



Reinforcement Learning for Photometric Tuning Improves Ovarian Lesion Segmentation

Quoc-Vi Tran¹, Duy-Hai Vu², Phi Dinh Huynh³ and Nguyen Viet Hung⁴

ABSTRACT

Image segmentation is a fundamental task in computer vision and has become an essential component of many real-world applications, particularly in medical image analysis. In ovarian tumor diagnosis, accurate ultrasound image segmentation can assist clinicians by reducing the time required for lesion delineation and facilitating faster treatment planning. However, ultrasound imaging remains challenging because images typically exhibit low contrast, speckle noise, and poorly defined lesion boundaries. These characteristics substantially degrade the performance of automated segmentation models, and reduce the accuracy of tumor localization. In this study, we introduce a reinforcement learning (RL)-based method designed to improve image contrast, and effectively mitigate this problem. We apply a Proximal Policy Optimization (PPO) Agent that applies small, structure-preserving photometric transformations (contrast/gamma adjustment, noise reduction and contrast-limited adaptive histogram equalization) to the input image and a pretrained model (U-Net) to evaluate the output results. The Agent receives a distinguishable reward based on the improvement in the soft-Dice Similarity Coefficient (S-DSC) from one step to the next, with penalties for unnecessary actions and an explicit stopping action. This method improves S-DSC over the baseline U-Net without retraining or architectural changes to the segmentation set. The proposed method consistently improved segmentation performance across all evaluated cases. The most substantial improvement was observed in challenging ultrasound images, where the Dice score increased by up to 78%. These results demonstrate that adaptive, learned photometric adjustments can effectively enhance inference-time segmentation without requiring retraining or architectural modifications to the segmentation model. The approach is designed to be architecture-independent, easy to integrate with frozen segmenters, and to maintain full interpretability across step-by-step action traces.

Article information:

Keywords: Reinforcement Learning, Proximal Policy Optimization, Dice Similarity Coefficient, image Segmentation, Ovarian Tumor Ultrasound, Photometric Enhancement

Article history:

Received: November 25, 2025

Revised: March 12, 2026

Accepted: June 18, 2026

Published: July 11, 2026

(Online)

DOI: 10.37936/ecti-cit.2026203.264968

1. INTRODUCTION

Ovarian cancer remains a major global health burden with high mortality when detected late, while ultrasound is the most accessible first-line imaging tool in routine screening and triage [1]. However, ovarian ultrasound often exhibits low contrast, speckle noise, and blurry boundaries, which complicate tu-

mor delineation and slow down clinical reading. Fig. 1 illustrates two pathways: (A) conventional manual inspection that can be time-consuming for large volumes, and (B) an AI-assisted path that aims to accelerate and standardize analysis at scale. The most popular method for segmenting medical images is deep convolutional networks, particularly U-Net and its variations [2, 3]. However, segmentation per-

^{1,2}The authors are with the Hanoi University of Science and Technology, Hanoi, Vietnam, Email: vibvdkvp@gmail.com and hai.vuduy@hust.edu.vn

^{3,4}The authors are with the East Asia University of Technology, Bac Ninh, Vietnam, Email: 20222072@eaut.edu.vn and hungnv@eaut.edu.vn

⁴Corresponding author: hungnv@eaut.edu.vn.

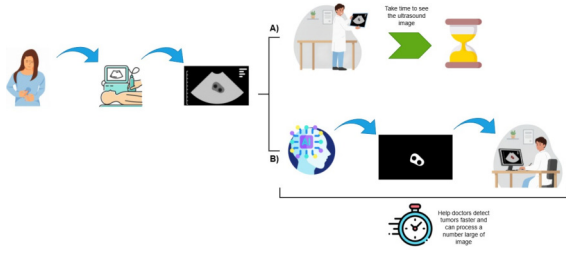


Fig.1: Traditional versus AI-assisted workflow. (A) Manual interpretation of ultrasound images can be time-consuming, particularly in clinical settings with high patient volumes. (B) An AI-assisted workflow automates tumor segmentation, reducing analysis time and improving diagnostic efficiency.

formance may degrade in low-contrast images or under domain shifts commonly encountered in multi-site ultrasound datasets. While test-time augmentation (TTA) aggregates predictions across simple intensity/geometry transforms [4–6], and test-time training (TTT) adapts model parameters on the fly using self-supervision [7], these strategies do not explicitly optimize per-case photometric enhancement to improve the performance of a frozen segmentation model during inference. More recently, studies in medical imaging have explored learning-based image enhancement and test-time adaptation to mitigate domain shifts. For instance, dynamic filtering and localized contrast adjustments have been successfully integrated into deep networks to stabilize feature representations across multi-site ultrasound scans [8, 9]. These sequential and adaptive enhancement frameworks demonstrate that targeted preprocessing can improve prediction robustness without requiring substantial architectural modifications. Related to our problem, reinforcement learning (RL) has been explored in medical image analysis for tasks such as landmark detection, registration, and view planning, showing that sequential decision-making can optimize downstream objectives under limited supervision [10]. However, there is little work that formulates test-time image enhancement itself as a sequential policy optimized by segmentation feedback. This gap is prominent in ovarian ultrasound, where subtle appearance shifts around lesion boundaries can have an outsized impact on voxel-wise overlap metrics.

In this work we propose a per-case, test-time enhancement policy learned with Proximal Policy Optimization (PPO). The agent observes the current image and selects subtle, label-preserving photometric adjustments (e.g., contrast adjustment, gamma correction, and Contrast-Limited Adaptive Histogram Equalization (CLAHE) [11]) to maximize the Dice score of a frozen U-Net without retraining the segmentation model. This design preserves the *in silico* safety of annotations (no geometric warping), exposes a stable reward signal, and yields interpretable action

traces. We evaluate on public ovarian tumor ultrasound data (MMOTU [12]) reflecting the multi-site, variable-appearance reality of this application.

