scientific reports

Check for updates

OPEN

Pixel level deep reinforcement learning for accurate and robust medical image segmentation

Yunxin Liu[1,2], Di Yuan[1,2], Zhenghua Xu[1,2]✉, Yuefu Zhan[3,4,8,9]✉, Hongwei Zhang[5], Jun Lu[5] & Thomas Lukasiewicz[6,7]

Existing deep learning methods have achieved significant success in medical image segmentation. However, this success largely relies on stacking advanced modules and architectures, which has created a path dependency. This path dependency is unsustainable, as it leads to increasingly larger model parameters and higher deployment costs. To break this path dependency, we introduce deep reinforcement learning to enhance segmentation performance. However, current deep reinforcement learning methods face challenges such as high training cost, independent iterative processes, and high uncertainty of segmentation masks. Consequently, we propose a Pixel-level Deep Reinforcement Learning model with pixel-by-pixel Mask Generation (PixelDRL-MG) for more accurate and robust medical image segmentation. PixelDRL-MG adopts a dynamic iterative update policy, directly segmenting the regions of interest without requiring user interaction or coarse segmentation masks. We propose a Pixel-level Asynchronous Advantage Actor-Critic (PA3C) strategy to treat each pixel as an agent whose state (foreground or background) is iteratively updated through direct actions. Our experiments on two commonly used medical image segmentation datasets demonstrate that PixelDRL-MG achieves more superior segmentation performances than the state-of-the-art segmentation baselines (especially in boundaries) using significantly fewer model parameters. We also conducted detailed ablation studies to enhance understanding and facilitate practical application. Additionally, PixelDRL-MG performs well in low-resource settings (i.e., 50-shot or 100-shot), making it an ideal choice for real-world scenarios.

Keywords Pixel-level optimization, Deep reinforcement learning, Medical image segmentation, A3C strategy

Accurate segmentation of medical images has extensive application and research value in medical research and practice fields, such as clinical diagnosis, pathological analysis, surgery planning, image information processing, and computer-aided surgery[1–4]. Recently, deep learning has been very successful in medical image segmentation, such as cardiac, liver, and spleen segmentation[5–7]. In particular, deep learning methods based on the U-Net-like architecture have gained a notable advantage in the field of medical image segmentation. Consequently, more and more research has focused on enhancing segmentation performance by adding various modules to the U-Net architecture. For example, Attention U-Net[8] integrates attention gates[9] into the expansive path to suppress irrelevant background information. U-Net++[10] introduces nested dense skip pathways to reduce the semantic gap between the feature maps of the encoder and decoder. Swin-Unet[11] replaces traditional convolution operations with Swin Transformer[12]. Although these methods indeed improve segmentation performance to some extent, they rely on incorporating advanced modules and architectures into the model. This not only increases the model parameters and makes deployment challenging but also fails to adequately address the problem of blurry segmentation masks at the boundaries. Besides, the challenge of segmentation tasks is delineating more precise

[1]State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Health Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, China. [2]Tianjin Key Laboratory of Bioelectromagnetic Technology and Intelligent Health, Hebei University of Technology, Tianjin, China. [3]The Third People's Hospital of Longgang District Shenzhen, Shenzhen, China. [4]The Seventh People's Hospital of Chongqing, No. 1, Village 1, Lijiatuo Labor Union, Banan District, Chongqing, China. [5]BigBear (Tianjin) Medical Technology Co., Ltd, Tianjin, China. [6]Institute of Logic and Computation, Vienna University of Technology, Vienna, Austria. [7]Department of Computer Science, University of Oxford, Oxford, United Kingdom. [8]Longgang Institute of Medical Imaging, Shantou University Medical College, Shenzhen, China. [9]Hainan Women and Children's Medical Center, Hainan, China. ✉email: zhenghua.xu@hebut.edu.cn; zyfradiology@hainmc.edu.cn

boundaries, which differs from detection tasks in that identifying the subject is not the primary challenge in segmentation. *Therefore, we consider whether we should abandon the existing popular improvement ideas and explore a more reasonable and feasible strategy for improving segmentation.*

Based on the inspiration that the optimized annotation process of doctors from coarse to fine can help determine clear boundaries, we hope the model to emulate this more reasonable and interpretable process, Therefore, we introduce deep reinforcement learning to model it. Similarly, some researchers have also introduced deep reinforcement learning to iteratively refine segmentation performance. These methods can be roughly divided into two categories: the first is the user interaction by adding new manual label constraints to the segmentation masks to improve segmentation performance[13]. Another one is an iteratively-refined method that optimizes the segmentation mask by slowly changing the classification confidence of a coarse segmentation mask[14]. However, these methods have several problems: (i) **High training cost.** Most of these methods require pre-training a segmentation model to provide rough segmentation masks. Moreover, user interaction methods require experts to manually add hints for each segmentation mask, significantly increasing the model's cost. (ii) **Independent iterative process.** Although these methods are also gradual iterative refinements, they always optimize the segmentation mask of each refinement step in isolation without effectively utilizing global information and surrounding pixel information, and the iterative process is overly lengthy. (iii) **High uncertainty of segmentation mask.** These methods are derived from dense segmentation probability plots through a threshold value to binary prediction results that can lead to quantization errors and loss of accuracy. It can be observed that the above are the current shortcomings of deep reinforcement learning segmentation methods, which may also be the reasons why they are not as popular in segmentation tasks as deep-learning-based methods. *Therefore, we attempt to design a more concise deep reinforcement learning segmentation method to directly generate high-precision segmentation masks.*

