

# 000 001 002 003 004 005 006 007 PI-CONTROLLED UNCERTAINTY FOR STEADY-STATE 008 ERROR ELIMINATION IN ULTRASOUND IMAGE SEG- 009 MENTATION 010 011 012

013 **Anonymous authors**  
014 Paper under double-blind review  
015  
016  
017  
018  
019  
020  
021  
022  
023  
024  
025  
026  
027  
028  
029  
030

## ABSTRACT

031 Accurate segmentation of anatomical structures from medical ultrasound images  
032 is essential for reliable diagnosis, yet conventional training losses often leave per-  
033 sistent steady-state errors, especially along ambiguous boundaries. These losses  
034 act as control variables generated by a proportional controller, since they respond  
035 only to instantaneous discrepancies and lack the memory required to correct long-  
036 term deviations. To overcome this limitation, we rethink segmentation training as  
037 a closed-loop control system where uncertainty acts as the control variable. Build-  
038 ing on this perspective, we introduce a proportional-integral (PI) control mecha-  
039 nism that integrates both present and historical error signals into the optimization  
040 process, enabling the model to systematically eliminate steady-state errors and  
041 deliver sharper, more reliable boundary predictions. Unlike existing uncertainty-  
042 based approaches that rely solely on fixed loss terms, our method provides a prin-  
043 cipled mechanism to incorporate dynamic feedback into training. The framework  
044 is model-agnostic and introduces no additional inference overhead, making it di-  
045 rectly compatible with real-time segmentation backbones. Extensive experiments  
046 on clinical medical ultrasound datasets demonstrate consistent improvements over  
047 state-of-the-art baselines. These results confirm that our framework offers an ef-  
048 fective solution for eliminating steady-state errors in medical ultrasound image  
049 segmentation under challenging conditions. Our code is available at <https://anonymous.4open.science/r/PI-control-uncertainty-B82C>.  
050  
051  
052  
053

## 1 INTRODUCTION

035 Accurate semantic segmentation of anatomical structures, particularly from ultrasound imaging,  
036 provides crucial quantitative support for clinical decision-making, including disease diagnosis, treat-  
037 ment planning, and prognostic monitoring (Azad et al., 2024; Tiwari et al., 2025; Zhang et al., 2020).  
038 A large number of learnable methods have achieved significant results in recent years (Wang et al.,  
039 2021; Li et al., 2025; Hu et al., 2025). Nevertheless, due to the scattering and attenuation char-  
040 acteristics of ultrasound waves in tissues, medical ultrasound images suffer from more severe low  
041 contrast, speckle noise, and ambiguous tissue boundaries compared to CT and MRI (Lee et al., 2022;  
042 Zamzmi et al., 2021; Gowda & Clifton, 2025). These image quality issues lead to higher uncertainty  
043 in segmentation tasks, making it crucial to learn uncertainty and leverage uncertainty to improve  
044 segmentation performance.

045 Methods for quantifying predictive uncertainty have been explored to enhance model reliability  
046 (Zhou et al., 2024; Mucsányi et al., 2024; Abdar et al., 2023; Judge et al., 2023; Zhou et al., 2025).  
047 They typically rely on static loss functions to learn uncertainty, for instance, by predicting the par-  
048 ameters of a probability distribution (Liu et al., 2022; He et al., 2019; Dong et al., 2025; Duenias et al.,  
049 2025). Unfortunately, this paradigm leads to a critical failure mode: the emergence of a steady-state  
050 segmentation mask prediction error, where a persistent discrepancy between the model’s prediction  
051 and the ground truth remains, even after training converges. We identify that this limitation arises  
052 because conventional loss functions are analogous to simple proportional (P) controllers in classical  
053 control theory (Franklin et al., 2010; Åström & Hägglund, 2006). They address only the instant-  
054 aneous error of each training step without memory of past failures, rendering them incapable of  
055 eliminating systematic, steady-state errors.

To address the above problems, We rethink existing uncertainty learning mechanisms through a closed-loop control framework. We analyze and reveal key insights into why P-control-like loss functions fail and how a more sophisticated control strategy can succeed. Inspired by these insights, we design a novel framework where the predicted segmentation mask is the controlled variable, driven to match the ground truth setpoint. Crucially, we define the uncertainty itself as the control variable, which is actively manipulated by our proposed controller to achieve precise error correction. Although some works, like PIDNet (Xu et al., 2023), incorporate PID control into network architecture to reduce feature fusion overshoot, their loss functions remains static. By training with static loss functions that capture only instantaneous error, these approaches inherently lack memory of historical information and cannot address the fundamental issue of steady-state error.

Furthermore, we introduce a Proportional-Integral (PI) controller as the core of our framework. Unlike P-control-like static losses, our PI controller leverages a crucial integral term to accumulate historical errors for persistently challenging regions. In this way, our controller can generate an adaptive and escalating correction specifically targeted at stubborn, long-term deviations. Therefore, our framework can effectively eliminate the steady-state error that plagues conventional methods, ensuring the model converges to a more accurate solution. In addition, our method can be seamlessly integrated into various segmentation backbones without extra inference cost.

Our main contributions are as follows: ① We rethink segmentation optimization as a control problem, applying uncertainty control directly to learning dynamics rather than conventional architectural design. This framework uniquely defines the segmentation mask as the controlled variable and the uncertainty as the control variable, offering a new lens for resolving persistent training errors. ② We design and implement a PI controller that integrates historical error information into the optimization process. This controller generates a dynamic, escalating corrective signal to precisely and effectively eliminate the steady-state errors that traditional methods fail to address. ③ We evaluate our method on two distinct ultrasound datasets with different challenges: MEIS with blurred boundaries and TN3K with variable nodule characteristics, demonstrating significant performance improvements over state-of-the-art approaches.

## 2 RELATED WORKS

Foundational models like YOLACT (Bolya et al., 2019) and its successor for ultrasound, RAMEM (Tseng et al., 2024), provide efficient architectural baselines. Performance is further pushed by enhancing network components, such as introducing explicit boundary operators (Lin et al., 2023), transformer-based designs (Pei et al., 2022), advanced decoders (Wazir & Kim, 2025), or specialized loss functions to handle data imbalance (Xu et al., 2025). Concurrently, uncertainty quantification is often addressed through Bayesian methods (Gal & Ghahramani, 2016) or Deep Ensembles (Lakshminarayanan et al., 2017). Control theory is also applied, primarily to network architecture design (Girum et al., 2021; Xu et al., 2023) or as an inspiration for optimizers An et al. (2018). Further details on related works are provided in Appendix A.4.

However, these approaches share a critical limitation: they are all optimized using static losses that lack a dynamic mechanism to correct for the steady-state errors that persist during training. Integral control is a mechanism specifically designed to eliminate persistent errors by accumulating historical information. However, while some methods apply control theory to network structure, its potential remains underexplored in the context of the optimization process itself. Our work identifies this fundamental gap and proposes a PI-controlled uncertainty mechanism to directly address stubborn boundary inaccuracies by rethinking the training dynamics.

## 3 METHOD

### 3.1 PROBLEM FORMULATION

Different image segmentation methods use various neural network outputs to represent uncertainty, learning from the discrepancy between model predictions ( $M_{\text{pred}}$ ) and ground truth ( $M_{\text{gt}}$ ). At each step, the loss function calculates the error  $e(t) = |M_{\text{gt}} - M_{\text{pred}}|$ , and the optimizer updates model parameters using the gradient to minimize this error. From a control theory perspective, this resembles a feedback control system where the corrective signal is proportional to the instantaneous error

108  $e(t)$ , similar to a Proportional (P) controller  $u(t) = K_p \cdot e(t)$ , where  $K_p$  is the proportional gain.  
 109 The comparison between conventional learning processes and our PI-control framework is shown in  
 110 Fig. 1.

111 The fundamental deficiency of this P-control-like mechanism is its inherent inability to eliminate  
 112 systematic steady-state error. To illustrate this limitation, consider medical image segmentation  
 113 with ambiguous tissue boundaries. After training, a model’s prediction for a ambiguous pixel may  
 114 converge to 0.55 probability despite a ground truth of 1, yielding a constant error  $e(t) = 0.45$ . This  
 115 persistent error has negligible influence on the loss function minimization compared to the much  
 116 larger number of easily-classified pixels. Consequently, the optimizer provides insufficient updates  
 117 to resolve this deviation, resulting in steady-state segmentation mask prediction error. Standard  
 118 training paradigms, behaving like P-controllers, cannot eliminate such persistent errors due to their  
 119 reliance solely on instantaneous error signals without error accumulation mechanisms.

120

## 121 3.2 OVERALL FRAMEWORK

123 To resolve the steady-state error in segmentation tasks, which arises from conventional loss func-  
 124 tions, we reframe the training process of deep neural networks as a closed-loop feedback control  
 125 system, as shown in Fig. 1. The core of this framework is a PI controller that acts as a dynamic con-  
 126 struction mechanism, generating the final control variable by integrating a modulated error signal  
 127 with the model’s predictive uncertainty. It is composed of the following key components.

128

### 129 3.2.1 CONTROLLED OBJECT AND FEEDBACK

131 Within our control system, the controlled object is the model itself. Our work is built upon an exist-  
 132 ing framework, RAMEM, which is based on the real-time instance segmentation model, YOLACT,  
 133 demonstrates impressive performance in M-mode echocardiography. A detailed description of the  
 134 RAMEM is provided in Appendix A.2. It is composed of three main components: a backbone, a  
 135 Feature Pyramid Network (FPN), and the prediction heads. The backbone utilizes UPANet for fea-  
 136 ture extraction. The prediction heads consist of two primary sub-networks: the Prediction Module  
 137 and the ProtoNet.

138 The Prediction Module predict a set of linear combination coefficients for each potential object  
 139 instance. Concurrently, the ProtoNet generates a series of prototype masks. The final segmentation  
 140 mask for each instance, denoted as  $M_{\text{pred}}$ , is then produced by linearly combining these prototype  
 141 masks with the corresponding instance’s coefficients.  $M_{\text{pred}}$  is the system’s Controlled Variable and  
 142 is compared with the Ground Truth Mask  $M_{\text{gt}}$ , which acts as the Setpoint. This comparison yields  
 143 an error signal,  $e(t) = |M_{\text{gt}} - M_{\text{pred}}|$ , quantifying the model’s performance deviation and is fed back  
 144 into our PI controller.