Our contributions are fourfold:

- We define a compact, physically meaningful action space of optical measurements that is label-preserving.
- We demonstrate consistent improvements over the robust U-Net baseline on ovarian ultrasound, reporting significant enhancements and delving into subtle changes in the enhanced images.
- We couple PPO with a ViT-S/16 feature backbone for the policy/value heads, capturing global ultrasound context that is critical for boundary recovery under speckle and low contrast.
- We provide a reproducible evaluation protocol at the case level (Dice & IoU), along with strong segmentation baselines trained on identical splits and trend/ablation plots for fair comparison.

The structure of the paper is as follows. Section 2 reviews relevant literature and positions our work within the existing research landscape. Section 3 introduces the reinforcement learning formulation adopted in this study. Section 4 outlines the dataset, model architecture, and training procedure. Section 5 presents the experimental results along with a detailed discussion. Finally, Section 6 summarizes the key findings and concludes the paper.

2. RELATED WORK

Many studies approach ultrasound segmentation by improving either the image quality before inference or the network used for segmentation. Classical enhancement methods remap intensities or attenuate speckle to make boundaries more visible. Contrastlimited adaptive histogram equalization (CLAHE) raises local contrast while capping amplification to avoid noise blow-up [11]. Speckle-reducing anisotropic diffusion smooths homogeneous regions without erasing edges [13]. Median and bilateral filters play a similar role. These preprocessing pipelines are simple and robust; however, they lack adaptability because the same preprocessing strategy is applied to every image, regardless of the requirements of the downstream segmentation model.

On the model side, U-Net remains the standard encoder-decoder for biomedical segmentation thanks to its skip connections and data efficiency [2]. Transfer learning with residual encoders such as ResNet-34 improves convergence and accuracy on small clinical datasets [14]. In practice, many implementations rely on open libraries (e.g., Segmentation Models PyTorch) that package tested backbones and decoders with consistent preprocessing [15].

A different line of work learns which image operations to apply. In classification, AutoAugment

searches for augmentation policies that maximize validation accuracy, showing that data-driven selection of bounded photometric/geometric edits can beat handcrafted rules [16]. Similarly, test-time adaptation (TTA) methods dynamically adapt model predictions at inference time, providing substantial practical benefits with minimal computational overhead in safety-critical clinical applications [5, 17]. In medical imaging, researchers have applied reinforcement learning (RL) to sequential decision problems such as landmark localization and navigation in volumetric scans, where the agent learns short action sequences with an explicit stopping rule [18]. PPO is widely used in reinforcement learning because its clipped objective promotes stable optimization [21]. Meanwhile, Vision Transformers provide rich global contextual representations that are well suited for policy and value networks [20].

Our work is situated where these threads converge. Rather than retraining the segmentation model or relying on a fixed preprocessing strategy, we keep the U-Net frozen and train a lightweight RL policy to sequentially select bounded, label-preserving photometric adjustments during inference, including contrast adjustment, brightness adjustment, gamma correction, CLAHE, and denoising. This design borrows the safety and simplicity of classical enhancement, the reliability of modern encoders, and the adaptivity of learned policies. The Agent decides when to stop, adds negligible runtime, and preserves anatomy, aligning with clinical constraints while targeting cases where the baseline model struggles.

3. THEORETICAL BASIS

This section presents the proposed PPO-based training framework and the corresponding mathematical formulation for the image enhancement process. We investigate whether the performance of a fixed U-Net [2] segmenter can be improved by applying bounded photometric adjustments to the input image before prediction. This question is particularly relevant to ovarian ultrasound, where low contrast, speckle noise, and poorly defined lesion boundaries can substantially affect segmentation confidence. We formulate test-time enhancement as a short sequential decision process in which the agent applies one label-preserving photometric operation at each step and terminates when no further adjustment is considered beneficial. Throughout this process, the segmentation model remains frozen. This formulation enables case-specific input adaptation without retraining the segmenter while providing interpretable action sequences for subsequent analysis.

We define a compact action space comprising label-preserving photometric and local filtering operations. Actions are selected sequentially, while each operation type can be applied at most once per episode. Let the normalized input image be $I \in [0, 1]^{H \times W}$,

and let J denote the transformed output image.

- **Contrast scaling** (rescale intensity around zero offset):

$$J = \text{clip}(\alpha, I, 0, 1). \quad (1)$$

Where α is the contrast gain (e.g., $\alpha \in [0.85, 1.15]$) and $\text{clip}(\cdot, 0, 1)$ clamps values into $[0, 1]$.

- **Linear brightness** (uniform shift of all pixels):

$$J = \text{clip}(I + \beta, I, 0, 1). \quad (2)$$

Where β is the brightness offset (e.g., $\beta \in [-0.05, 0.05]$).

- **Gamma correction** (nonlinear reweighting of dark vs. bright tones):

$$J = I^\gamma. \quad (3)$$

Where $\gamma < 1$ brightens darker regions and $\gamma > 1$ suppresses them (e.g., $\gamma \in [0.8, 1.2]$).

- **CLAHE** (local histogram equalization with clipping to limit noise amplification):

$$J(p) = (L - 1) \frac{\sum_{k=0}^{I(p)} \min(H_{\Omega_p}(k), c)}{\sum_{j=0}^{L-1} \min(H_{\Omega_p}(j), c)} \quad (4)$$

Where p indexes a pixel; Ω_p is the local window around p ; $H_{\Omega_p}(k)$ is the histogram count at gray level k within Ω_p ; c is the clip limit; and L is the number of gray levels (e.g., $L = 256$). The sum forms the clipped cumulative histogram, normalized to $[0, L - 1]$.

- **Mild denoising/sharpening** (reduce speckle while preserving or enhancing edges). **Median filter:**
- **Median filter:**

$$J(p) = \text{median} \{I(p) | q \in N_k(p)\}. \quad (5)$$

Where $N_k(p)$ is a neighborhood $k \times k$ centered on p .

Bilateral filter:

$$J(p) = \frac{1}{W_p} \sum_{q \in N_k(p)} \exp\left(-\frac{\|p - q\|^2}{2\sigma_s^2} - \frac{(I(p) - I(q))^2}{2\sigma_r^2}\right) I(p). \quad (6)$$

Where σ_s and σ_r control the spatial and range kernels, respectively, and W_p is the normalization term (sum of weights).