To this end, novel proposes a new **Pixel**-level **D**eep **R**einforcement **L**earning model with pixel-by-pixel **M**ask **G**eneration (PixelDRL-MG) using a dynamic iterative update policy, which does not require a user interaction or rough segmentation mask, and can directly segment the regions of interest by inputting the original images. Specifically, to address the issue of high training costs, our model is a direct end-to-end model. We designed a Pixel-level Asynchronous Advantage Actor-Critic (PA3C) for segmentation tasks based on Asynchronous Advantage Actor-Critic (A3C)[15]. In this model, the policy network can directly select actions to change the current agent's state (i.e., whether a pixel belongs to the foreground or background) without any user intervention or coarse segmentation masks. Then, to tackle the problem of independent iterative processes, in PixelDRL-MG, each pixel is considered an agent, and its value represents its current state. The model takes different actions based on the current state to obtain a new state. Through multiple iterations, the segmentation results gradually approach the ground truth. To make full use of the information in each iteration and improve the accuracy of action selection and reward computation, we introduce a Self-Attention Module (SAM) for global information, allowing each agent to gather more global information. For local information, we simply introduce dilated convolutions to help each agent expand its receptive field, considering the states of neighboring pixels, which also makes the model more concise without the need to add extra modules. Finally, to address the issue of high uncertainty of segmentation masks, our method differs from the traditional outputting of the segmentation mask using the probability map with a threshold value, we use the policy network to directly implement the action of setting zero (background) or doing nothing (object) for each pixel, which can alleviate the problems of quantization error and precision loss. Our experimental results also prove that the pixel-by-pixel mask generation method achieves better segmentation performance than the latest U-Net-based deep learning and deep-reinforcement-learning-based methods, especially in terms of segmentation accuracy at the boundaries. Additionally, our model uses significantly fewer parameters.

In summary, this work's main contributions are as follows:

- We identify the limitations in current deep-learning and deep-reinforcement-learning segmentation methods, and then propose a new pixel-level deep reinforcement learning model based on PixelDRL-MG to alleviate these problems and achieve more accurate and robust medical image segmentation.
- In PixelDRL-MG, we designed a Pixel-level Asynchronous Advantage Actor-Critic tailored for segmentation tasks. Each pixel is treated as an agent, and the policy network directly outputs the state of each pixel. Through multiple iterations, the segmented image evolves from coarse to fine-grained segmentation masks.
- Extensive experiments are conducted on two public medical image segmentation datasets, whose results show the effectiveness of PixelDRL-MG. Specifically, our model outperforms common deep learning and deep reinforcement learning methods in terms of segmentation performance with fewer parameters, especially at the segmentation boundaries, with a 3.4% (resp. 3.0%) higher BIoU than the second-best method on the Cardiac dataset (resp. Brain dataset). Furthermore, we validated the effectiveness of our model design through ablation experiments. Finally, we showed that our model can achieve superior segmentation results even on datasets with extremely limited data constraints (i.e., 50-shot or 100-shot).

The rest of this paper is organized as follows. Section 2 briefly reviews related works for deep learning and deep reinforcement learning for medical image segmentation. Section 3 describes our proposed model, including a detailed description of each module. Section 4 shows the details of the applications and experimental results. Section 5 summarizes the social benefits of our method and future research directions. Finally, Section 6 summarizes this paper.

Related work

Deep learning based medica image segmentation

In recent years, deep learning methods have seen growing use in medical image segmentation. This application assists in computer-aided diagnoses, aiding physicians in subsequent evaluations and treatments[7]. For instance, FCN[16] introduces an innovative method employing transpose convolutional up-sampling within a skip architecture for semantic segmentation. U-Net[17] adopts a structure characterized by both contracting and expansive paths, incorporating skip connections to combine deep and coarse features with shallow and fine features. Recent research has yielded several enhanced versions of U-Net. Attention U-Net[8] innovatively integrates attention gates[9] into its expansive path, effectively suppressing irrelevant background information responses and heightening the sensitivity of pancreas features through weight assignment. U-Net++[10] employs a nested, dense skip pathway approach, connecting encoder and decoder sub-networks to reduce the certain semantic gap between their respective feature maps. ResUNet++[18] is specifically designed for colonoscopic image segmentation, incorporating three enhanced modules into the U-Net framework to boost segmentation performance. UNet3+[19] generates intermediate outputs through bilinear up-sampling, facilitating the learning of hierarchical representations from fully aggregated feature maps. nnU-Net [20] is a deep learning-based segmentation method that automatically configures itself, including preprocessing, network architecture, training, and post-processing, to adapt to any new task in the biomedical field. Swin-Unet[11] uses the Swin Transformer[12] block, patch merging layer, and patch expanding layer to build a U-Net-like architecture for medical image segmentation. Although these methods improve segmentation performance to some extent, they rely on stacking better-performing modules or architectures into the U-net model. And TransUNet[21] unites the Transformer as a powerful encoder with the U-Net architecture to improve fine-grained details by restoring local spatial information. However, the performance improvement achieved through this approach is also limited.

Most current deep learning for medical image segmentation is based on U-Net. To show the strength of our PixelDRL-MG in medical image segmentation (including extreme data constraints) without the state-of-the-art improved modules (only adding simple modules), we select eight of the most commonly used representative and relatively new fully supervised U-Net-based models as baselines: FCN[16], U-Net[17], Attention U-Net[8], U-Net++[10], ResUNet++[18], U-Net3+[19], Swin-UNet[11], and TransUNet[21].

Deep reinforcement learning based medical image segmentation

Inspired by the successful application of deep reinforcement learning in playing video games[22], many deep reinforcement learning algorithms applied in medical image analysis[23], especially medical image segmentation[24], have been widely explored and applied[5]. To solve the problem that a threshold is nontrivial to obtain in deep learning that uses image thresholding to create segmentation,[25] and[26] propose to explore the optimal threshold by using deep reinforcement learning.[27] first introduces the concept of context-specific segmentation, making the model adaptable to both the defined objective function and the user's intent and prior knowledge for medical image segmentation. In recent years, more and more work has focused on using an iterative refinement[28] based on deep reinforcement learning to improve the segmentation performance in medical images. For instance,[13] proposes a multi-agent reinforcement learning approach with user interaction introduced. It aims to capture voxel dependencies for medical image segmentation while reducing the exploration space to a manageable size.[14] proposes an end-to-end policy strategy of deep reinforcement learning. This method emulates the progressive delineation of a region of interest (ROI) on medical images, starting from a coarse result and refining it into a finer result, mimicking the behavior of physicians. The model trains an agent to perform ROI segmentation, characterizing the action as a set of continuous parameters, and leverages the policy gradient method to learn how to segment images in a continuous action space. However, these methods are not only not automatic (requiring user interaction or providing coarse segmentation results), but also the iterative process is relatively independent (not considering the global and surrounding pixel information), and the segmentation uncertainty caused by the threshold value is not solved.