145

### 146 3.2.2 PI-CONTROLLED UNCERTAINTY

148 The raw error signal  $e(t)$  is fed into our designed PI controller. Unlike a classic PI controller that  
 149 only processes an error signal, our proposed controller is an integrated module designed to generate  
 150 the final control variable  $L_{\text{uncertainty}}$  by simultaneously considering historical error dynamics and the  
 151 model’s predictive uncertainty. It consists of four components: the proportional term, ensuring a  
 152 rapid response to current changes; the integral term, dedicated to eliminating long-term systemic  
 153 biases; an Uncertainty Estimation Module, which predicts the uncertainty scale parameter  $b$  for each  
 154 instance; and a Laplace distribution to output the control variable  $L_{\text{uncertainty}}$ .

155 The proportional term focuses on the current instantaneous error. To make the controller more  
 156 efficient, instead of using a global average error, we focus its attention on the model’s most uncertain  
 157 hard pixels, which is analogous to a form of online hard example mining. We first compute the per-  
 158 pixel predictive uncertainty (measured as the variance of a Bernoulli distribution),  $U = 4 \cdot M_{\text{pred}} \cdot$   
 159  $(1 - M_{\text{pred}})$ , which is maximized when the predicted probability  $M_{\text{pred}}$  is close to 0.5. We then  
 160 select the top- $k$  pixels with the highest uncertainty to form a hard set  $\mathcal{H}$ . This strategy prevents  
 161 the controller’s decision from being diluted by a large number of easily segmented background  
 pixels, thereby concentrating its efforts on solving the real challenges. The proportional term  $P(t)$

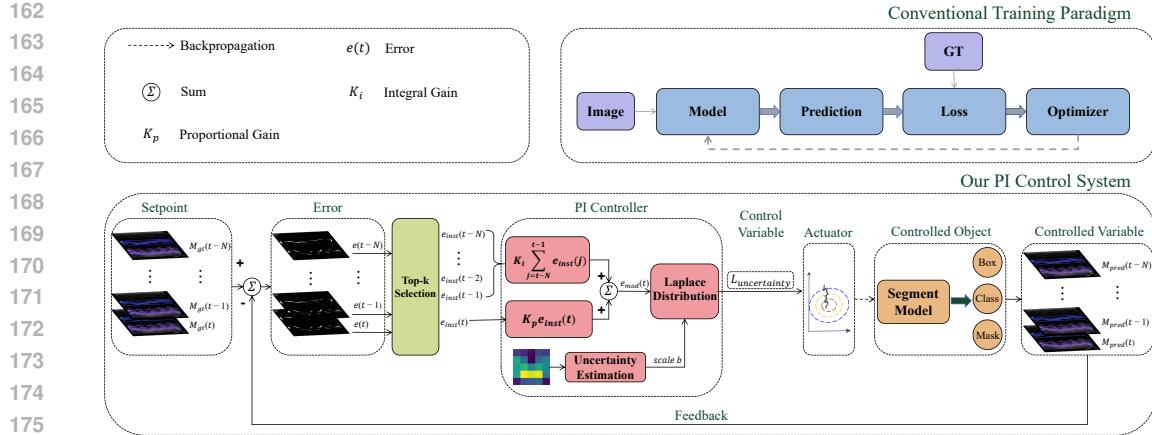


Figure 1: **Overview of the Proposed PI Control Framework.** This figure contrasts the conventional training paradigm with our control system approach. **Top:** The conventional training process, where the loss is driven only by the instantaneous error. **Bottom:** Our proposed PI Control System. This closed-loop framework introduces a novel PI controller as its core. An error signal, calculated as the difference between the ground truth (Setpoint) and the model’s predictions over time, is processed by the controller. The controller internally performs (1) top-k selection on hard pixels to compute the proportional error term, (2) summation of historical errors for the integral term, and (3) learns an uncertainty parameter  $b$  via a dedicated estimation module. These components are then synthesized using a Laplace distribution to formulate the final  $L_{\text{uncertainty}}$  (control variable). This signal is used by the optimizer (Actuator) to update the segmentation model (Controlled Object), effectively minimizing persistent steady-state errors.

is defined as the mean error over these hard pixels

$$P(t) = e_{\text{inst}}(t) = \frac{1}{k} \sum_{p \in \mathcal{H}} |M_{\text{gt}}^{(p)} - M_{\text{pred}}^{(p)}|, \quad (1)$$

where  $e_{\text{inst}}(t)$  represents the instantaneous instance error for the current  $t$ , calculated as the average discrepancy over the set of  $k$  most uncertain pixel indices  $p$  within the hard set  $\mathcal{H}$ .

The integral term is the key to eliminating steady-state error, as it introduces memory into the system. In classical continuous control theory, the integral action is defined by the integral of the error signal over time:  $I(t) = \int_0^t e(\tau) d\tau$ . This mechanism ensures that any persistent, non-zero error, no matter how small, will eventually produce a significant corrective action. Since the training of a deep neural network is a discrete-time process that proceeds in epochs, we approximate this continuous integral with its discrete counterpart: a summation of errors over a finite time window. This provides the system with a practical “memory” of recent performance. Specifically, for each instance in the training set, we maintain a deque  $D$  of length  $N$  to store its average instance error,  $e_{\text{inst}}(t), e_{\text{inst}}(t-1), \dots, e_{\text{inst}}(t-N)$ , over the past  $N$  training epochs. The hyperparameter  $N$  determines the system’s memory length; a smaller value makes the controller more responsive to recent trends, while a larger value provides a more stable estimate of long-term systemic bias. The integral term  $I(t)$  is defined as the sum of these historical errors

$$I(t) = \sum_{j=t-N}^{t-1} e_{\text{inst}}(j), \quad (2)$$

where  $t$  represents the  $t$ -th epoch.

The final modulated error signal  $e_{\text{mod}}(t)$  is the weighted sum of the two terms  $e_{\text{mod}}(t) = K_p \cdot P(t) + K_i \cdot I(t)$ , where  $K_p$  and  $K_i$  are the proportional and integral gain hyperparameters, respectively. When an instance’s error persists,  $P(t)$  may not change significantly, but  $I(t)$  will accumulate, causing  $e_{\text{mod}}(t)$  to escalate and form a targeted, ever-increasing corrective signal.

The controller’s output,  $e_{\text{mod}}(t)$ , along with an uncertainty scale parameter  $b$  predicted by the Uncertainty Estimation Module, is used to dynamically construct a supplementary loss term, which we call the PI-Controlled Uncertainty Loss ( $L_{\text{uncertainty}}$ ). The Uncertainty Estimation Module consists of a 3x3 convolutional layer followed by a softplus activation function. The convolutional layer learns to extract features relevant to uncertainty from the shared feature map, while the softplus activation ensures that its output, the scale parameter  $b$ , is strictly positive, as mathematically required for Laplace distributions.

In our framework, this  $L_{\text{uncertainty}}$  is defined as the system’s control variable,  $u(t)$ . It is dynamically constructed by modeling the PI-modulated error signal  $e_{\text{mod}}(t)$  with a Laplace distribution. The choice of the Laplace distribution over the more common Gaussian distribution is motivated by its heavier tails and sharper peak, which make it more robust to handling the potential outliers that can be generated by the PI controller’s integral term, especially when  $e_{\text{mod}}(t)$  becomes large. The general probability density function (PDF) of a Laplace distribution is given by

$$f(x | \mu, b) = \frac{1}{2b} \exp\left(-\frac{|x - \mu|}{b}\right), \quad (3)$$

where  $\mu$  is the location parameter (mean) and  $b$  is the scale parameter, which  $2b^2$  the variance of the distribution. We set the mean  $\mu$  to zero, as our goal is to drive the error to zero, and we treat our modulated error  $e_{\text{mod}}(t)$  as a sample  $x$  drawn from this distribution.

For a batch containing  $M$  positive instances, we assume the modulated errors for each instance are independent. The model is trained to predict a unique scale parameter  $b^{(i)}$  for each instance  $i$ , which quantifies the uncertainty. The joint PDF for observing the set of modulated errors  $\{e_{\text{mod}}^{(i)}\}_{i=1}^M$  is the product of the individual Laplace PDFs. With  $\mu = 0$ , this becomes

$$f(\{e_{\text{mod}}^{(i)}\}_{i=1}^M, \{b^{(i)}\}_{i=1}^M) = \prod_{i=1}^M \frac{1}{2b^{(i)}} \exp\left(-\frac{|e_{\text{mod}}^{(i)}(t)|}{b^{(i)}}\right). \quad (4)$$

To train this probabilistic model, we maximize the likelihood of the observed data, which is equivalent to minimizing the Negative Log-Likelihood (NLL). Taking the negative logarithm of the joint PDF gives us our final loss formulation for  $L_{\text{uncertainty}}$

$$L_{\text{uncertainty}} = -\log(f) = \sum_{i=1}^M \left( \log(2b^{(i)}) + \frac{|e_{\text{mod}}^{(i)}(t)|}{b^{(i)}} \right). \quad (5)$$

This uncertainty-based modeling mechanism is ingeniously reflected in the loss function’s design. The scale parameter  $b^{(i)}$  serves as a direct indicator of the model’s confidence. This relationship is mathematically grounded, as the entropy of the Laplace distribution is  $1 + \log(2b)$ . A larger scale parameter corresponds to higher entropy and thus lower model confidence. This creates a dynamic trade-off during training. On one hand, a larger  $b^{(i)}$  value reduces the penalty from the position deviation term  $|e_{\text{mod}}^{(i)}(t)|/b^{(i)}$ , providing more lenient error tolerance for predictions with high, persistent uncertainty. On the other hand, the  $\log(2b^{(i)})$  term acts as a regularizer that penalizes excessive uncertainty, preventing the model from simply increasing  $b$  to ignore all errors. The pseudo-code of our PI-Controlled Uncertainty Learning mechanism is presented in Algorithm 1 in Appendix A.3.

The  $L_{\text{uncertainty}}$  is backpropagated through an optimizer, which can be viewed as the Actuator. The optimizer calculates gradients and updates the weights of the model. This update alters the model’s output  $M_{\text{pred}}$ , which in turn generates a new error signal in the next iteration, thus forming a closed control loop.