Unsharp masking:

$$J = \text{clip}(I + \kappa(I - (G_\sigma * I, 0, 1))). \quad (7)$$

where G_σ denotes a Gaussian kernel with standard deviation σ , $*$ denotes convolution, and $\kappa > 0$ controls the sharpening strength.

All operations are limited to photometric or local contrast adjustments that preserve image geometry. Their magnitudes are kept small, and each operation is applied at most once per episode to avoid over-enhancement.

To guide the Agent, we rely on a stable overlap signal from the segmenter: the soft Dice between its probabilistic mask $p = f\theta(x) \in [0, 1]^{H \times W}$ and the ground truth y (available only during training):

$$\text{Dice}(p, y) = \frac{2 \sum_i p_i y_i + \varepsilon}{\sum_i p_i + \sum_i y_i + \varepsilon} \quad (8)$$

At each step t , the reward encourages progress and discourages long trajectories:

$$r_t = \alpha(d_t - d_{t-1}) - \lambda_{step} + \beta[d_t - \tau]. \quad (9)$$

Where d_t is the current soft Dice, λ_{step} is a small time penalty, and τ is a stopping target. This shaping turns subtle but helpful intensity shifts into a dense learning signal and naturally promotes early stopping once the image is “good enough”.

4. METHODOLOGY

This section presents the data and experimental setup, then details how we train the U-Net model, Agent, and RL environment.

We use the OTU 2D ovarian ultrasound dataset [23]. All images are normalized to $[0, 1]$ and resized to 512×512 pixels. The dataset is split into a development set of 1,000 images and an independent test set of 469 images. We further partitioned the development set into 800 training images and 200 validation images. The validation set was strictly used for hyperparameter tuning of the PPO Agent and for early stopping during the U-Net training phase to prevent overfitting.

On the one hand, Windows 11 Pro, a 12th Gen Intel(R) Core(TM) i7-12700 (2.10 GHz) processor, 477 GB of storage, 32 GB of RAM, and a 12 GB graphics card (NVIDIA GeForce RTX 3060) were the machine specifications used for the trials. On the other hand, PyTorch and Stable-Baselines3 were the primary libraries utilized in this study.

4.1 U-Net Training

We configure the U-Net model using the hyperparameters listed in Table 1. The model employs a ResNet-34 backbone pretrained on ImageNet, with the input image size fixed at 512×512 pixels.

Table 1: U-Net configuration used in our experiments.

Hyperparameter	Value	Note
Architecture	SMP U-Net	Segmentation Models PyTorch
Encoder backbone	ResNet-34 (ImageNet)	encoder='resnet34', pretrained
Encoder depth	5	Downsampling stages
Decoder channels	256-128-64-32	Default SMP decoder plan
Input channels	1	Grayscale ultrasound
Output classes	1	Binary mask
Activation	None	Sigmoid only at evaluation
Upsampling	Transposed conv	Default in U-Net decoder
Skip connections	Enabled	Encoder-decoder concatenation
Normalization	BatchNorm2d	In decoder blocks
Input size	512×512	Resized, no geometry change
Optimizer & LR	Adam, 1×10^{-4}	Initial learning rate

Additionally, we display the training procedure as illustrated in Fig. 2. First, we divide the data into three subsets: training, validation, and testing (train and val will be in the same file as shown in the figure), then the data will be fed into the model to perform the training process (batch size=8, epochs=100), and finally we will obtain the segmentation model for the OTU 2D dataset.

4.2 Our Method Proposed Setup

In this part, we can see that Fig. 3 visually depicts the process we use to train the Agent policy, each iteration of which proceeds as follows:

- Step 1: The 512×512 input image is normalized and then fed into a ViT-S/16 backbone pretrained on ImageNet. To accommodate the larger spatial resolution, the standard ViT-S/16 positional embeddings are bilinearly interpolated to match the input grid size.
- Step 2: The actor outputs a categorical distribution over the available label-preserving actions, including contrast adjustment, brightness adjustment, gamma correction, CLAHE, median filtering, bilateral filtering, unsharp masking, and STOP. During training, the agent samples an action to encourage exploration, whereas during evaluation, the action with the highest probability is selected deterministically.
- Step 3: The selected correction is applied to the input image, producing the image processed by the RL Agent.
- Step 4: The enhanced image is passed to the frozen segmentation model to produce a probabilistic segmentation mask. During training, the soft Dice score of the current prediction is

computed against the ground-truth mask and compared with the score obtained at the previous step. The resulting improvement provides the primary reward signal, together with a small step penalty and an optional bonus when the target quality threshold is reached. The episode terminates when the agent selects the STOP action, the target threshold is achieved, or the maximum number of steps K is reached.

Table 2: Key PPO hyperparameters used for the RL Agent.

Hyperparameter	Symbol/Setting	Value
Policy backbone	ViT-S/16	ImageNet pretrained
Parallel environments	n_envs	8
Rollout length per env	n_steps	256
Total timesteps	—	1.5×10^6
Learning rate	lr	1×10^{-4}
Discount factor	γ	0.99
GAE parameter	λ	0.95
Clip range	—	0.10
Reward coefficients	$\alpha, \beta, \tau, \lambda_{\text{step}}$	1.0, 0.5, 0.95, 0.01

- Step 5: We calculate the advantage/profit and

update the critic-Agent using PPO.

- Finally, during inference, the agent selects actions deterministically. The final segmentation result is obtained when the agent selects the STOP action or reaches the maximum number of steps K . This procedure requires no ground-truth information and therefore supports a causal and deployable inference pipeline. Table 2 and Alg. 1 summarize the main PPO hyperparameters.