Inspired by PixelRL[29], using the asynchronous advantage actor-critic (A3C)[15] as the backbone to achieve denoise, we propose a new pixel-level deep reinforcement learning model using the designed pixel-by-pixel mask generation (PixelDRL-MG) using a dynamic iterative update policy. Although PixelRL does not completely solve the above problems and is not used for medical image segmentation, it enhances pixel-level task performance by incorporating information not only from the target pixel but also from its neighboring pixels. Our PixelDRL-MG has also achieved a significant improvement in the pixel-level segmentation task, because (i) our model not only considers the surrounding pixel information, but also considers the global information, which solves the problem of the independent iterative process of existing methods, (ii) our model does not need to set a threshold, and the policy network directly selects the action of setting zero or doing noting for each pixel, and (iii) our model does not require user interaction or provides coarse segmentation results, and the segmentation results are automatically generated by the policy network, which is more objective.

To prove that our model is more effective than common deep reinforcement learning, we choose five deep reinforcement learning methods, namely, AMP-DRL[24], DQN[30], Double DQN[31], Dueling DQN[32], and AC[33], as baselines; and to prove that our pixel-by-pixel mask generation is more effective than the method of using a threshold to select actions, we apply PixelRL[29] to medical image segmentation in our ablation experiments.

Methodologies

To solve the problems that the current iterative refinement methods based on deep reinforcement learning are not automatic models, are not independent iterative processes, and come with high uncertainty of the segmentation mask relying on the threshold, we propose a pixel-level deep reinforcement learning model with pixel-by-pixel mask generation (PixelDRL-MG) using a dynamic iteration update policy. As shown in Fig. 1,

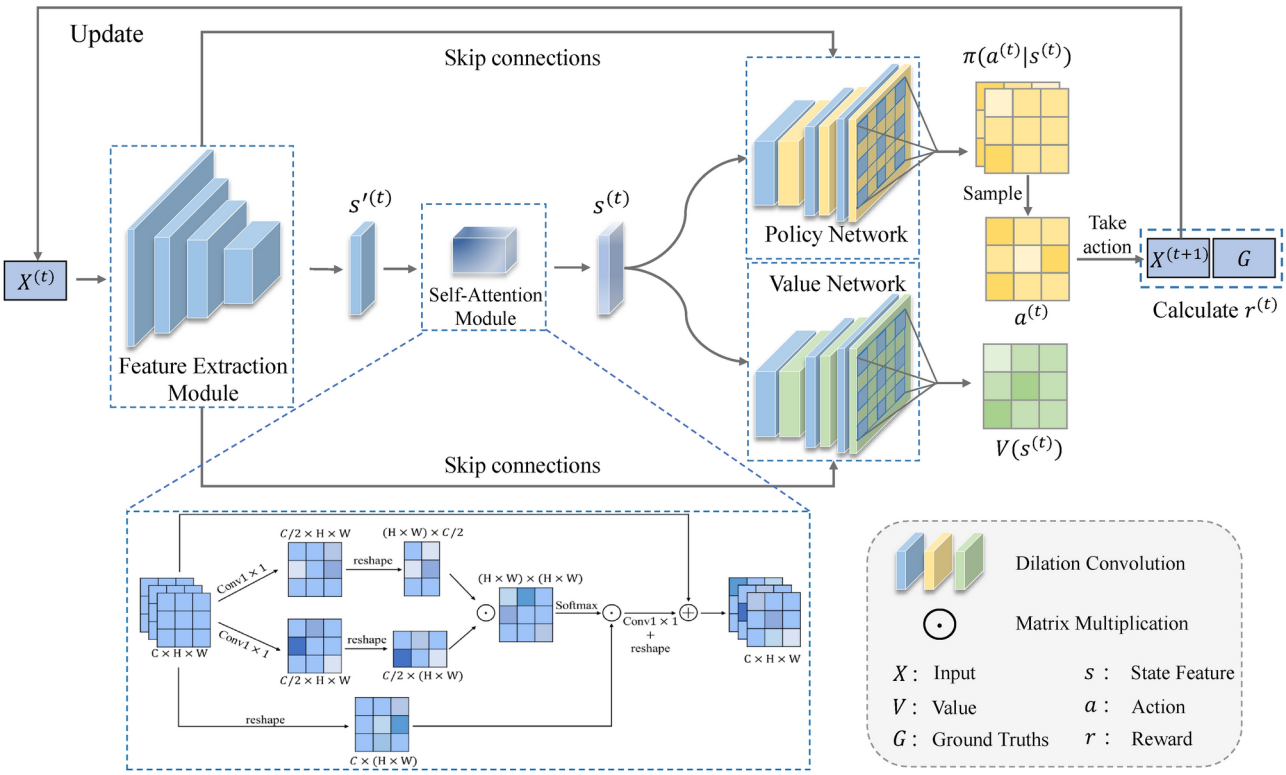

Fig. 1. The framework of the proposed PixelDRL-MG. The PixelDRL-MG architecture is based on PA3C combining VGG16, SAM, and DC. $X^{(t)}$ is the temporary input in step t, G is the label. $a^{(t)}$ is sampled from the policy $\pi : a^{(t)} \sim \pi(a^{(t)}|s^{(t)})$.