### 3.3 OVERALL LOSS FUNCTION

The final training loss function is a weighted sum of the original model RAMEM’s losses and our newly introduced PI-controlled uncertainty loss. We intentionally retain the static mask loss ( $L_{\text{mask}}$ ) alongside our new loss term, as they perform distinct and complementary roles. The standard static mask loss, a combination of BCE and Dice loss, is crucial for efficiently learning the features of the segmentation target, especially in the early stages of training.

270 In contrast, our PI-controlled uncertainty loss,  $L_{\text{uncertainty}}$ , acts as a segmentation rectifier. Its primary  
 271 role is to deal with the small subset of hard pixels that cause the steady-state error, which the static  
 272 mask loss is ill-equipped to handle. To ensure training stability, we adopt a two-stage strategy where  
 273 the  $L_{\text{uncertainty}}$  term is introduced only after the preliminary training with mask loss reaches an initial  
 274 feasible state. This allows the model to first learn the basic features of the task under the guidance  
 275 of the mask loss before our PI controller begins its fine-grained correction process.

276 The overall loss function for end-to-end training is thus formulated as  
 277

$$278 L_{\text{total}} = \lambda_{\text{mask}} \cdot L_{\text{mask}} + \lambda_{\text{uncertainty}} \cdot \frac{1}{M} L_{\text{uncertainty}} + \lambda_{\text{cls}} \cdot L_{\text{cls}} + \lambda_{\text{bbox}} \cdot L_{\text{bbox}}, \quad (6)$$

279 where  $M$  is the number of positive instances in the batch, and  $\lambda$  terms are the respective loss weights.  
 280 It is worth noting that our task is formulated as instance-level lesion segmentation rather than purely  
 281 semantic segmentation. Since our framework is built upon the RAMEM instance segmentation  
 282 pipeline, we retain the original classification and bounding box regression losses ( $L_{\text{cls}}$  and  $L_{\text{bbox}}$ )  
 283 from the baseline. Our PI-controlled uncertainty loss is only attached to the mask branch and does  
 284 not alter the detection components, ensuring that the underlying instance segmentation architecture  
 285 and inference procedure remain unchanged. By integrating our PI-controlled loss as an additional  
 286 component, our method acts as a rectifier to the standard training process. This dynamic adjustment  
 287 also provides valuable guidance for feature learning, directing the model to focus more on reliable  
 288 features while being cautious with uncertain ones. This enhancement is achieved without adding  
 289 any computational cost at inference time, as the PI controller and uncertainty loss are only active  
 290 during training.

## 291 4 EXPERIMENT

292 We conduct experiments on two ultrasound datasets to validate our proposed PI-controlled training  
 293 framework against multiple state-of-the-art methods. Our experimental evaluation is designed to  
 294 answer the following key questions. Q1: Does the proposed PI-controlled framework outperform  
 295 existing state-of-the-art methods in medical ultrasound segmentation? Q2: Can the proposed PI  
 296 controller reduce segmentation errors on ambiguous boundaries compared to conventional methods?  
 297 Q3: What are the individual contributions of the Proportional (P) and Integral (I) components of our  
 298 controller? Do both these two components contribute to ultrasound image segmentation? Q4: Is the  
 299 proposed framework applicable to different segmentation tasks and anatomical structures?  
 300

### 301 4.1 EXPERIMENT SETUP

302 We validate our framework on two distinct and publicly available ultrasound datasets: MEIS, which  
 303 presents challenges with blurred boundaries in M-mode echocardiography, and TN3K, which in-  
 304 volves significant variability in nodule characteristics in B-mode images. For fair comparison, all  
 305 methods are trained using the SGD optimizer. Our PI controller is configured with a proportional  
 306 gain  $K_p = 2.0$ , an integral gain  $K_i = 0.1$ . We evaluate performance using a suite of standard met-  
 307 rics, including Precision, Recall, Dice Similarity Coefficient (DSC), HD95 and COCO-style mean  
 308 Average Precision for masks (Mask-mAP) and boxes (Box-mAP). Comprehensive details regarding  
 309 data preprocessing, augmentation strategies, learning rate schedules, and evaluation protocols are  
 310 provided in Appendix A.5, A.6, and A.7.

### 311 4.2 COMPARISON WITH STATE-OF-THE-ART METHODS(Q1,Q4)

312 We conduct a quantitative comparison of our method against several advanced instance segmentation  
 313 models, including OSFOMER, CTO, YOLACT, BALANCE, MCADS, and our baseline, RAMEM,  
 314 on both the MEIS and TN3K datasets.

315 **Results on MEIS Dataset.** As shown in Table 1, our method achieves the best performance across  
 316 most key metrics. It obtains the highest Precision (87.63%), DSC (87.55%), and the smallest bound-  
 317 ary error measured by HD95 (13.59). While OSFOMER records a marginally higher recall (88.26%  
 318 vs. our 87.97%), it achieves significantly lower precision (85.77% vs. our 87.63%). In contrast,  
 319 our method strikes a superior balance, leading to more reliable and accurate segmentation. Instance-  
 320 level results in Table 3 further show that our method surpasses all baselines in Mask-mAP, Box-  
 321 mAP, and Avg-mAP. This quantitative superiority is visually corroborated by the qualitative results

324  
 325 Table 1: The results on the MEIS dataset. The best results are highlighted in **bold**, and the second  
 326 best are underlined.

Methods	Recall	Precision	DSC	HD95
OSFOMER(ECCV’22)	<b>88.26±1.81</b>	85.77±2.08	86.69±1.91	17.39±8.68
CTO(IPMI’23)	87.53±2.05	86.05±1.79	86.48±1.92	17.04±7.51
YOLACT(ICCV’19)	87.34±1.73	87.14±2.16	86.92±1.90	<u>13.85±5.37</u>
BALANCE(AAAI’25)	87.48±1.35	87.02±1.99	<u>86.94±1.67</u>	15.48±6.41
MCADS(CVPR’25)	87.40±2.25	86.93±1.83	<u>86.79±2.06</u>	16.79±9.30
RAMEM(J-BHI’24)	87.31±2.33	<u>87.25±2.01</u>	86.91±1.84	14.17±5.17
<b>Ours</b>	<u>87.97±2.13</u>	<b>87.63±1.97</b>	<b>87.55±1.69</b>	<b>13.59±5.72</b>

335  
 336  
 337 Table 2: The results on the TN3K dataset. The best results are highlighted in **bold**, and the second  
 338 best are underlined.

Methods	Recall	Precision	DSC	HD95
OSFOMER(ECCV’22)	<b>91.24±0.52</b>	76.86±0.69	79.98±0.54	63.36±0.73
CTO(IPMI’23)	<u>85.89±0.37</u>	83.09±0.61	82.29±0.51	<u>42.86±2.02</u>
YOLACT(ICCV’19)	84.43±0.38	80.88±0.57	<u>79.88±0.25</u>	<u>46.29±2.26</u>
BALANCE(AAAI’25)	84.12±0.66	81.54±0.43	80.45±0.49	54.02±1.91
MCADS(CVPR’25)	84.03±0.47	82.22±0.37	81.17±0.42	49.17±3.08
RAMEM(J-BHI’24)	83.97±0.98	<u>84.68±0.82</u>	81.91±0.64	44.79±2.23
<b>Ours</b>	84.52±0.70	<b>85.51±0.57</b>	<b>82.55±0.31</b>	<b>42.40±1.63</b>

348  
 349  
 350 in Fig. 2. For the challenging cases, baseline methods like RAMEM and YOLACT often produce  
 351 fragmented or discontinuous masks. Our model, however, consistently generates predictions that  
 352 are both complete and precisely aligned with the ground truth boundaries, a direct demonstration of  
 353 the PI controller’s effectiveness in resolving local ambiguities.

354 **Results on TN3K Dataset.** To validate the adaptability of our framework (Q4), we evaluate it on the  
 355 TN3K thyroid nodule dataset. The results are detailed in Table 2 and 3. Our approach achieves the  
 356 top performance in Precision (85.51%), DSC (82.55%), HD95 (42.40), and all mAP-based metrics,  
 357 culminating in the highest Avg-mAP of 49.20%. This quantitative superiority is visually corroborated  
 358 by the qualitative results presented in Fig. 3. For challenging cases involving low contrast,  
 359 irregular shapes, or multiple nodules, baseline methods often yield segmentations with substantial  
 360 boundary leakage or missed detections. In contrast, our model consistently generates more complete  
 361 masks that more accurately delineate the ground truth contours, highlighting the PI controller’s ef-  
 362 ficacy in adapting to diverse nodule morphologies. The performance on this distinct anatomical  
 363 structure and imaging modality highlights the versatility of our control approach, suggesting its  
 364 potential to resolve persistent errors across different medical ultrasound segmentation tasks. For  
 365 additional qualitative comparisons, please refer to Appendix A.8.1 and A.8.2.

#### 366 4.3 ABLATION STUDY(Q3)

367 To deconstruct the contributions of our core components and answer Q3, we conducted an ablation  
 368 study on both the MEIS and TN3K datasets, as shown in Fig. 4. We compared our full method (Ours)  
 369 against two variants: one removing the integral term of the PI controller (Ours-PI) and another  
 370 removing the PI controller and uncertainty loss (Ours-PI-uncertainty).

371 The results demonstrate that both components are crucial for optimal performance, working in syn-  
 372 ergy to address different challenges. On the MEIS dataset, with its characteristic blurred bound-  
 373 aries, the integral term showed a more pronounced impact, underscoring its strength in correcting  
 374 systematic, steady-state errors. On the TN3K dataset, which features high nodule variability, the  
 375 contributions of both the PI control and the uncertainty were more balanced and closely matched.  
 376 This highlights their collaborative effectiveness in handling tasks with high data variance.

378

379  
380  
Table 3: Mask-mAP, Box-mAP and Avg-mAP on MEIS and TN3K datasets. The best results are  
highlighted in **bold**, and the second best are underlined.