Algorithm 1 PPO-based enhancement

Require: Dataset $\mathcal{D}$, frozen segmenter f_w , policy π_θ , value V_ϕ , horizon T

```

1: Action space: {contrast, brightness, gamma, CLAHE, denoise, STOP}
2: State is the current image
3: for  $u = 1, 2, \dots$  do
4:   Sample  $(x, y) \sim \mathcal{D}$ 
5:    $s_0 \leftarrow \text{normalize}(x)$ 
6:   for  $t = 0$  to  $T - 1$  do
7:     Sample  $a_t \sim \pi_\theta(\cdot | s_t)$ 
8:      $s_{t+1} \leftarrow \text{APPLY}(s_t, a_t)$ 
9:     Compute reward  $r_t = R(s_t, s_{t+1}; y, f_w)$ 
10:    if  $a_t = \text{STOP}$  then
11:      break
12:    end if
13:  end for
14:  Compute advantages with GAE( $\gamma, \lambda$ )
15:  Update  $(\theta, \phi)$  using PPO (clip  $\epsilon$ )
16: end for

```

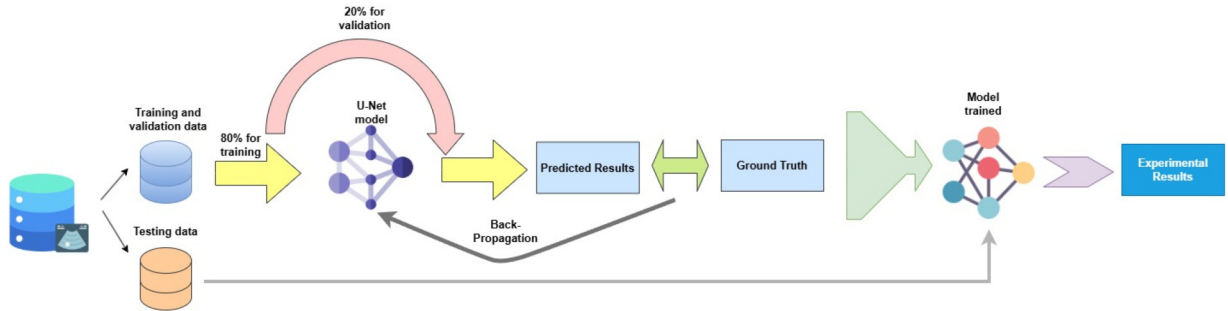


Fig.2: U-Net Training Process.

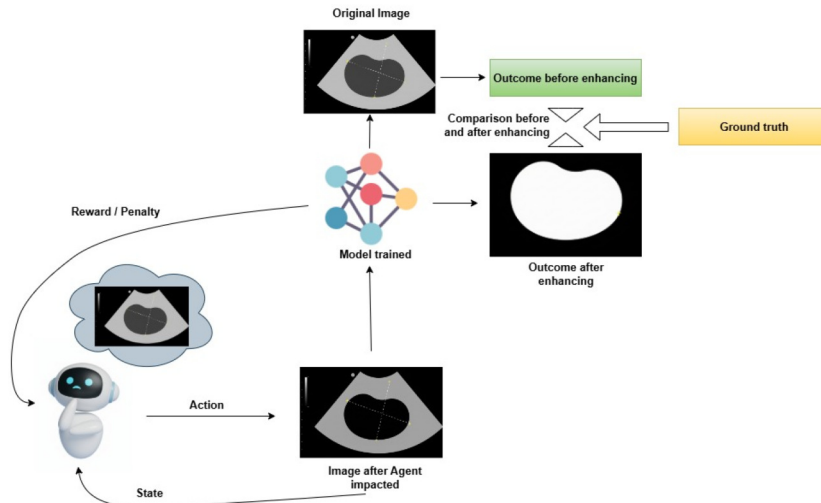


Fig.3: Overview Our Method Process.

4.3 Evaluation Metrics

We use the soft Dice from Eq. 8 as the base measure and report the following per-case and dataset-level quantities.

Baseline Dice (raw image). The soft Dice of the frozen U-Net on the unmodified input:

$$D_{base}^{(n)} = \text{Dice} \left(P_{raw}^{(n)}, Y^{(n)} \right). \quad (10)$$

Agent Dice (at stopping action). The soft Dice of the frozen U-Net on the image produced at the Agent's chosen STOP action (or terminal step T):

$$D_{agent}^{(n)} = \text{Dice} \left(P_T^{(n)}, Y^{(n)} \right). \quad (11)$$

Per-case improvement. The absolute gain achieved by the Agent on case n:

$$\Delta D^{(n)} = D_{agent}^{(n)} - D_{base}^{(n)}. \quad (12)$$

Mean improvement (dataset-level). The average gain across all N cases:

$$\overline{\Delta D} = \frac{1}{N} \sum_{n=1}^N \Delta D^{(n)}. \quad (13)$$

Median improvement (dataset-level). The median of per-case gains:

$$\tilde{\Delta D} = \text{median} \left\{ \Delta D^{(n)} \right\}_{n=1}^N. \quad (14)$$

Fraction of improved cases. The proportion of cases with gains at or above a threshold τ :

$$\rho_\tau = \frac{1}{N} \sum_{n=1}^N \mathbb{1} \left[\Delta D^{(n)} \geq \tau \right], \tau \in \{0.05, 0.10\}. \quad (15)$$

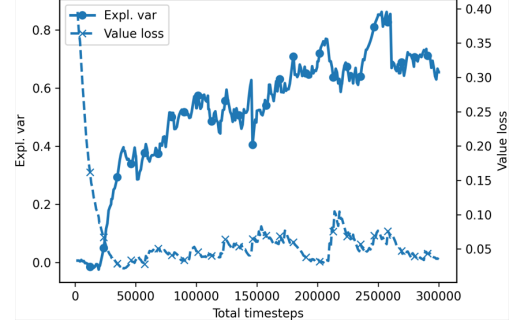
Where n indexes cases ($n = 1, \dots, N$), K is the maximum Agent steps per episode, $P_{raw}^{(n)}$ and $P_t^{(n)}$ denote the frozen U-Net probability maps obtained from the raw image and the image after step t , respectively, and $y_n \in \{0, 1\}^{H \times W}$ denotes the ground-truth mask.