PixelDRL-MG combines a feature extraction module (we use VGG16[34] here; theoretically, any module that can extract features can be used) to extract image feature information, a self-attention module (SAM) to obtain more global information, a dilation convolution[35](DC) to obtain surrounding pixel information, and the Pixel-level Asynchronous Advantage Actor-Critic (PA3C) to directly select the optimal action. In PixelDRL-MG, the input image $X^{(t)}$ at time step t is first fed into the feature extraction module to obtain image information $s'^{(t)}$, and then it is passed through SAM to further obtain global information $s^{(t)}$, and then the feature map containing global information $s^{(t)}$ is input to the policy network and value network that contain DC in each layer, respectively, and finally, the policy network directly selects the action of setting zero (background) or doing nothing (object) according to the current state of each pixel to update the pixel state. The model goes through multiple iterations until the segmentation map approaches the ground truths.

PixelDRL-MG architecture

Our PixelDRL-MG is designed based on Pixel-level Asynchronous Advantage Actor-Critic (PA3C), whose structure is similar to Asynchronous Advantage Actor-Critic (A3C), but we can replace PA3C with any backbone based on deep reinforcement learning such as Deep-Q-Network-based and Actor-Critic-based backbones. Here, we mainly introduce the structure of the proposed PixelDRL-MG, including the feature extraction module (i.e., VGG16), SAM for obtaining global information and the PA3C framework.

Feature extraction module

We choose the VGG16 with half the channel count and dilation convolutions as the forward feature map extractor for its strong feature representation capabilities and compatibility with the subsequent two networks for concatenation. Specifically, the initial 23 layers, ranging from conv1-1 to conv4-3, are utilized to generate feature maps scaled to 1/2, 1/4, 1/8, and 1/16 of the original input size. Subsequently, three source layers (i.e., conv1-2, conv2-2, and conv3-3), containing multi-scale features and diverse semantic information, are concatenated by both the policy and value networks after the self-attention module.

Self-attention module

Inspired by self-attention[36], we add a self-attention module (SAM) after the feature extraction module to enhance image feature information, emphasize object features within each image, and expand the receptive field, thereby enhancing network performance. Given that the original image exhibits more pronounced image features, richer object details, accurate pixel similarities and differences, and a consistent distribution of image features, we employ self-attention to capture long-distance dependency relationships in the original image. The structure of SAM is shown in Fig. 1.

Specifically, if f_{att} are the convolutional features output by the feature extraction module, then the attention map M_{att} is:

$$M_{att} = Softmax(W \copyright f_{att} + f_{att}), \tag{1}$$

where W represents the weights of the 1×1 convolution layer, $\copyright$ denotes the convolution operation, and $Softmax$ corresponds to the softmax activation function.

Pixel-level asynchronous advantage actor-critic

We extend the asynchronous advantage actor-critic (A3C) for the problems of current iterative refinement methods on medical image segmentation. It can be well aimed at pixel-level tasks in images, which we call Pixel-level Asynchronous Advantage Actor-Critic (PA3C). Moreover, we introduce the dilation convolution[37] in each layer of both networks, to solve the problem of image resolution reduction and information loss caused by down-sampling, and it can also capture the state of its neighboring pixels. Here, we provide a concise overview of the training algorithm for A3C, which is an actor-critic method employing a policy network and a value network. The policy network encourages the agent to take better actions, and the value network scores more accurately based on status. The parameters of these networks are denoted as θ_p and θ_v, respectively. Both two networks utilize the current state $s^{(t)}$ at time step t, which is the feature map containing global information, as their input. Then the value network produces the value $V(s^{(t)})$, which reflects the anticipated total rewards for the state $s^{(t)}$ and provides insight into the desirability of the current state. The computation of the gradient for the value network parameters θ_v is as follows:

$$R^{(t)} = r^{(t)} + \gamma r^{(t+1)} + \gamma^2 r^{(t+2)} + \cdots + \gamma^{(n)} r^{(t+n)}, \tag{2}$$

$$d\theta_v = \bigtriangledown_{\theta_v} (R^{(t)} - V(s^{(t)}))^2, \tag{3}$$

where γ represents the discount factor. Besides, the policy network generates the policy $\pi(a^{(t)}|s^{(t)})$ for selecting action $a^{(t)} \in \delta$. As a result, the policy network has output channels equal to $|\delta|$. The computation of the gradient for the policy network parameters θ_p is as follows:

$$A(a^{(t)}, s^{(t)}) = R^{(t)} - V(s^{(t)}), \tag{4}$$

$$d\theta_p = -\bigtriangledown_{\theta_p} \log \pi(a^{(t)}|s^{(t)}) A(a^{(t)}, s^{(t)}). \tag{5}$$

where $A(a^{(t)}, s^{(t)})$ is the advantage function, which represents the advantage of taking an action in the current state relative to the average expectation, and $V(s^{(t)})$ is subtracted to reduce the variance of the gradient.

On the basis of A3C, we propose PA3C, which improves the policy and the value network, so that it can directly act on each pixel in the image, unlike other methods[29,38] that can only fine-tune pixel values. Our model and action design enable us to directly generate pixel values for segmentation. To achieve this, we keep the output resolution of the policy and value networks the same as the input image. In the value network, each pixel value in the output value $V(s^{(t)})$ corresponds to the input image one-to-one, where each pixel value represents the state value of each pixel at this time. In the policy network, the resolution of the output action $a^{(t)}$ is also consistent with the input image, but the channel is different, and the number of channels is determined by the number of optional actions. However, to prevent an excessive increase in the number of model parameters and algorithm complexity, we keep the dimension of the action relatively small. For the medical image segmentation, we only design two actions, that is, set to zero (i.e., background) or doing nothing (i.e., the model's output is denoted as the segmented object), instead of the traditional setting threshold range to adjust the value of each pixel. The motivation for this modification is (i) to ensure the objectivity of the outputs, thereby preventing further calculation loss and segmentation error caused by artificially setting and adjusting the threshold; and (ii) only designing two actions can also reduce the calculation cost. Consequently, the output channels of the policy network $|\delta|$ is 2 under this setting. In the output action $a^{(t)}$, the same pixel position under different channels represents the probability of each action selected by this pixel.

Input: Global counter T_{max}; thread step counter t_{max}; training original images and labels (X, G); outputs of segmentation Y; discount rates γ; and segmentation network M parameterized with θ_s.
Output: M.