Methods	MEIS			TN3K		
	Mask-mAP	Box-mAP	Avg-mAP	Mask-mAP	Box-mAP	Avg-mAP
OSFOEMER	54.58 $\pm$ 6.39	67.73 $\pm$ 5.27	61.16 $\pm$ 5.15	48.81 $\pm$ 0.71	44.24 $\pm$ 0.66	46.53 $\pm$ 0.69
YOLACT	55.08 $\pm$ 5.65	66.98 $\pm$ 4.99	61.03 $\pm$ 5.10	49.41 $\pm$ 0.85	44.54 $\pm$ 0.79	46.97 $\pm$ 0.79
RAMEM	55.67 $\pm$ 5.41	68.83 $\pm$ 4.67	62.25 $\pm$ 4.83	50.80 $\pm$ 0.47	46.51 $\pm$ 0.60	48.66 $\pm$ 0.52
<b>Ours</b>	<b>57.03<math>\pm</math>3.87</b>	<b>72.67<math>\pm</math>3.24</b>	<b>64.85<math>\pm</math>2.96</b>	<b>51.13<math>\pm</math>0.34</b>	<b>47.27<math>\pm</math>0.48</b>	<b>49.20<math>\pm</math>0.42</b>

381

382

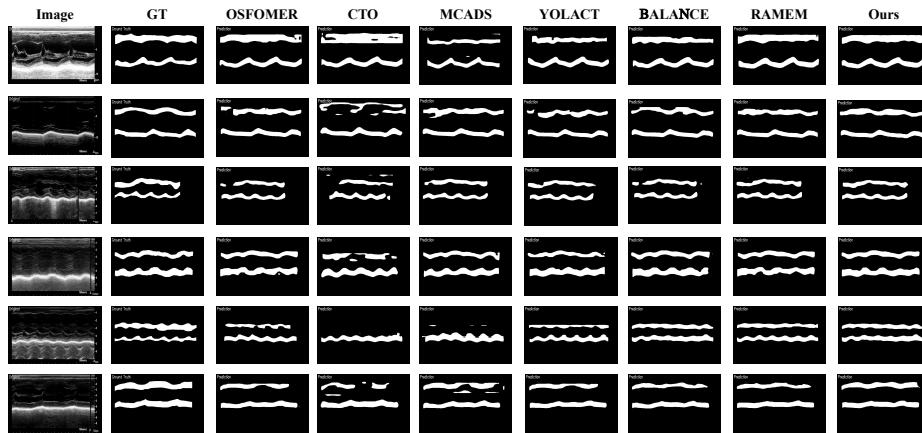
383

384

385

386

387



388

389

390

391

392

393

394

395

396

397

398

399

400

401

402

403

404

405  
Figure 2: Visualization of segmentation prediction results on the MEIS dataset for different methods.  
406

407

408

409

410

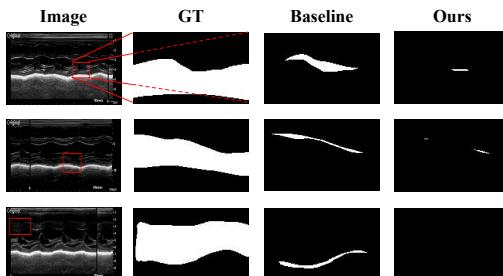
## 4.4 VISUALIZATION OF STEADY-STATE ERROR ELIMINATION(Q2)

411

412 To visually demonstrate that our PI-controlled  
413 framework effectively eliminates steady-state  
414 error, we design a qualitative comparison ex-  
415 periment, with results shown in Fig. 6. We  
416 select several challenging samples where the  
417 baseline model performed poorly and com-  
418 pared its error map against that of our method.  
419 More visual examples are presented in Ap-  
420 pendix A.8.3.

421

422 In the figure, each row represents a test case.  
423 The first column shows the original input, with  
424 red boxes highlighting the most challenging  
425 regions. The third column, displaying the  
426 baseline’s error map, clearly exhibits persistent  
427 bright spots within these boxed regions, which  
428 correspond to the steady-state error. In com-  
429 parison, the fourth column shows the error map of  
430 our method, where the error in the same regions  
431 is significantly suppressed and close to zero. The  
432 visualization results show that our PI controller  
433 can precisely target and eliminate the steady-state  
434 errors that conventional methods fail to resolve,  
435 thereby enhancing segmentation reliability. In ad-  
436 dition to the qualitative error maps, we also provide  
437 a quantitative measurement of steady-state error  
438 and visualization showing how model attention changes  
439 under PI-Control in Appendix A.10.

440  
Figure 6: Visualization of segmentation error re-  
441 sults on the MEIS dataset for different methods.  
442 Red boxes highlight the challenging regions.  
443

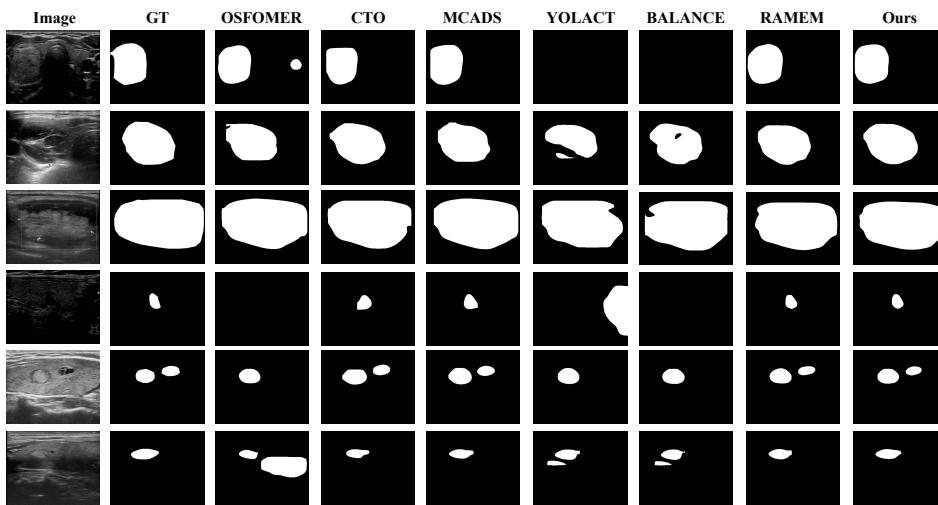


Figure 3: Visualization of segmentation prediction results on the TN3K dataset for different methods.

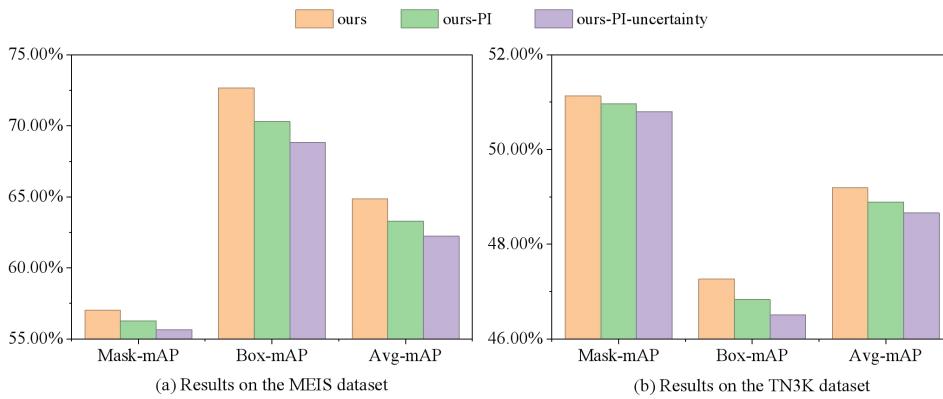


Figure 4: Ablation study. We compare the full model (ours) against a variant without the integral term (ours-PI) and a variant without both the PI control and the uncertainty loss (ours-PI-uncertainty), which represents the RAMEM baseline.

## 4.5 PARAMETER SENSITIVITY ANALYSIS

To investigate the sensitivity of our PI controller to its hyperparameters, the proportional gain ( $K_p$ ) and the integral gain ( $K_i$ ), we conduct parameter sensitivity analysis. As presented in Fig. 5, we vary one parameter while keeping the others fixed to observe the impact on model performance.

The results indicate that the optimal performance is achieved around  $K_p = 2.0$  and  $K_i = 0.1$ . While performance slightly degrades as the parameters deviate from the optimal values, the changes are small. Crucially, the performance across the parameter variation range remains significantly superior to the baseline without the PI controller. This demonstrates that our method exhibits good robustness to hyperparameter selection, facilitating its deployment and tuning in practical applications. Beyond the controller gains, we further analyse the sensitivity with respect to the number of hard pixels  $k$ , the history length  $N$  used in the integral term and likelihood. The results are summarized in Appendix A.11.

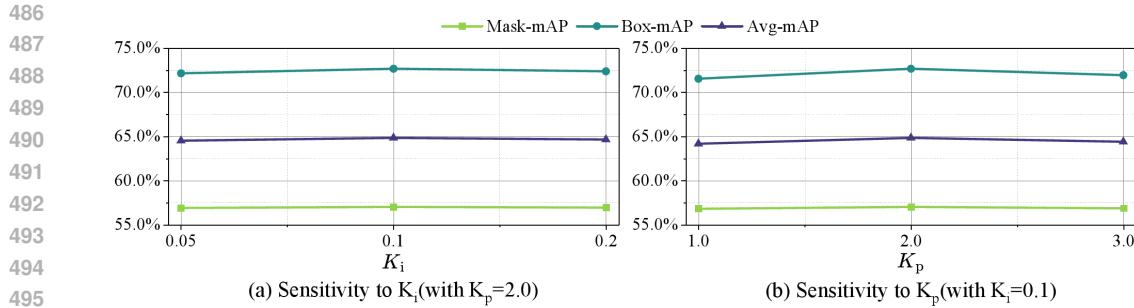


Figure 5: Parameter sensitivity analysis for the PI controller gains,  $K_p$  and  $K_i$ , on the MEIS dataset. Plot (a) shows the results of varying the integral gain  $K_i$  while holding  $K_p = 2.0$ . Plot (b) shows the results of varying the proportional gain  $K_p$  while holding  $K_i = 0.1$

## 5 CONCLUSION

In this paper, we present a novel training framework that rethinks segmentation through the lens of classical control theory. We identify that conventional loss functions are equivalent to proportional controllers, making them incapable of eliminating steady-state errors in challenging segmentation tasks. To address this, we designed a PI controlled uncertainty integrated into the training process. This model-agnostic approach leverages an integral controller to accumulate a memory of historical errors, generating an adaptive corrective signal to resolve stubborn inaccuracies. Experiments on the MEIS and TN3K ultrasound datasets validate our method’s effectiveness, showing superior performance over state-of-the-art baselines and visually confirming the elimination of steady-state errors. Moreover, since our training process is explicitly formulated as a closed-loop feedback system, the PI-controlled uncertainty mechanism is conceptually compatible with real-time or continual adaptation scenarios (e.g., test-time adaptation under domain shift). We leave such extensions to online and cross-domain settings as promising directions for future work.