5. RESULTS AND DISCUSSION

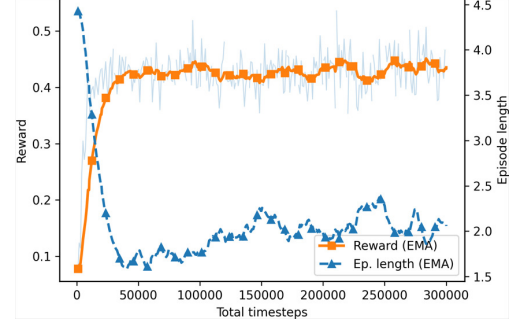
5.1 Experimental Results

This section presents the empirical results and analyzes how the learned policy translates a small set of photometric edits into segmentation gains.

Fig. 4 shows some of the values we measured during Agent training. Fig. 4a shows that the explained variance fluctuates between approximately 0.6 and 0.9, while the value loss remains low after its initial decline. These trends indicate that the critic adequately models the observed returns and that PPO training remains stable. We do not observe signs of collapse or out-of-control updates (e.g., persistent loss



(a) Critic health (explained variance, value loss).



(b) Reward & episode length.

Fig.4: Outcomes during training: (a) Critic health; (b) Reward & episode length.

spikes); together with the smooth reward stability, this supports the reliability of the learned policy. On the other hand, Figure 4b shows the evolution of the training reward and episode length. The reward increases rapidly and stabilizes at approximately 0.40-0.45. Meanwhile, the average episode length initially decreases from about 4-4.5 steps to approximately one step, indicating that the agent learns to stop when further enhancement is unlikely to be beneficial. It subsequently stabilizes at around two steps, suggesting that some images benefit from short action sequences, such as mild gamma correction followed by CLAHE. This pattern confirms that the step penalty is balanced: short trajectories dominate, but the policy still requires an extra operation when it is beneficial.

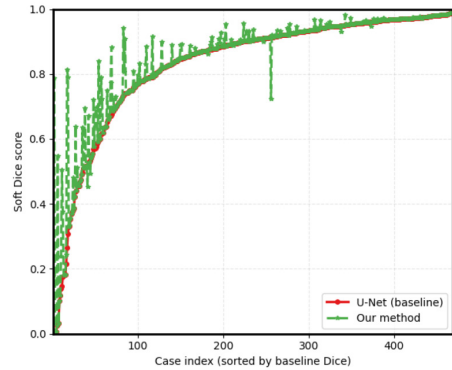


Fig.5: Comparison results between baseline vs Our Agent (Case per case).

Fig. 5 compares the per-case soft Dice scores of the baseline U-Net and the proposed RL-guided enhancement method. We observe that in certain highly challenging cases, the baseline model fails to identify any target regions (yielding a Dice score of 0). However, after enhancement by the agent, the Dice score improves by up to 78% in the most challenging cases, while other cases show consistent gains of at least 1–2%, with no performance degradation relative to the baseline. Overall, the graph shows broad case-specific improvements rather than isolated outliers, indicating that the learned policy generalizes consistently across cases.

Additionally, we add comparisons with other baselines:

- UNet++ extends U-Net with nested, densely connected skip pathways and deep supervision to reduce the semantic gap between encoder and decoder features, thereby improving robustness in medical image segmentation [21].
- FPN constructs a top-down feature pyramid with lateral connections, yielding multi-scale, semantically strong maps at marginal extra cost; it is widely used as a general-purpose encoder for dense prediction [22].
- DeepLabv3+ combines Atrous Spatial Pyramid Pooling for multi-scale context with a lightweight decoder and depthwise separable convolutions, improving boundary quality without heavy post-processing [23].
- PSPNet introduces pyramid pooling to aggregate global scene context across multiple receptive field sizes, boosting pixel accuracy in challenging scenes [24].
- LinkNet is an efficient encoder-decoder with streamlined skip paths that preserves accuracy while reducing parameters and FLOPs, making it attractive for fast inference [25].

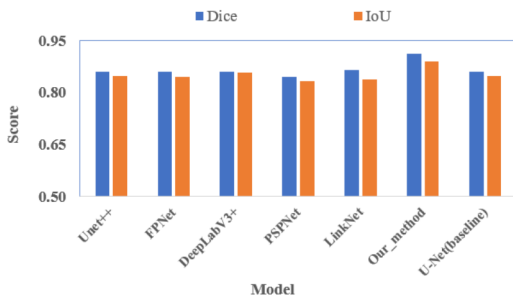


Fig.6: Dice and IoU Performance: Proposed Method vs. Baselines.

To ensure no bias or error during training and evaluation, we train all the models above with the parameters as tuned in Table 1. Fig. 6 reports the average Dice and IoU values on the legacy dataset. Conventional baselines cluster tightly around 0.85 ± 0.02 Dice and 0.82 ± 0.03 IoU, while our RL-guided enhancement achieves 0.91 ± 0.01 Dice and

0.88 ± 0.02 IoU. It is important to note that this is a contextual comparison; it demonstrates that a standard U-Net augmented with our RL preprocessing can outperform more complex architectures (such as UNet++ or DeepLabv3+) that rely solely on standard inputs, highlighting the value of adaptive test-time enhancement. The improvement over the baseline is statistically significant ($p < 0.001$) according to the paired Wilcoxon signed-rank test. This enhancement is consistent across cases, with the largest improvements on low-contrast scans, where the Agent selectively enhances local contrast around the lesion boundary. Notably, we keep the U-Net segmentation backbone fixed; the gain arises solely from learned, label-preserving photometric corrections. These results indicate that controllable preprocessing can recover boundary evidence that standard pipelines fail to exploit, translating into higher overlap scores without compromising stability. We evaluated a nonadaptive baseline consisting of U-Net with a fixed CLAHE preprocessing step (using standard parameters: clip limit 2.0, tile grid 8×8). This static enhancement achieved a Dice score of 0.86 ± 0.02 , which is significantly lower than the 0.91 ± 0.01 achieved by our RL-guided approach. These results show that adaptively sequencing and tuning photometric operations for each case is critical for recovering complex lesion boundaries.