Sample a training batch (x, g) from the (X, G) pool
Assume global shared parameter vectors θ_p, θ_v, and θ_s
Assume thread-specific parameter vectors θ'_p, θ'_v, and θ'_s
Initialize thread step counter $t \leftarrow 1$
for $T = 1, \ldots, T_{max}$ **do**
  Reset gradients: $d\theta_p \leftarrow 0$, $d\theta_v \leftarrow 0$, and $d\theta_s \leftarrow 0$
  Synchronize thread-specific parameters $\theta'_p = \theta_p$, $\theta'_v = \theta_v$, and $\theta'_s = \theta_s$
  $t_{start} = t$
  Obtain state $s_i^{(t)}$ for $\forall i$
  for $t = t_{start}, \ldots, t_{max}$ **do**
    Perform $a_i^{(t)}$ based on policy $\pi(a_i^{(t)}|s_i^{(t)}; \theta'_p)$ for $\forall i$
    Accumulate reward $r_i^{(t)}$ and receive new state $s_i^{(t+1)}$ for $\forall i$
    $t \leftarrow t+1,\ T \leftarrow T+1$
    Obtain the outputs of segmentation y corresponding to x
    Calculate the segmentation metrics between y and g
  end for
  for $\forall$ i $R_i = \begin{cases} 0 & for \quad terminal\ s_i^{(t)} \\ V(s_i^{(t)}; \theta'_v) & for \quad non-terminal\ s_i^{(t)} \end{cases}$
  for $n \in \{t-1, \ldots, t_{start}\}$ **do**
    Calculate a discount reward $R_i \leftarrow r_i^{(n)} + \gamma R_i$
    Accumulate gradients of the policy network w.r.t. θ'_p:
    $d\theta_p \leftarrow d\theta_p - \bigtriangledown_{\theta'_p} \frac{1}{N} \sum_{i=1}^{N} \log \pi(a_i^{(n)}|s_i^{(n)}; \theta'_p)(R_i - V(s_i^{(n)}; \theta'_v))$
    Accumulate gradients of the value network w.r.t. θ'_v:
    $d\theta_v \leftarrow d\theta_v + \bigtriangledown_{\theta'_v} \frac{1}{N} \sum_{i=1}^{N} (R_i - V(s_i^{(n)}; \theta'_v))^2$
    Accumulate gradients of the segmentation network w.r.t. θ'_s:
    $d\theta_s \leftarrow d\theta_s - \bigtriangledown_{\theta_s} \frac{1}{N} \sum_{i=1}^{N} \log \pi(a_i^{(n)}|s_i^{(n)}; \theta_s)(R_i - V(s_i^{(n)}; \theta'_v)) + \bigtriangledown_{\theta_s} \frac{1}{N} \sum_{i=1}^{N} (R_i - V(s_i^{(n)}; \theta'_v))^2$
  end for
  Update the parameters θ_p, θ_v, and θ_s using $d\theta_p$, $d\theta_v$, and $d\theta_s$, respectively.
  Update M using gradient descent w.r.t. θ_s
end for

Algorithm 1. Training PixelDRL-MG Using a Dynamic Iterative Update Policy

Dynamic iterative update policy

In this subsection, we outline our dynamic iterative update policy. We represent the i-th pixel in the medical image I (with a total of N pixels, where $i = 1, 2, \ldots, N$) as I_i. Each pixel in the image is an agent, characterized by a policy denoted as $\pi_i(a_i^{(t)}|s_i^{(t)})$, where $a_i^{(t)} \in A$ and $s_i^{(t)}$ represent the action and state of the i-th agent at time step t. In this work, A signifies the action set (i.e., 0 representing the background and doing nothing representing keeping the model output), and $s_i^{(0)} = I_i$. The agents receive subsequent states $s^{(t+1)} = (s_1^{(t+1)}, \ldots, s_N^{(t+1)})$ and rewards $r^{(t)} = (r_1^{(t)}, \ldots, r_N^{(t)})$ by executing actions $a^{(t)} = (a_1^{(t)}, \ldots, a_N^{(t)})$. The aim of pixel-level deep reinforcement learning is to learn the optimal policies $\pi = (\pi_1, \ldots, \pi_N)$ to maximize the mean of the total expected rewards across all pixels of the medical image:

$$\pi^* = \underset{\pi}{argmax}\ \ E_\pi\Big(\sum_{t=0}^{\infty} \gamma^t\ \overline{r}^{(t)}\Big), \tag{6}$$

$$\overline{r}^{(t)} = \frac{1}{N}\ \sum_{i=1}^{N} r_i^{(t)}, \tag{7}$$

$\overline{r}$ stands for the mean of the rewards $r_i^{(t)}$.

Pixel-level deep reinforcement learning has an extremely large number of agents N ($> 10^5$). As a result, typical multi-agent learning solutions[39] are not directly applicable to our work. Additionally, these agents are arranged on a 2D image plane. To enhance agent performance, we introduce the dynamic iterative update policy in our work. Here we discuss the one-step learning case (i.e., $n = 1$ in original A3C) for ease of understanding.

Employing dilation convolutions to expand the receptive field and enhance network performance, our policy network and value network consider the current *i*-th pixel $s_i^{(t)}$ and neighboring pixels when generating the policy π and value V at the *i*-th pixel. Namely, the action $a_i^{(t)}$ influences the state $s_i^{(t+1)}$ and the policy and value within $N(i)$ (representing the local window centred around the *i*-th pixel) at the subsequent time step. Consequently, the gradients for the policy network and value network are calculated as follows:

$$R_i^{(t)} = r_i^{(t)} + \gamma \sum\nolimits_{j \in N(i)} \omega_{i-j} V(s_j^{(t+1)}), \tag{8}$$

$$d\theta_v = \bigtriangledown_{\theta_v} \frac{1}{N} \sum\nolimits_{i=1}^{N} (R_i^{(t)} - V(s_i^{(t)}))^2, \tag{9}$$

$$A(a_i^{(t)}, s_i^{(t)}) = R_i^{(t)} - V(s_i^{(t)}), \tag{10}$$

$$d\theta_p = -\bigtriangledown_{\theta_p} \frac{1}{N} \sum\nolimits_{i=1}^{N} \log \pi(a_i^{(t)}|s_i^{(t)}) A(a_i^{(t)}, s_i^{(t)}), \tag{11}$$

where ω_{i-j} represents the weight that signifies the extent to which we take into account the values V of neighboring pixels at the subsequent time step $(t+1)$. ω values are essentially convolution filter weights, which can be learned concurrently with the parameters θ_p and θ_v. The gradient computation for each network parameter involves averaging gradients across all pixels, where each pixel *i* is identified by a 2D coordinate.