## ETHICS STATEMENT

This research utilizes two publicly available medical imaging datasets: MEIS and TN3K. Our study did not involve direct interaction with human subjects, and we did not collect any new patient data. The primary goal of our work is to enhance the accuracy of medical ultrasound image segmentation, which can serve as a beneficial tool to aid clinical diagnosis and decision-making. We acknowledge that any machine learning model trained on specific datasets may carry inherent biases. Future work should investigate the generalizability of our method across diverse patient populations and imaging hardware to ensure fairness and robustness.

## REPRODUCIBILITY STATEMENT

To ensure the reproducibility of our work, we provide the following details. **Code:** An anonymized implementation of our method is available at: <https://anonymous.4open.science/r/PI-control-uncertainty-B82C>. **Datasets:** We conduct experiments on two publicly available datasets: MEIS and TN3K. A complete description of the data preprocessing steps for both datasets is provided in Appendix A.5. **Implementation Details:** Our framework is implemented using PyTorch. Full details of the experimental setup, including the learning rate schedule, optimizer, batch size, and data augmentation techniques, are described in Appendix A.7. The key hyperparameters for our proposed PI controller ( $K_p$  and  $K_i$ ) are also detailed in Section 4.1, with a sensitivity analysis presented in Section 4.5. **Evaluation:** The evaluation metrics and protocols used to report our results are detailed in Section 4.1 and Appendix A.6.

540 REFERENCES  
541

542 Moloud Abdar, Soorena Salari, Sina Qahremani, Hak-Keung Lam, Fakhri Karray, Sadiq Hussain,  
543 Abbas Khosravi, U. Rajendra Acharya, Vladimir Makarenkov, and Saeid Nahavandi. Uncer-  
544 taintyFuseNet: Robust uncertainty-aware hierarchical feature fusion model with ensemble monte  
545 carlo dropout for covid-19 detection. *Information Fusion*, 90:364–381, 2023.

546 Wangpeng An, Haoqian Wang, Qingyun Sun, Jun Xu, Qionghai Dai, and Lei Zhang. A PID con-  
547 troller approach for stochastic optimization of deep networks. In *2018 IEEE/CVF Conference on*  
548 *Computer Vision and Pattern Recognition*, pp. 8522–8531, 2018.

549 Karl Johan Åström and Tore Hägglund. *Advanced PID control*. ISA-The Instrumentation, Systems  
550 and Automation Society, 2006.

552 Reza Azad, Ehsan Khodapanah Aghdam, Amelie Rauland, Yiwei Jia, Atlas Haddadi Avval, Afshin  
553 Bozorgpour, Sanaz Karimijafarbigloo, Joseph Paul Cohen, Ehsan Adeli, and Dorit Merhof. Medi-  
554 cal image segmentation review: The success of u-net. *IEEE Transactions on Pattern Analysis*  
555 and *Machine Intelligence*, 46(12):10076–10095, 2024.

556 Daniel Bolya, Chong Zhou, Fanyi Xiao, and Yong Jae Lee. YOLACT: Real-time instance segmen-  
557 tation. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pp. 9156–9165,  
558 2019.

560 Min Dong, Ating Yang, Zhenhang Wang, Dezhen Li, Jing Yang, and Rongchang Zhao. Uncertainty-  
561 aware consistency learning for semi-supervised medical image segmentation. *Knowledge-Based*  
562 *Systems*, 309:112890, 2025.

563 Daniel Duenias, Brennan Nichyporuk, Tal Arbel, and Tammy Riklin Raviv. Hyperfusion: A hyper-  
564 network approach to multimodal integration of tabular and medical imaging data for predictive  
565 modeling. *Medical Image Analysis*, 102:103503, 2025.

567 Jacob G. Elkins and Farbod Fahimi. Control-theoretic techniques for online adaptation of deep  
568 neural networks in dynamical systems. *ArXiv*, abs/2402.00761, 2024.

569 Gene F Franklin, J David Powell, Abbas Emami-Naeini, et al. *Feedback control of dynamic systems*,  
570 volume 10. Pearson Upper Saddle River, NJ, 2010.

572 Yarin Gal and Zoubin Ghahramani. Dropout as a bayesian approximation: Representing model  
573 uncertainty in deep learning. In *Proceedings of The 33rd International Conference on Machine*  
574 *Learning*, pp. 1050–1059, 2016.

575 Kibrom Berihu Girum, Gilles Créhange, and Alain Lalande. Learning with context feedback loop  
576 for robust medical image segmentation. *IEEE Transactions on Medical Imaging*, 40(6):1542–  
577 1554, 2021.

579 Shreyank N. Gowda and David A. Clifton. CC-SAM: SAM with cross-feature attention and context  
580 for ultrasound image segmentation. In *ECCV*, pp. 108–124, 2025.

581 Yihui He, Chenchen Zhu, Jianren Wang, Marios Savvides, and Xiangyu Zhang. Bounding box  
582 regression with uncertainty for accurate object detection. In *2019 IEEE/CVF Conference on*  
583 *Computer Vision and Pattern Recognition (CVPR)*, pp. 2883–2892, 2019.

585 Jiongtong Hu, Wufeng Xue, Jun Cheng, Yingying Liu, Wei Zhuo, and Dong Ni. EchoONE: Seg-  
586 menting multiple echocardiography planes in one model. In *Proceedings of the IEEE/CVF Con-  
587 ference on Computer Vision and Pattern Recognition (CVPR)*, pp. 5207–5216, 2025.

588 Mohamed Jasim Mohamed, Bashra Kadhim Olewi, Ahmad Taher Azar, and Ahmed Redha  
589 Mahlous. Hybrid controller with neural network PID/FOPID operations for two-link rigid robot  
590 manipulator based on the zebra optimization algorithm. *Frontiers in Robotics and AI*, 11, 2024.

592 Thierry Judge, Olivier Bernard, Woo-Jin Cho Kim, Alberto Gomez, Agisilaos Chartsias, and Pierre-  
593 Marc Jodoin. Asymmetric contour uncertainty estimation for medical image segmentation. In  
594 *MICCAI 2023*, pp. 210–220, Cham, 2023.

594 Simon Kohl, Bernardino Romera-Paredes, Clemens Meyer, Jeffrey De Fauw, Joseph R Ledsam,  
 595 Klaus Maier-Hein, SM Eslami, Danilo Jimenez Rezende, and Olaf Ronneberger. A probabilistic  
 596 u-net for segmentation of ambiguous images. *Advances in neural information processing systems*,  
 597 2018.

598 Balaji Lakshminarayanan, Alexander Pritzel, and Charles Blundell. Simple and scalable predictive  
 599 uncertainty estimation using deep ensembles. In *Advances in Neural Information Processing  
 600 Systems*, volume 30, 2017.

602 Haeyun Lee, Moon Hwan Lee, Sangyeon Youn, Kyungsu Lee, Hah Min Lew, and Jae Youn Hwang.  
 603 Speckle reduction via deep content-aware image prior for precise breast tumor segmentation in  
 604 an ultrasound image. *IEEE Transactions on Ultrasonics, Ferroelectrics, and Frequency Control*,  
 605 69(9):2638–2650, 2022.

606 Chenxin Li, Xinyu Liu, Wuyang Li, Cheng Wang, Hengyu Liu, Yifan Liu, Zhen Chen, and Yixuan  
 607 Yuan. U-KAN makes strong backbone for medical image segmentation and generation. *Proceed-  
 608 ings of the AAAI Conference on Artificial Intelligence*, 39(5):4652–4660, 2025.

610 Yi Lin, Dong Zhang, Xiao Fang, Yufan Chen, Kwang-Ting Cheng, and Hao Chen. Rethinking  
 611 boundary detection in deep learning models for medical image segmentation. In *Information  
 612 Processing in Medical Imaging*, pp. 730–742, Cham, 2023.

613 Jiawei Liu, Jing Zhang, and Nick Barnes. Modeling aleatoric uncertainty for camouflaged object  
 614 detection. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer  
 615 Vision (WACV)*, pp. 1445–1454, 2022.

617 Bálint Mucsányi, Michael Kirchhof, and Seong Joon Oh. Benchmarking uncertainty disentangle-  
 618 ment: Specialized uncertainties for specialized tasks. In *Advances in Neural Information Pro-  
 619 cessing Systems*, volume 37, pp. 50972–51038, 2024.

620 Jialun Pei, Tianyang Cheng, Deng-Ping Fan, He Tang, Chuanbo Chen, and Luc Van Gool. OS-  
 621 Former: One-stage camouflaged instance segmentation with transformers. In *Computer Vision –  
 622 ECCV 2022*, pp. 19–37, Cham, 2022.

624 Murat Sensoy, Lance Kaplan, and Melih Kandemir. Evidential deep learning to quantify classifica-  
 625 tion uncertainty. In *Advances in Neural Information Processing Systems*, volume 31, 2018.

626 Ashutosh Tiwari, Soumya Mishra, and Tsung-Rong Kuo. Current AI technologies in cancer diag-  
 627 nóstics and treatment. *Molecular Cancer*, 24(1):159, 2025.

629 Ching-Hsun Tseng, Shao-Ju Chien, Po-Shen Wang, Shin-Jye Lee, Bin Pu, and Xiao-Jun Zeng. Real-  
 630 time automatic m-mode echocardiography measurement with panel attention. *IEEE Journal of  
 631 Biomedical and Health Informatics*, 28(9):5383–5395, 2024.

633 Jiacheng Wang, Lan Wei, Liansheng Wang, Qichao Zhou, Lei Zhu, and Jing Qin. Boundary-aware  
 634 transformers for skin lesion segmentation. In *MICCAI*, pp. 206–216, 2021.

635 Saad Wazir and Daeyoung Kim. Rethinking decoder design: Improving biomarker segmentation  
 636 using depth-to-space restoration and residual linear attention. In *Proceedings of the IEEE/CVF  
 637 Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 30861–30871, 2025.

639 Feilong Xu, Feiyang Yang, Xiongfei Li, and Xiaoli Zhang. A unified loss for handling inter-class  
 640 and intra-class imbalance in medical image segmentation. *Proceedings of the AAAI Conference  
 641 on Artificial Intelligence*, 39(8):8842–8850, 2025.