Furthermore, the RL Agent introduces only minimal computational overhead. The average inference time for the full enhancement pipeline is 0.12 ± 0.05 seconds per image, compared to 0.04 ± 0.01 seconds for the baseline U-Net, measured on the NVIDIA GeForce RTX 3060 GPU specified in Section 4.

To address potential skewness from extreme cases (e.g., the +78% gain), we analyzed the per-case improvement across the test set. The median improvement (ΔD) was 0.05, confirming a consistent positive shift rather than outlier-driven artifacts. Furthermore, 52% of cases achieved a Dice gain ≥ 0.05 ($p(0.05) = 0.52$), and 18% improved by ≥ 0.10 ($p(0.10) = 0.18$). This robust distribution underscores the generalized efficacy of our learned policy.

In Fig. 7, we also give an example of the output image after being upgraded by Agent as in Fig. 7a, if we look at it with the naked eye, we will not be able to clearly see the difference between the two photos before and after because the Agent does not actually change the color or shape of the image. Still, it modifies contrast through operations such as CLAHE, which may appear subtle to the human eye, but the deep feature extractor computationally amplifies these photometric modifications during processing. On the other hand, we also provide Fig. 7b, which is the image that helps verify that our Agent performs label-preserving photometric corrections instead of changing the structure, we visualize the mapping between the original image and the enhanced

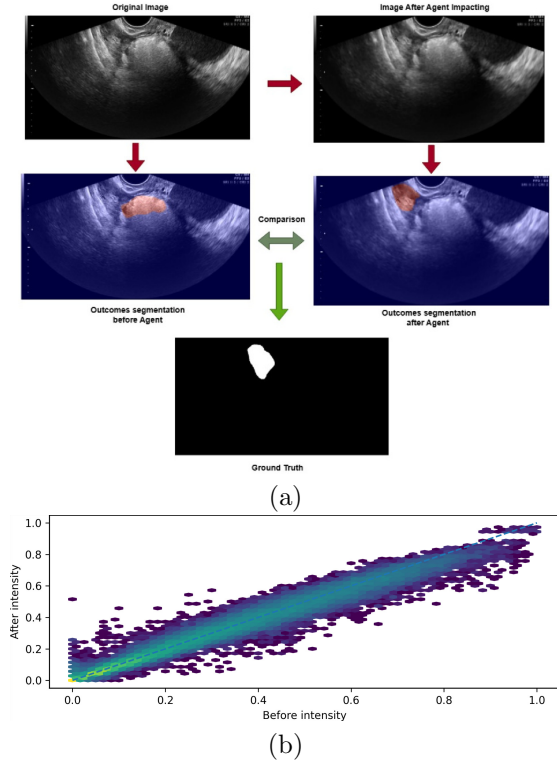


Fig. 7: A typical example of Agent’s photo enhancement results: (a) image before and after Agent impactation; (b) A graph measuring the distribution of joint intensity between the original image and the image after going through the Agent.

image as a distribution of joint intensity (hexbin); the x-axis is the gray level before correction, the y-axis is the gray level after correction, and the dashed line marks the consistency. Most pixels fall within a narrow diagonal band, indicating a monotonic, near-uniform transformation. This band shows a slight upward shift in the midtones ($\approx 0.1 - 0.9$), consistent with a slight brightness/contrast increase, while both poles hug $(0, 0)$ and $(1, 1)$, indicating limited edits without clipping.

5.2 Discussion

In this section, we discuss the practical implications of the results, the reasons underlying the method’s effectiveness, its limitations and potential risks, and considerations related to integration, safety, and generalizability in clinical workflows.

The Agent acts as a lightweight preprocessor, preserving the label before the freeze segmentation, requiring only a few photometric corrections per image. Because the operations are bounded and structure-preserving, anatomical geometry remains unchanged, and the resulting images remain compatible with existing measurement and reporting tools. In fact, the module can be inserted as a toggle before inference or applied automatically when segmentation confidence is low.

During validation (Fig. 5), the enhancement stream consistently tracks and often exceeds the baseline on a case-by-case basis, with the largest qualitative gains on low-contrast or noisy scans. The training curves (Fig. 4a and Fig. 4b) show a stable policy that favors short correction sequences and a well-performing critic. The joint intensity analysis in Fig. 7b reveals a monotonic, near-uniform mapping with a slight upward shift in the midtone range. This transformation enhances boundary contrast for the frozen U-Net without substantially amplifying noise.

The edits are limited in size and type (contrast/brightness/gamma/CLAHE/denoising), which limits over-enhancement and maintains scale. For implementation, we recommend recording each selected action and its parameters so that every prediction is verifiable and reproducible. Displaying the original and enhanced images side by side also helps the radiologist grasp the information.

Degradation can occur in images with severe bright spots, heavy shadows, or extremely blurred boundaries, while global edits can amplify noise. In these cases, tightening the action limit, penalizing the step slightly more aggressively, or using a confidence-based stopping rule will help to minimize overcorrection. Vendor- or transducer-specific preprocessing can also alter the intensity statistics; it is advisable to calibrate the action limit per location.

By design, we decouple this enhancement module from the segmentation architecture; it preserves the anatomical labels and can theoretically serve as a plug-and-play preprocessor for different backbones without retraining them, though we currently validate the approach only on the U-Net architecture. However, intensity distributions vary between scanners and locations, so multi-center validation and stratification by vendor/transducer/preset are needed to assess robustness and potential bias.

Runtime overhead is low (a few scalar operations and one additional pass of the segmenter). The module can run on the CPU if needed and does not require access to patient metadata. For clinical use, we propose:

- Applying only when the initial confidence drops below a threshold;
- Recording the before/after images and the sequence of actions;
- Presenting a simple “revert” control.