In deep reinforcement learning, the design of the reward function is particularly crucial as it determines the specific direction for model optimization. In this task, we aim for the reward function to guide the model toward generating progressively better segmentation results. In other tasks, the most straightforward approach is to reward the model with $+1$ when it performs well and -1 when its performance is poor. However, unlike game tasks, there are no direct evaluation criteria in this task. Therefore, we design the reward function such that when the current segmentation result is better than the previous one, a positive reward is provided, and conversely, a negative reward is given when the segmentation result worsens. Additionally, if the reward is simply a constant, it would not be conducive to encouraging the model to make significant progress, nor would it effectively penalize severe errors. Therefore, for the reward $\mathrm{r}^{(t)}$ at each step *t*, it is calculated by the difference between the segmentation map of the previous step $f^{(t-1)}$ and the ground truth *G*, subtracted by the difference between the segmentation map of the current step $f^{(t)}$ and the ground truth *G*. This design both balances the rewards for different actions and encourages the model to optimize in a more favorable direction. The specific formula is as follows:

$$\mathrm{r}^{(t)} = ||f^{(t-1)} - G||^2 - ||f^{(t)} - G||^2 \tag{12}$$

Since we treat each pixel as an independent agent, we redefine Eqs. (10) to (12) in matrix form as follows:

$$d\theta_v = \bigtriangledown_{\theta_v} \frac{1}{N} 1^T \{(\mathrm{R}^{(t)} - \mathrm{V}(\mathrm{s}^{(t)})) \odot (\mathrm{R}^{(t)} - \mathrm{V}(\mathrm{s}^{(t)}))\} 1, \tag{13}$$

$$\mathrm{A}(\mathrm{a}^{(t)}, \mathrm{s}^{(t)}) = \mathrm{R}^{(t)} - \mathrm{V}(\mathrm{s}^{(t)}), \tag{14}$$

$$d\theta_p = -\bigtriangledown_{\theta_p} \frac{1}{N} 1^T \{\log \boldsymbol{\pi}(\mathrm{a}^{(t)}|\mathrm{s}^{(t)}) \odot \mathrm{A}(\mathrm{a}^{(t)}, \mathrm{s}^{(t)})\} 1, \tag{15}$$

where $\mathrm{R}^{(t)}, \mathrm{r}^{(t)}, \mathrm{V}(s^{(t)}), \mathrm{A}(\mathrm{a}^{(t)}, \mathrm{s}^{(t)})$ and $\boldsymbol{\pi}(\mathrm{a}^{(t)}|\mathrm{s}^{(t)})$ are the matrices whose (i_x, i_y)-th elements are $R_i^{(t)}, r_i^{(t)}, V(s_i^{(t)}), A(a_i^{(t)}, s_i^{(t)})$ and $\pi(a_i^{(t)}|s_i^{(t)})$, respectively. 1 represents a vector filled with ones, where each element equals one. The symbol $\odot$ represents element-wise multiplication.

Similarly to the gradients of θ_p and θ_v, the gradient using the matrix form for the parameters of segmentation network θ_s is computed as follows:

$$\begin{aligned} d\theta_s = &-\bigtriangledown_{\theta_s} \frac{1}{N} 1^T \{\log \boldsymbol{\pi}(\mathrm{a}^{(t)}|\mathrm{s}^{(t)}) \odot \mathrm{A}(\mathrm{a}^{(t)}, \mathrm{s}^{(t)})\} 1 \\ &+ \bigtriangledown_{\theta_s} \frac{1}{N} 1^T \{(\mathrm{R}^{(t)} - \mathrm{V}(\mathrm{s}^{(t)})) \odot (\mathrm{R}^{(t)} - \mathrm{V}(\mathrm{s}^{(t)}))\} 1. \end{aligned} \tag{16}$$

Much like conventional policy gradient algorithms, the initial component of $d\theta_s$ promotes an increased expected total reward, while the second component serves as a regularization factor to prevent the deviation of R_i from the predicted $V(s_i^{(t)})$ by the convolution operation. In contrast to traditional deep reinforcement learning methods, which are primarily designed for gaming environments (such as deep Q-learning and actor-critic algorithms), traditional deep reinforcement learning approaches are tailored for games that inherently possess well-defined reward functions and action spaces. However, this is not the case in other domains, which is why deep reinforcement learning is rarely applied to image processing tasks, particularly in the context of medical image segmentation. The challenge in medical image segmentation lies in delineating more precise boundaries (i.e., achieving accurate pixel-level segmentation). Therefore, although we adopt the intuition and mechanisms

of the original A3C algorithm, the A3C used in this study has been redesigned to better suit the requirements of medical image segmentation tasks.

Specifically, we first designed a reasonable reward function based on the medical segmentation task to facilitate algorithm convergence and encourage better segmentation performance. Secondly, we defined new actions according to the requirements of the segmentation task (i.e., whether a pixel belongs to the foreground or background). Finally, we proposed a novel agent definition strategy, where each patch is treated as an independent agent, rather than treating the game player as the agent, as in the original A3C. This approach helps the model more accurately determine which pixels belong to the segmentation target and allows them to take independent actions, thereby achieving precise pixel-level segmentation. Algorithm 1 summarizes the training processes of the proposed PixelDRL-MG. Figure 2 illustrates the flowchart of the proposed PixelDRL-MG algorithm.