642 Jiacong Xu, Zixiang Xiong, and Shankar P. Bhattacharyya. PIDNet: A real-time semantic segmen-  
 643 tation network inspired by PID controllers. In *2023 IEEE/CVF Conference on Computer Vision  
 644 and Pattern Recognition (CVPR)*, pp. 19529–19539, 2023.

646 Ghada Zamzmi, Li-Yueh Hsu, Wen Li, Vandana Sachdev, and Sameer Antani. Harnessing ma-  
 647 chine intelligence in automatic echocardiogram analysis: Current status, limitations, and future  
 directions. *IEEE Reviews in Biomedical Engineering*, 14:181–203, 2021.

648 Siyuan Zhang, Shan Wu, Shaoqiang Shang, Xuewei Qin, Xin Jia, Dapeng Li, Zhiwei Cui, Tianqi Xu,  
 649 Gang Niu, Ayache Bouakaz, and Mingxi Wan. Detection and monitoring of thermal lesions in-  
 650 duced by microwave ablation using ultrasound imaging and convolutional neural networks. *IEEE*  
 651 *Journal of Biomedical and Health Informatics*, 24(4):965–973, 2020.

652 Jiazhou Zhou, Xu Zheng, Yuanhuiyi Lyu, and Lin Wang. ExACT: Language-guided conceptual  
 653 reasoning and uncertainty estimation for event-based action recognition and more. In *Proceedings*  
 654 *of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pp. 18633–  
 655 18643, 2024.

656 X Zhou, B Sheil, S Suryasentana, and P Shi. Uncertainty-guided u-net for soil boundary segmen-  
 657 tation using monte carlo dropout. *Computer-Aided Civil and Infrastructure Engineering*, 40(10):  
 658 1370–1387, 2025.

## 661 662 A APPENDIX

### 663 664 A.1 THE USE OF LLMs

665 The language in this paper is polished by a large language model (LLM) to enhance clarity and  
 666 readability. The author remains responsible for the final content and academic integrity.

### 667 668 A.2 BASELINE

669 Our work builds upon the RAMEM framework, an efficient real-time instance segmentation model  
 670 adapted from YOLACT for the specific challenges of M-mode echocardiography. Its good perfor-  
 671 mance and high speed make it a suitable backbone for our control-theoretic improvements.

672 The overall architecture of RAMEM follows the single-stage design of YOLACT, consisting of three  
 673 core components: a feature extraction backbone, a Feature Pyramid Network (FPN), and parallel  
 674 prediction heads. Unlike YOLACT, which is based on a traditional ResNet backbone, RAMEM  
 675 utilizes its proprietary UPANet V2 as the backbone. Its core innovation is the introduction of the  
 676 Panel Attention mechanism, designed to efficiently build a global receptive field. Since targets in  
 677 echocardiograms (such as ventricular walls) typically occupy a large portion of the image pixels, the  
 678 limited local receptive field of traditional CNNs struggles to capture complete structural information.  
 679 Panel Attention uses a lossless pixel-unshuffle operation to transfer spatial information to the channel  
 680 dimension, achieving local-to-global attention with low computational overhead and thus resolving  
 681 this issue. This enables the model to segment large objects more accurately.

682 RAMEM’s prediction heads consist of two parallel sub-networks: the ProtoNet and the Prediction  
 683 Module. ProtoNet is responsible for generating a series of high-quality, generic prototype masks  
 684 from a single high-resolution feature map. This process is independent of the number of instances  
 685 in the image, making it highly computationally efficient. The Prediction Module, on the other hand,  
 686 operates on multiple feature levels from the FPN, predicting class confidence, bounding boxes, and  
 687 a set of coefficients for each detected potential instance.

688 Finally, the segmentation mask for each instance (which is the controlled variable  $M_{\text{pred}}$  in our  
 689 system) is produced by linearly combining the prototype masks from the ProtoNet with the corre-  
 690 sponding instance’s coefficients predicted by the Prediction Module. This efficient design allows  
 691 RAMEM to achieve real-time processing while maintaining high accuracy, and our proposed PI  
 692 controller is designed to regulate and optimize this final output.

### 693 694 A.3 ALGORITHM

695 We provide the pseudocode for our proposed PI-Controlled Uncertainty Learning mechanism. Al-  
 696 gorithm 1 details the step-by-step process of computing the proportional and integral error terms,  
 697 generating the modulated error signal, and constructing the final PI-Controlled Uncertainty Loss for  
 698 each training batch, as described in Section 3.2.2.



756 A.4.2 CONTROL APPROACHES IN DEEP LEARNING  
757

758 Control theory, particularly principles of feedback and optimization, has been increasingly integrated  
759 into deep learning to stabilize training dynamics and enhance performance. For instance,  
760 some approaches draw direct parallels between optimizers and controllers. An et al. (2018) formalized  
761 this connection by proposing a full Proportional-Integral-Derivative (PID) optimizer. This  
762 optimizer incorporates a derivative term (the change in gradient) to anticipate updates and accelerate  
763 convergence. In robotics, Jasim Mohamed et al. (2024) developed hybrid control structures that  
764 combine neural networks with PID controllers for precise trajectory tracking.

765 LFB-Net (Girum et al., 2021) formulates segmentation as a recurrent process with an explicit context  
766 feedback loop, where the output is encoded and fed back to refine subsequent predictions. Some  
767 other works focus on online parameter adaptation. Elkins & Fahimi (2024) models a neural network  
768 as a continuous-time dynamical system and applies the Super-Twisting Algorithm (STA) to derive  
769 update rules for the network’s final layer, thereby guaranteeing error convergence.

770 However, despite these advances, existing methods often struggle to mitigate severe low contrast,  
771 speckle noise, and ambiguous tissue boundaries. Furthermore, the potential of integral control,  
772 which is a core component of PI controllers designed to eliminate persistent steady-state errors,  
773 remains underexplored in the context of segmentation. Our work designs PI-controlled uncertainty  
774 to address the stubborn boundary inaccuracies that many recently proposed segmentation methods  
775 struggle to handle.

776  
777 A.5 DATASETS  
778

779 To validate the effectiveness of our proposed method, we conduct experiments on two distinct and  
780 publicly available ultrasound datasets: MEIS and TN3K. The two selected datasets represent two  
781 distinctly different challenges in medical ultrasound segmentation. MEIS consists of M-mode tem-  
782 poral images with relatively regular structures but blurred boundaries, while TN3K comprises B-  
783 mode static images where nodule shape, size, and contrast vary dramatically.

784 The M-mode Echocardiography of Interventricular Septum (MEIS) dataset is a comprehensive col-  
785 lection of 2,639 M-mode ultrasound images from 923 de-identified subjects, with a resolution of  
786 1024×768 pixels. The dataset covers two standard cardiac views: the Aortic Valve (AV) and the  
787 Left Ventricle (LV). For the purposes of our study, we selected a specific subset of 763 images cor-  
788 responding to the Left Ventricle (LV) view. In this LV view, the key annotated structures are the  
789 Interventricular Septum (IVS) and the Left Ventricular Posterior Wall (LVPW). The thicknesses of  
790 these structures, along with the Left Ventricular Internal Diameter (LVID), are crucial for deriv-  
791 ing key cardiac function indicators such as End-Diastolic Volume (EDV) and End-Systolic Volume  
792 (ESV). To isolate the relevant M-mode data for our segmentation task, the images were preprocessed  
793 to remove the B-mode components, resulting in a final resolution of 1024×418. The dataset’s blurred  
794 boundaries and subtle structural details provide a significant challenge for precise segmentation.

795 The Thyroid Nodule 3000 (TN3K) dataset, provided by the Zhujiang Hospital of Southern Medical  
796 University, is highly representative of real-world clinical practice. It comprises 3,493 ultrasound  
797 images collected from 2,421 different patients between January 2016 and August 2020. This dataset  
798 presents a high degree of complexity, making it an excellent benchmark. The images feature varying  
799 dimensions, and approximately 9% of them contain two or more thyroid nodules, increasing the  
800 difficulty of the automatic segmentation task. Each image is accompanied by expert-provided, pixel-  
801 level annotations. For our experiments, we adhere to the official data split, which designates 2,879  
802 images for the training set and the remaining 614 images for the testing set.

803 A.6 EVALUATION METRICS  
804

805 To provide a comprehensive assessment of our method’s performance, we employ a suite of standard  
806 metrics for both pixel-level segmentation accuracy and instance-level detection and segmentation  
807 quality. Before detailing the metrics, we define the fundamental terms based on a pixel-wise com-  
808 parison between the predicted segmentation mask ( $M_{\text{pred}}$ ) and the ground truth mask ( $M_{\text{gt}}$ ): True  
809 Positive (TP) is the number of pixels correctly classified as the target object; False Positive (FP) is  
the number of pixels incorrectly classified as the target object (background predicted as foreground);

810 False Negative (FN) is the number of pixels incorrectly classified as background (foreground predicted as background).  
 811

812 From these, we derive Precision, which measures the accuracy of positive predictions (TP/(TP+FP)),  
 813 and Recall, which measures the model’s ability to identify all actual positive pixels (TP/(TP+FN)).  
 814 To balance these two, we use the Dice Similarity Coefficient (DSC), a widely used metric in medical  
 815 imaging that measures the overlap between the  $M_{\text{pred}}$  and the  $M_{\text{gt}}$ . It is calculated as  
 816

$$817 \text{DSC} = \frac{2 \times |M_{\text{pred}} \cap M_{\text{gt}}|}{|M_{\text{pred}}| + |M_{\text{gt}}|} = \frac{2 \times \text{TP}}{2 \times \text{TP} + \text{FP} + \text{FN}}. \quad (7)$$

820 In addition to region-overlap metrics, we further evaluate the boundary localization accuracy using  
 821 the 95th percentile Hausdorff Distance (HD95), a distance-based metric that measures the  
 822 largest segmentation discrepancy after excluding extreme outliers. Formally, given two boundary  
 823 point sets  $P$  and  $G$  extracted from  $M_{\text{pred}}$  and  $M_{\text{gt}}$ , the directed Hausdorff distance is defined as  
 824  $h(P, G) = \max_{p \in P} \min_{g \in G} \|p - g\|$ . The HD95 is computed as the 95th percentile of the symmetric  
 825 bidirectional distances:  
 826

$$827 \text{HD95} = \text{Percentile}_{95} \left( \{ h(P, G), h(G, P) \} \right), \quad (8)$$

828 providing a robust indicator of boundary accuracy that is particularly relevant for ultrasound images  
 829 with ambiguous or fuzzy contours.  
 830