Promising extensions include adding lightweight local contrast operators, learning confidence-based stopping criteria, estimating gate edit uncertainty, and evaluating alternative RL or offline RL variants. Beyond 2D, similar ideas could be tested on 3D volumes or other modalities, as long as the monotonic, bounded edit constraint is maintained.

In summary, the experimental evidence supports a practical, low-risk approach to improving segmentation quality in difficult ultrasound cases by learning

when and how much to edit images before freezing the model, without changing anatomy or workflow.

A current limitation of this study is the absence of a stratified performance analysis by tumor pathology, because some categories contain too few samples to support meaningful statistical evaluation. Furthermore, while the proposed framework demonstrates compelling performance gains, several aspects warrant further investigation.

First, a comprehensive ablation study exploring the individual contributions of action subsets (e.g., isolating CLAHE versus sharpening filters), the sensitivity of the reward-shaping coefficients (α, β, τ), and the impact of alternative lightweight policy backbones remains outside the scope of this initial proof-of-concept. Second, although the Agent achieves a significant performance leap on the OTU 2D dataset, cross-dataset validation on external, multi-center cohorts is essential to fully evaluate the policy's generalizability across diverse transducer manufacturers and scanning protocols. Lastly, the current evaluation is limited to technical performance. Future studies should include rigorous clinical validation with radiologist assessments to quantify time savings and changes in diagnostic confidence under real-world clinical conditions.

Consequently, future research will focus on expanding multi-center clinical trials, executing exhaustive ablation experiments, and conducting a rigorous quantitative comparison against advanced Test-Time Augmentation (TTA) and Test-Time Training (TTT) strategies to solidify the clinical feasibility of this real-time enhancement pipeline.

6. CONCLUSION

As stated in Section 1, this study tests the hypothesis that small, label-preserving photometric edits selected by an RL agent can improve the segmentation performance of a fixed U-Net without altering anatomy or retraining the backbone. The Section 5 confirms this expectation (+78% improvement at best, 1-2% increase across all cases on average). The validation trend shows a consistent case-wise lift of the enhanced stream over the baseline; the training curves indicate a stable policy that prefers short edit sequences; and the joint intensity analysis evidences a near-identity, monotonic mapping—precisely the behavior anticipated for safe, structure-preserving enhancement.

This approach is lightweight and practical because it requires no retraining of the segmentation model, introduces minimal computational overhead, and supports verification by recording the selected actions and their parameters. However, performance may degrade on scans with numerous spots or extremely blurry boundaries, and current results lack stratification by tumor pathology. Future work includes adding lightweight local contrast operators

and confidence-based stop rules, conducting multi-center evaluations, and performing rigorous comparisons with standard TTA and TTT strategies. We also intend to provide comprehensive runtime proof to support its use in clinical procedures requiring rapid processing. In short, learning when and how little to edit provides a simple, safe way to exploit segmentation benefits in challenging ultrasound cases without altering subsequent modeling or clinical procedures.

AUTHOR CONTRIBUTIONS

Conceptualization, Quoc-Vi Tran, Nguyen Viet Hung, and Duy-Hai Vu; methodology, Quoc-Vi Tran and Phi Dinh Huynh; software, Quoc-Vi Tran; validation, Quoc-Vi Tran and Nguyen Viet Hung; formal analysis and investigation, Phi Dinh Huynh; data curation, Quoc-Vi Tran; writing—original draft preparation, Quoc-Vi Tran, Nguyen Viet Hung, and Phi Dinh Huynh; writing—review and editing, Nguyen Viet Hung and Phi Dinh Huynh; visualization, Quoc-Vi Tran and Duy-Hai Vu; supervision, Duy-Hai Vu. All authors have read and agreed to the published version of the manuscript.

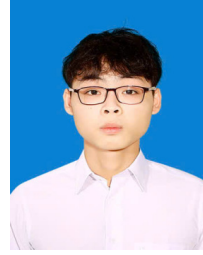
References

- [1] H. Sung, J. Ferlay, R. L. Siegel, M. Laversanne, I. Soerjomataram, A. Jemal and F. Bray, "Global cancer statistics 2020: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries," *CA: A Cancer Journal for Clinicians*, vol. 71, no. 3, pp. 209–249, 2021.
- [2] O. Ronneberger, P. Fischer and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)*, LNCS, vol. 9351, Springer, pp. 234–241, 2015.
- [3] W. Jiangtao, N. I. R. Ruhaiyem and F. Panpan, "A comprehensive review of U-Net and its variants: Advances and applications in medical image segmentation," *IET Image Processing*, vol. 19, no. 1, p. e70019, 2025.
- [4] N. Hung, T. T. Huong, N. Tan, T. Doan and N. Nam-Hoang, "Machine learning applications for delivery time prediction and freight planning," *Informatics and Automation*, vol. 24, no. 5, pp. 1379–1407, 2025.
- [5] N. Moshkov, L. Bonte, N. Vandewiele, J. M. S. Sánchez, W. Alkema, J. Dambre and F. Wyffels, "Test-time augmentation for deep learning-based cell segmentation," *Scientific Reports*, vol. 10, no. 1, p. 5068, 2020.
- [6] P. Conde, C. Premevida and P. Coimbra, "Adaptive-TTA: Accuracy-consistent weighted test time augmentation method for the uncertainty calibration of deep learning classifiers," in *Proceedings of the British Machine Vision Conference (BMVC)*, p. 869, 2022.

