess. A gentle introduction to deep learning in medical image processing. *Zeitschrift für Medizinische Physik*, 29(2):86–101, 2019.

[25] Lena Maier-Hein, Annika Reinke, Philipp Godau, Mahdieh Diba Tizabi, Florian Büttner, et al. Metrics reloaded: Pitfalls and recommendations for image analysis validation. *arXiv preprint arXiv:2206.01653*, 2022.

[26] Andrew Murray et al. Deep learning in medical imaging: Overfitting issues. *Medical Image Analysis*, 2022.

[27] Judy Pang, Eline Smit, Janne von Spiczak, Christakis P Loizou, Sven Plein, Imran Ali, Anilkumar Arujuna, Wenjia Bai, Mikel Nunez-Garcia, Paul M Matthews, et al. Automated myocardium segmentation in cardiac magnetic resonance images using fully convolutional networks. *Medical Physics*, 46(11):5168–5180, 2019.

[28] Yicui Peng, Hao Chen, Chingsheng Lin, Guo Huang, Jinrong Hu, Hui Guo, Bin Kong, Shu Hu, Xi Wu, and Xin Wang. Uncertainty-Aware explainable recommendation with large language models. *IJCNN*, 2024.

[29] Caroline Petitjean and Jean-Nicolas Dacher. A review of segmentation methods in short axis cardiac mr images. *Medical image analysis*, 15(2):169–184, 2011.

[30] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. U-Net: Convolutional networks for biomedical image segmentation. *International Conference on Medical image computing and computer-assisted intervention*, pages 234–241, 2015.

[31] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. Attention is all you need. *Advances in neural information processing systems*, 30, 2017.

[32] Xin Wang, Shu Hu, Heng Fan, Hongtu Zhu, and Xin Li. Neural radiance fields in medical imaging: Challenges and next steps. *arXiv preprint arXiv:2402.17797*, 2024.

[33] Xin Wang and Hongtu Zhu. Artificial intelligence in image-based cardiovascular disease analysis: A comprehensive survey and future outlook. *arXiv:2402.03394*, 2024.

[34] Ziyang Wang and Chao Ma. Weak-Mamba-UNet: Visual mamba makes CNN and ViT work better for scribble-based medical image segmentation. *arXiv preprint arXiv:2402.10887*, 2024.

[35] Ziyang Wang, Jian-Qing Zheng, Yichi Zhang, Ge Cui, and Lei Li. Mamba-UNet: UNet-Like pure visual mamba for medical image segmentation. *arXiv preprint arXiv:2402.05079*, 2024.

[36] Xujiang Zhao, Feng Chen, Shu Hu, and Jin-Hee Cho. Uncertainty aware semi-supervised learning on graph data. *Advances in Neural Information Processing Systems*, 33:12827–12836, 2020.

[37] Xujiang Zhao, Shu Hu, Jin-Hee Cho, and Feng Chen. Uncertainty-based decision making using deep reinforcement learning. In *2019 22th International Conference on Information Fusion (FUSION)*, pages 1–8. IEEE, 2019.

[38] H Zheng, Z Lin, L Wang, and V S Sheng. Deep learning-based segmentation using an architectural prototype with multi-scale convolutional network. *Journal of Digital Imaging*, 33(3):581–597, 2020.

[39] Huiyu Zhou, Yuyin Yang, Kai Chen, Jie Huang, Yucheng Pan, Yingwei Wang, Xin Xie, Liansheng Yang, and Huazhong Shu. nnFormer: Volumetric medical image segmentation via a 3D transformer. *arXiv preprint arXiv:2109.03201*, 2021.