831 For end-to-end instance segmentation performance, we adhere to the standard COCO evaluation  
 832 protocol, which is based on the mean Average Precision (mAP). This protocol’s foundation is the  
 833 Intersection over Union (IoU) metric, calculated as  $\text{IoU}(A, B) = |A \cap B| / |A \cup B|$ , which quantifies  
 834 the overlap between a predicted instance (A) and a ground truth instance (B). The mAP score  
 835 is obtained by averaging the Average Precision (AP) across a range of IoU thresholds (from 0.5 to  
 836 0.95), rewarding models that are accurate at various levels of overlap quality. We report Box-mAP  
 837 and Mask-mAP, which are calculated using the IoU of bounding boxes and segmentation masks,  
 838 respectively, to evaluate detection and segmentation quality. Finally, we report the Avg-mAP, which  
 839 is the mean of Box-mAP and Mask-mAP, providing a single, comprehensive score for overall per-  
 840 formance.  
 841

#### A.7 IMPLEMENTATION DETAILS

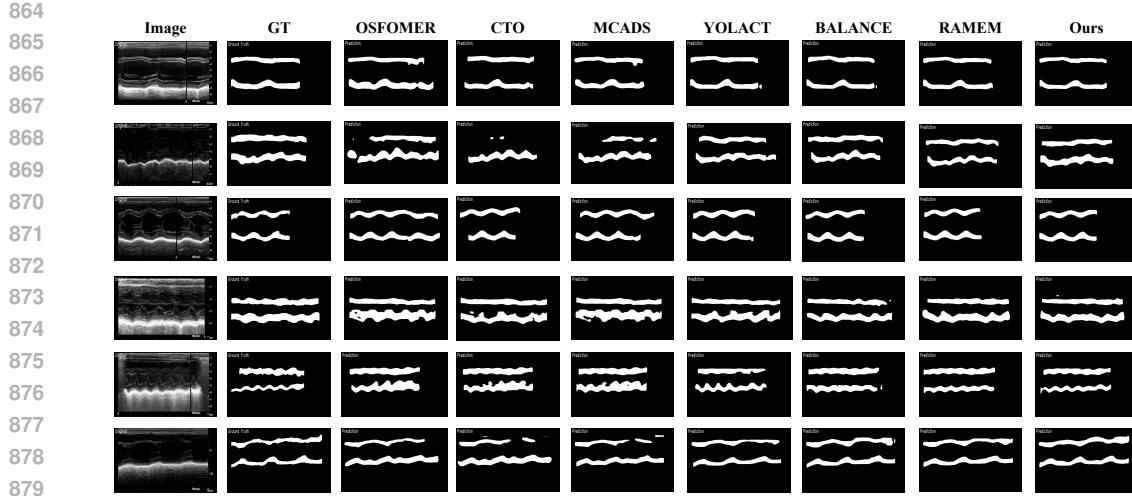
842 All experiments are conducted using PyTorch 1.11.0 and Python 3.8 (Ubuntu 20.04) on a single  
 843 NVIDIA RTX 4090 GPU with CUDA 11.3. We employ the Stochastic Gradient Descent (SGD)  
 844 optimizer with an initial learning rate of  $1 \times 10^{-3}$ , a momentum of 0.9, and a weight decay of  
 845  $5 \times 10^{-4}$ . The learning rate is decayed by a factor of 0.1 at step 13,000. All models are trained for  
 846 a maximum of 22,000 steps with a batch size of 3. The weights for the overall loss function (Eq. 6)  
 847 are all set to 1.  
 848

849 We apply several data augmentation techniques during training. These include random adjustments  
 850 to brightness and contrast, as well as horizontal flipping with a probability of 0.5. All input images  
 851 are resized to a uniform dimension of  $544 \times 544$ , with scale-padding used to maintain the aspect  
 852 ratio. Pixel values are normalized to the  $[0, 1]$  range before being fed into the network. To ensure  
 853 fair comparison, these enhancement methods are also used in the baseline experiments.  
 854

855 For our proposed PI control mechanism, the key hyperparameters are set as follows: the proportional  
 856 gain  $K_p$  is set to 2.0, and the integral gain  $K_i$  is set to 0.1. The controller’s memory is configured  
 857 with a history length of  $N_{\text{epochs}} = 5$ . For the proportional term’s hard pixel mining, we select the  
 858 top-k pixels, where k was set to 500. Following our two-stage training strategy, the PI-controlled  
 859 uncertainty loss is activated after the 30th epoch. For fair comparison, all experiments of baseline  
 860 methods are conducted using the same configuration as the original papers.  
 861

#### A.8 ADDITIONAL QUALITATIVE RESULTS

862 This section provides additional qualitative results to supplement the findings presented in the main  
 863 paper. We showcase more visual comparisons on the MEIS and TN3K datasets to demonstrate the  
 864 effectiveness of our proposed method on a wider range of challenging cases.  
 865



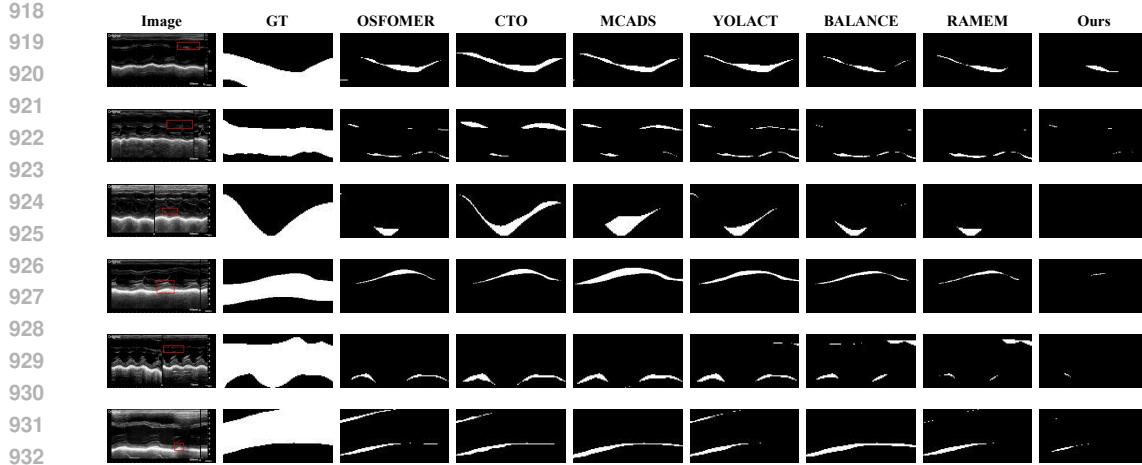


Figure 9: Additional visualization of segmentation error results on the MEIS dataset for different methods. Red boxes highlight the challenging regions.

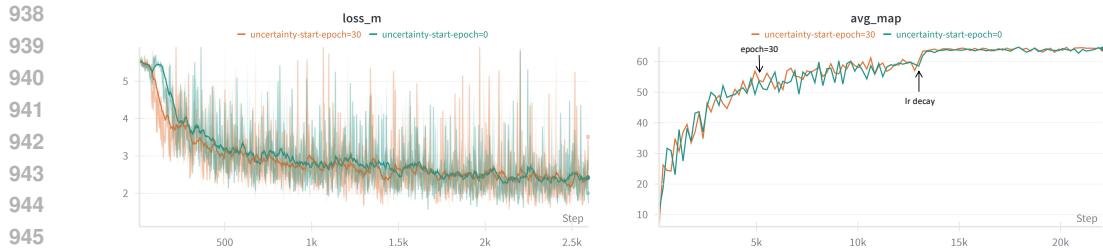


Figure 10: Impact of the uncertainty loss start epoch on training dynamics and performance. (Left) Convergence of mask loss in early training steps. (Right) Evolution of Average mAP throughout the training process.

## A.8.2 MORE EXAMPLES ON THE TN3K DATASET

Fig.8 displays additional results on the TN3K dataset, focusing on nodules with low contrast, irregular shapes, or small sizes. Methods like OSFOMER, BALANCE, and YOLACT often only capture a fraction of the nodule or miss it entirely. Others, such as CTO and RAMEM, may capture the general location but suffer from boundary leakage and fail to segment the precise, often irregular, contours of the nodules. Our method demonstrates a superior ability to adhere to the true nodule boundaries. Even in cases where useful features are extremely sparse, our approach produces a segmentation that is more accurate and faithful to the nodule’s true shape than baseline results, making it more reliable for potential clinical assessment.

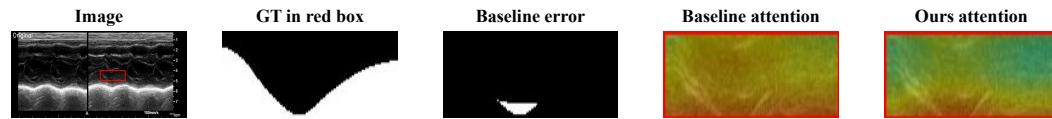


Figure 11: Feature-level attention visualization at epoch 30. The ambiguous ROI is marked by a red box. Left: original image and ground-truth mask. Middle: state-state error in baseline model. Right: JET heatmaps for the baseline and PI-Control.

972  
973 Table 4: Sensitivity of the PI controller to the number of hard pixels  $k$  on MEIS dataset.  
974  
975  
976  
977  
978

$k$	100	200	400	500	600	800	1000	2000
DSC	87.38	87.41	87.49	87.55	87.60	87.58	87.54	87.47
Mask-mAP	56.48	56.83	57.06	57.03	57.01	57.05	56.99	56.92
Box-mAP	71.75	72.17	72.54	72.67	72.66	72.71	72.61	72.36

979  
980 Table 5: Sensitivity of the PI controller to the history length  $N$  on MEIS dataset.  
981  
982  
983  
984  
985

$N$	1	3	5	7	9	12
DSC	87.12	87.29	87.55	87.59	87.51	87.52
Mask-mAP	56.36	56.64	57.03	57.08	57.04	56.98
Box-mAP	70.88	71.30	72.67	72.72	72.58	72.60

986  
987 A.8.3 VISUALIZATION OF STEADY-STATE ERROR ELIMINATION  
988

989 To supplement the analysis in the main text, Fig. 9 provides a direct visualization of steady-state  
990 error elimination on the MEIS dataset. The images in this figure represent the absolute difference  
991 between the model’s prediction and the ground truth, where white areas indicate segmentation er-  
992 rror. The red boxes highlight regions where conventional methods struggle. As shown, all baseline  
993 methods exhibit significant residual errors (bright white regions), which are visual manifesta-  
994 tions of the steady-state error that persists after training converges. In contrast, the error maps for our  
995 method are almost entirely black in these same regions. This provides visual evidence that our PI-  
996 controlled framework successfully targets and eliminates the persistent errors that plague standard  
997 training paradigms, leading to more reliable and accurate segmentation.