Experiments

We have extensively evaluated our proposed PixelDRL-MG through a series of experiments. We first provide information about datasets, experimental settings, evaluation metrics, and baselines. Then, to verify that our model utilizes a pixel-level mask generation to improve the segmentation performance more effectively than utilizing the state-of-the-art modules, we conduct an extensive experimental study to compare the performance of PixelDRL-MG with deep-learning-based and deep-reinforcement-learning-based methods. Additionally, we conduct ablation studies to further demonstrate the necessity and effectiveness of each module in our model. Finally, to further demonstrate our model's robustness, we execute supplementary experiments under extreme data constraints.

Datasets

We evaluated our model using two public medical image segmentation datasets, which are shown in Table 1. These datasets exhibit features such as small size, small objects, and intricate segmentation details, so they are more reflective of the attributes found in contemporary medical images.

Cardiac[40] is a public MRI dataset for segmenting the heart, comprising 20 cases. The images within each case are of dimensions 320×320, encompassing a varying slice count ranging from 90 to 130. The challenge in segmenting this dataset stems from its limited size and the substantial variations in the segmentation objects.

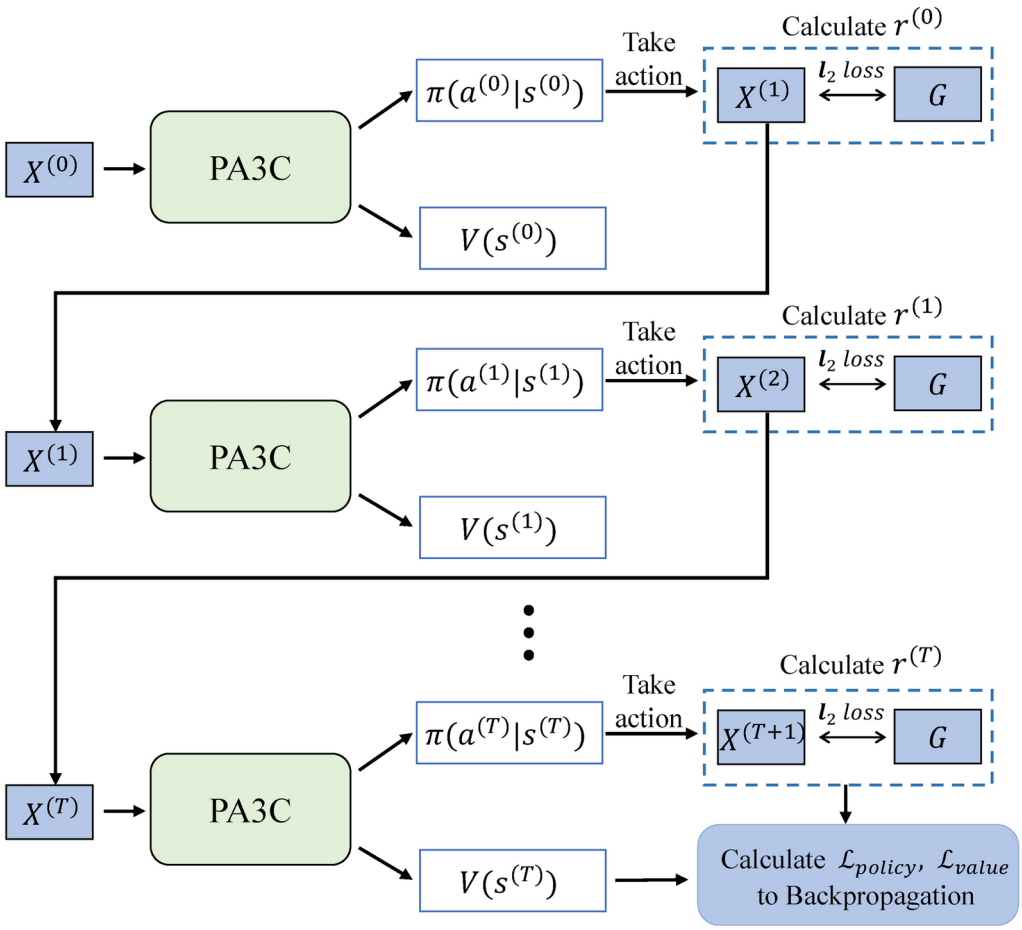

Fig. 2. Flowchart of the proposed PixelDRL-MG algorithm..

Datasets	Quantity	Image size	Modality	Challenge	Source
Cardiac[40]	2,271	320×320	MRI	Small objects with large variability	King's College London
Brain[41]	3,929	256×256	MRI	Extremely irregular segmentation edges	The Cancer Imaging Archive

Table 1. Dataset Information.

Brain[41] is a publicly available MRI dataset designed for automated tumor segmentation, comprising 110 cases. Segmentation masks are validated by a board-certified radiologist at Duke University. Each case varies in its scanning mechanism, resulting in images of dimensions 256×256 with a slice count ranging from 40 to 176. The segmentation task on this dataset is notably challenging due to the presence of intricate segmentation details in the object boundaries.

To train and evaluate our proposed method, We perform simple preprocessing on two medical image datasets as follows. First, the Cardiac dataset's 3D images are converted into 2D images by taking transverse sections in a slice-by-slice manner [42]. Then, negative samples, which lack segmentation objects, are removed from both datasets. This results in the utilization of 1, 293 Cardiac images and 1, 168 Brain images in our experiments. Finally, the data for both datasets is partitioned into training (70%), validation (10%), and testing (20%) sets.