- [7] Y. Sun, X. Wang, Z. Liu, J. Miller, A. A. Efros and M. Hardt, "Test-time training with self-supervision for generalization under distribution shifts," in *Proceedings of the International Conference on Machine Learning (ICML)*, PMLR, vol. 119, pp. 9229–9248, 2020.
- [8] N. Karani, E. Erdil, K. Chaitanya and E. Konukoglu, "Test-time adaptable neural networks for robust medical image segmentation," *Medical Image Analysis*, vol. 68, p. 101907, Feb. 2021. [Online]. Available: <http://dx.doi.org/10.1016/j.media.2020.101907>
- [9] R. J. G. van Sloun, R. Cohen and Y. C. Eldar, "Deep Learning in Ultrasound Imaging," in *Proceedings of the IEEE*, vol. 108, no. 1, pp. 11–29, Jan. 2020.
- [10] M. Hu, J. Zhang, L. Matkovic, T. Liu and X. Yang, "Reinforcement learning in medical image analysis: Concepts, applications, challenges, and future directions," *arXiv preprint arXiv:2206.14302*, 2022.
- [11] K. Zuiderveld, "Contrast limited adaptive histogram equalization," in *Graphics Gems IV*, P. S. Heckbert, Ed. Academic Press, pp. 474–485, 1994.
- [12] Q. Zhao, S. Lyu, W. Bai, L. Cai, B. Liu, G. Cheng, M. Wu, X. Sang, M. Yang and L. Chen, "MMOTU: A multi-modality ovarian tumor ultrasound image dataset for unsupervised cross-domain semantic segmentation," *arXiv preprint arXiv:2207.06799*, 2022.
- [13] Y. Yu and S. T. Acton, "Speckle reducing anisotropic diffusion," in *IEEE Transactions on Image Processing*, vol. 11, no. 11, pp. 1260–1270, Nov. 2002.
- [14] K. He, X. Zhang, S. Ren and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Las Vegas, NV, USA, pp. 770–778, 2016.
- [15] P. Reangsuntea, A. Reangsuntea, P. Boonsrimuang, K. Mori and H. Kobayashi, "Iterative Based Time Domain Equalization Method for DFTS-OFD in Highly Mobile Environments," *ECTI-CIT Transactions*, vol. 11, no. 1, pp. 28–39, Jul. 2017.
- [16] E. D. Cubuk, B. Zoph, D. Mané, V. Vasudevan and Q. V. Le, "AutoAugment: Learning Augmentation Strategies From Data," *2019 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Long Beach, CA, USA, pp. 113–123, 2019.
- [17] D. Wang, E. Shelhamer, S. Liu, B. Olshausen and T. Darrell, "Tent: Fully test-time adaptation by entropy minimization," *arXiv preprint arXiv:2006.10726*, 2021.
- [18] F.-C. Ghesu, B. Georgescu, T. Zach, Y. Zheng, J. Petersen and D. Comaniciu, "An artificial agent for anatomical landmark detection in medical images," in *International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI)* pp. 229–237, 2017.
- [19] J. Schulman, F. Wolski, P. Dhariwal, A. Radford, and O. Klimov, "Proximal policy optimization algorithms," *arXiv preprint arXiv:1707.06347*, 2017.
- [20] A. Dosovitskiy *et al.*, "An image is worth 16x16 words: Transformers for image recognition at scale," in *Proceedings of International Conference on Learning Representations*, 2021.
- [21] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support*, vol. 11045, pp. 3–11, 2018.
- [22] T. -Y. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan and S. Belongie, "Feature Pyramid Networks for Object Detection," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 936–944, 2017.
- [23] G. -V. Nguyen and T. Huynh-The, "Enhancing Aerial Semantic Segmentation With Feature Aggregation Network for DeepLabV3+," in *IEEE Geoscience and Remote Sensing Letters*, vol. 21, pp. 1–5, Art no. 2504205, 2024.
- [24] L.-C. Chen, Y. Zhu, G. Papandreou, F. Schroff and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Computer Vision – ECCV 2018*, vol. 11211, pp. 833–851, 2018.
- [25] H. Zhao, J. Shi, X. Qi, X. Wang and J. Jia, "Pyramid Scene Parsing Network," *2017 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, Honolulu, HI, USA, pp. 6230–6239, 2017.



Quoc-Vi Tran received the B.Sc. degree in Electrical Power Systems in 2007 and the M.Sc. degree in Biomedical Engineering in 2013 from Hanoi University of Science and Technology (HUST), Vietnam. He is currently working at Vinh Phuc General Hospital, Vietnam. His research interests include ovarian tumor analysis in ultrasound images, medical image processing, deep learning, computer vision, artificial intelligence, and healthcare applications of intelligent systems.



Phi Dinh Huynh is an undergraduate student at the Faculty of Information Technology, East Asia University of Technology, and a Research Assistant at the Mobilab Laboratory since 2023. His research focuses on computer networking, computer vision, data mining, and intelligent systems. He is particularly interested in integrating autonomous vehicles with AI-driven technologies to enhance smart mobility and intelligent communication systems. With strong enthusiasm for research, she continues to explore the synergy between machine learning, computer vision, and networked systems.



Duy-Hai Vu is an associate professor of biomedical engineering and director of the Biomedical Electronics Center at Hanoi University of Science and Technology. He received his Ph.D. degree from Hanoi University of Science and Technology in 2013 and became an associate professor of biomedical engineering in 2016. He has published more than 60 papers, book chapters, and patents on biomedical engineering. His research interests include biosignal processing and its applications, medical devices and technologies, integrated electronic medical records, medical IoT, and AI for healthcare. Manufacturing equipment used in the COVID-19 pandemic: BKVM-HF1 (HFNC), BK-Vent (Ventilator), BK-NICO (ICG Cardiac Output Measurement Equipment).



Nguyen Viet Hung received the B.Sc. degree in Informatics Pedagogy from the Faculty of Engineering Technology, Ha Tinh University, Vietnam, in 2012, and the M.Sc. degree in Information Technology from the Faculty of Information Technology, Ho Chi Minh City University of Technology, Vietnam, in 2016. He received the Ph.D. degree in Telecommunications Engineering from the Hanoi University of Science and Technology, Vietnam, in 2025. He also received a B.A. degree in English Language from Thai Nguyen University, Vietnam, in 2026. He is currently a Lecturer and the Director of the Information Technology Program at the International Training and Cooperation Institute, East Asia University of Technology, Vietnam, a position he has held since 2025. His research interests include multimedia communications, network security, artificial intelligence, traffic engineering in next-generation networks, machine learning, deep learning, computer vision, data and web mining, autonomous vehicles, QoE/QoS assurance for network services, green networking, and related applications.