998 A.9 STABILITY OF THE PI CONTROLLER AND DELAYED ACTIVATION  
999

1000 Fig. 10 illustrates the impact of the uncertainty loss start epoch on training dynamics. When the  
1001 PI-controlled uncertainty loss is activated from the beginning of training, the mask loss exhibits  
1002 slightly stronger oscillations in the early iterations. However, thanks to the bounded integration  
1003 horizon ( $N = 5$ ) and the small integral gain ( $K_i = 0.1$ ), these fluctuations are quickly damped  
1004 and the model converges to a similar or even better Avg-mAP compared to the delayed-activation  
1005 setting.

1006 To avoid unnecessary transient oscillations, we therefore adopt a two-stage training strategy in the  
1007 main experiments: the model is first optimized using only the static mask loss until a reasonable  
1008 initial solution is reached, and the PI-controlled uncertainty loss is then activated. Importantly, the  
1009 closed-loop feedback remains active whenever  $L_{\text{uncertainty}}$  is used, so this schedule implements a  
1010 stabilised PI controller rather than a simple curriculum without feedback.

1011  
1012 A.10 STEADY-STATE ERROR  
1013

1014 To quantitatively validate the steady-state error phenomenon, we manually selected a representative  
1015 ambiguous ROI on the MEIS dataset and measured the percentage of misclassified pixels within this  
1016 ROI across training epochs. Table 7 reports the results.

1017 The baseline exhibits a clear steady-state error plateau around 10%–12% after approximately 70  
1018 epochs, whereas our PI-controlled framework continues to decrease the residual error to below 1%,  
1019 directly demonstrating the effectiveness of the integral term in eliminating persistent segmentation  
1020 bias.

1021 To examine how the proposed PI-Control influences the model’s internal attention patterns during  
1022 training, we visualize the feature responses at epoch 30 for both the baseline model and the PI-  
1023 controlled model. We select a representative region of interest (ROI) containing persistent segmen-  
1024 tation ambiguity, highlighted by a red bounding box in the original input image. For each model,  
1025 we extract the corresponding FPN feature map, average it across channels, normalize it to the range  
[0, 1] using a sigmoid function, and map the result to a JET colormap in which blue denotes low

1026

1027 Table 6: Comparison between Laplace and Gaussian likelihoods on MEIS dataset.

Likelihood	DSC	Mask-mAP	Box-mAP
Gaussian	87.18	56.68	71.05
Laplace (ours)	<b>87.55</b>	<b>57.03</b>	<b>72.67</b>

1031

1032

1033

Table 7: Steady-state error (%) within an ambiguous ROI over training epochs on MEIS dataset.

Epoch	10	20	30	40	50	60	70	80	90	100
RAMEM	95.29	64.28	38.37	32.75	27.15	20.44	12.27	11.29	10.77	10.83
Ours	94.96	65.93	32.75	21.86	13.38	5.82	0.69	0.43	0.38	0.41

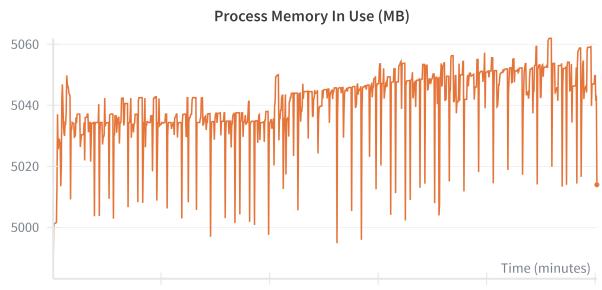
1038

1039

1040 response and red denotes high response. The visualization therefore reflects the relative activation  
 1041 strength and highlights spatial regions that attract stronger feature-level attention.

1042

1043



1044

1045

1046

1047

1048

1049

1050

1051

1052

1053

1054

1055

Figure 12: Process memory usage over time.

1056

1057

1058

1059

1060

1061

1062

1063

1064

1058 As shown in Fig. 11, the baseline model produces a diffuse and weak activation distribution: al-  
 1059 though the ambiguous area shows slightly elevated responses, the attention remains broadly spread  
 1060 and does not strongly emphasize the region responsible for the steady-state error. In contrast, the  
 1061 PI-controlled model exhibits a sharply localized high-response region that aligns precisely with the  
 1062 ambiguous ROI. This concentrated attention indicates that the proportional term identifies uncertain  
 1063 pixels while the integral term accumulates persistent residual errors and amplifies them over time,  
 1064 directing the network to focus on hard-to-correct regions more aggressively. These results provide  
 1065 direct visual evidence that the closed-loop PI mechanism reshapes the internal attention dynamics,  
 1066 enabling the model to attend to regions where errors remain unresolved across epochs.

1067

1068

### A.11 ADDITIONAL ANALYSIS OF PARAMETERS

1069

1070

1071

1072

1073

1074

1075

1076

1077

1078

1079

We conduct a comprehensive parameter analysis to examine how different components of the pro-  
 posed PI-controlled uncertainty framework affect segmentation performance. Specifically, we study  
 (1) the number of hard pixels  $k$  used in the top- $k$  selection, (2) the history length  $N$  of the integral  
 term, and (3) the choice of likelihood (Laplace vs. Gaussian) used for modelling the PI-modulated  
 error signal  $e_{\text{mod}}(t)$ . These parameters directly influence the behaviour of the closed-loop controller  
 and the robustness of the uncertainty modelling.

**Sensitivity to the number of hard pixels  $k$ .** Table 4 reports the performance on MEIS when varying  
 $k$ . The results show a smooth trend: performance improves as  $k$  increases from 100 to around 400–  
 800, and only slightly degrades when  $k$  becomes very large. This is intuitive—the proportional term  
 benefits from a moderate number of challenging pixels but including too many easy pixels weakens  
 its discriminative effect.

**Sensitivity to the history length  $N$ .** Table 5 summarises the effect of varying the integral window  
 size  $N$ . Performance improves when  $N$  increases from 1 to around 5–7, reflecting the benefit of

1080

1081 Table 8: Comparison with Probabilistic U-Net on the MEIS dataset. The best results are highlighted  
1082 in bold.

Method	Recall	Precision	DSC	HD95 ↓
Probabilistic U-Net	<b>88.63± 1.84</b>	85.52 ± 2.19	86.97 ± 1.82	16.25 ± 7.73
RAMEM	87.31 ± 2.33	87.25 ± 2.01	86.91 ± 1.84	14.17 ± 5.17
Ours	87.97 ± 2.13	<b>87.63± 1.97</b>	<b>87.55± 1.69</b>	<b>13.59± 5.72</b>

1083

1084

1085 accumulating a modest temporal history of residual errors. Larger  $N$  does not lead to further gains  
1086 as excessively long memory weakens the controller’s responsiveness, consistent with PI control  
1087 theory.1088 **Effect of the likelihood function.** Besides the PI-related parameters, we also compare the Laplace  
1089 likelihood used in the main paper with a Gaussian alternative that predicts a per-instance variance.  
1090 As shown in Table 6, the Laplace formulation consistently yields better performance. The heavier-  
1091 tailed Laplace distribution better models the PI-modulated errors, which may deviate from Gaussian  
1092 due to the controller’s targeted emphasis on hard examples and persistent error regions.

1093

1094

1095

1096

1097

1098

1099

1100 Fig. 12 plots the process memory usage over time with the PI controller. The PI-controlled model  
1101 stores only one scalar error statistic per positive instance for each of the  $N$  past epochs. For the  
1102 MEIS training set with  $M = 1174$  instances and  $N = 5$ , this corresponds to  $N \times M = 5870$   
1103 scalars.

1104

1105

1106

1107

1108

1109

## A.12 MEMORY

Fig. 12 plots the process memory usage over time with the PI controller. The PI-controlled model stores only one scalar error statistic per positive instance for each of the  $N$  past epochs. For the MEIS training set with  $M = 1174$  instances and  $N = 5$ , this corresponds to  $N \times M = 5870$  scalars.

Empirically, enabling the PI controller increases the peak GPU memory from about 5000 MB to 5060 MB, i.e., less than 1.2% overhead. Since only instance-level scalars are stored, this additional memory cost is independent of the image resolution or dimensionality, indicating that the method can scale to higher-resolution images and even 3D volumes without prohibitive memory growth.

## A.13 ADDITIONAL BASELINE COMPARISON AND CROSS-MODALITY EVALUATION

To further position our method against uncertainty-aware segmentation frameworks, we compare with Probabilistic U-Net(Kohl et al., 2018) on MEIS dataset. As shown in Table 8, our PI-controlled framework achieves higher DSC and Precision and lower HD95, demonstrating that incorporating PI feedback on uncertainty provides more effective error correction than modelling output distributions alone.

We additionally evaluate our framework on the ISIC 2018 dermoscopic lesion segmentation dataset, which differs substantially from ultrasound in imaging physics, contrast patterns and boundary characteristics. Table 9 and 10 reports the results. The consistent gains in DSC, HD95 and Avg-mAP indicate that the proposed PI-controlled optimization is not restricted to ultrasound, but can benefit segmentation under significant domain shifts.

1120

1121

1122

Table 9: Results on the ISIC 2018 dataset. The best results are highlighted in bold.

Methods	Recall	Precision	DSC	HD95↓
Baseline	<b>95.47±1.35</b>	84.29±1.98	87.42±0.41	207.56±12.36
Ours	94.95±0.87	<b>86.45±1.66</b>	<b>87.93±0.38</b>	<b>185.47±9.92</b>

1123

1124

1125

1126

1127

1128

1129

Table 10: mAP Results on the ISIC 2018 dataset. The best results are highlighted in bold.

Methods	Mask-mAP	Box-mAP	Avg-mAP
Baseline	66.83±3.18	60.42±3.74	63.62±3.45
Ours	<b>68.15±2.86</b>	<b>63.26±3.51</b>	<b>65.70±3.16</b>

1130

1131

1132

1133