Implementation details

Our experiments are implemented using PyTorch and run on an NVIDIA GeForce GTX 3090 GPU. We evaluate our model on two public datasets using the same experimental setup, including network architecture, and training hyperparameters. Below are the implementation details of the proposed PixelDRL-MG. The first 23 layers of VGG16 use 3×3 dilation convolution and the number of channels is half of the original and is applied as the front-end feature extractor. We use the popular optimizer *Adam*[43] to train the proposed model, where the learning rates used in Adam are set to $1e-3$, the learning rate drops by a factor of 0.9 every 25 epoch. Due to the machine's GPU memory limitation, the batch size is set to 2. We set the maximum epoch to 200, the length of each episode t_{max} to 10 and the discount rate γ to 0.95. For all deep-learning-based baselines, we use the settings following the original papers, among them, Swin-Unet uses swin-tiny in the original paper; and for all deep-reinforcement-learning-based baselines, we choose the U-Net with k kernels (where $k = 32 \times 2^i$, and i={1,2,3,4} indexes the down-sampling layer along with the decoder) as the policy network and value network.

Evaluation

To demonstrate the efficacy of our models, we employ several evaluation metrics, including the Dice coefficient (DICE), Positive Predictive Value (PPV), Sensitivity (SEN), Intersection over Union (IoU), 95% Hausdorff Distance (HD95)[44], and boundary IoU (BIoU)[45]. Specifically, DICE, which evaluates the overlap between predictions and ground truths, is a comprehensive metric combining SEN and PPV. PPV calculates the proportion of true positive samples among all predicted positive samples. SEN measures the likelihood of correctly classifying positive samples. IoU is a standard for assessing the accuracy of object correspondence within a given dataset. Boundary IoU (BIoU) is designed to measure the overlap of segmentation boundaries. Hausdorff Distance (HD), a frequently employed distance-based metric, is utilized in our study to mitigate the influence of a minute subset of outliers. Note that higher values for these metrics, except HD95, mean better performance. Formally,

$$
\begin{aligned}
DICE &= \frac{2 * TP + \epsilon}{T + P + \epsilon}, \qquad PPV = \frac{TP + \epsilon}{TP + FP + \epsilon}, \\
SEN &= \frac{TP + \epsilon}{TP + FN + \epsilon}, \quad IoU = \frac{TP + \epsilon}{T + P - TP + \epsilon}, \\
BIoU &= \frac{G_d \cap P_d}{G_d \cup P_d}, \\
HD95 &= max_{k95\%}[d(P, G), d(G, P)],
\end{aligned}
$$

where *TP* for true positive points, *FP* for false positive points, and *FN* for false negative points. *T* represents the number of ground-truth points for a specific class, while *P* is the count of predicted positive points. Additionally, *G* stands for the number of ground-truth positive points, P_d denotes the quantity of predicted positive boundary points, and G_d represents the number of ground-truth positive boundary points. The function $d(*)$ calculates the surface distance, and ϵ is a small constant (set to $1e-4$) used to prevent zero division.

Baselines

To evaluate the performances of the proposed PixelDRL-MG, the eight common U-Net-based deep learning image segmentation methods are selected as baselines. (i) **FCN**[16] employs transpose convolutional up-sampling within a skip architecture. (ii) **U-Net**[17] is the most commonly used and most influential medical image segmentation model, and (iii) Many works are improvements of U-Net. **Attention U-Net**[8] with attention gates, **ResUNet++**[18] with three improving modules, **U-Net++**[10] with a series of nested and dense skip pathways, **U-Net3+**[19] with bilinear up-sampling, **nnU-Net**[20] automates preprocessing and post-processing according to different task, **TransUNet**[21] combines Transformer as a powerful encoder, and **Swin-UNet**[11] uses the Swin Transformer[12] block, patch merging layer, and patch expanding layer to build a U-Net-like architecture.

Model	Cardiac						Brain						Params ↓
	DICE ↑	PPV ↑	SEN ↑	IoU ↑	BIoU ↑	HD95 ↓	DICE ↑	PPV ↑	SEN ↑	IoU ↑	BIoU ↑	HD95 ↓	
FCN	0.7259	0.6547	0.6155	0.6231	0.2177	8.9765	0.6269	0.7186	0.5786	0.5376	0.1800	14.6973	128.95M
U-Net	0.7586	0.7467	0.7330	0.6691	0.2428	7.6417	0.7220	0.7783	0.6734	0.6022	0.1932	13.2416	31.03M
Attention U-Net	0.7889	0.8014	0.8078	0.6857	0.2961	6.2739	0.7536	0.7956	0.6940	0.6402	0.2079	12.2775	32.43M
ResUNet++	0.7714	0.7774	0.8278	0.6839	0.2816	7.0957	0.7462	0.8095	0.6994	0.6330	0.2029	11.5932	14.48M
U-Net++	0.8177	0.8242	0.7883	0.7232	0.3515	4.9647	0.7781	0.8539	0.7114	0.6789	0.2360	10.2009	47.17M
U-Net3+	0.8250	0.8491	0.7977	0.7470	0.3746	4.8439	0.7952	0.8731	0.7465	0.6918	0.2413	10.1787	27.03M
nnU-Net	0.8305	0.8511	0.8185	0.7484	0.3614	4.9464	0.7988	0.8614	0.7285	0.7013	0.2397	11.1758	30.50M
SwinUnet	0.8290	0.8439	0.8314	0.7447	0.3422	5.3375	0.7910	0.8513	0.7328	0.6955	0.2276	11.3980	27.16M
TransUnet	0.8321	0.8586	0.8235	0.7492	0.3587	4.9740	0.7996	0.8553	0.7132	0.7064	0.2371	11.1686	105.32M
AMP-DRL	0.7786	0.7844	0.8068	0.6876	0.2885	6.8749	0.7485	0.8050	0.7011	0.6368	0.2063	12.2869	31.03M
DQN	0.7747	0.7840	0.8181	0.6844	0.3072	6.0283	0.7224	0.7816	0.6479	0.6070	0.1918	13.5471	7.77M
Double DQN	0.7835	0.7929	0.8103	0.6899	0.3105	5.8746	0.7432	0.7987	0.7081	0.6330	0.2077	12.0447	7.77M
Dueling DQN	0.7967	0.7993	0.7978	0.7073	0.3374	5.3818	0.7580	0.8087	0.7133	0.6526	0.2253	10.6835	7.77M
AC	0.8122	0.8104	0.8086	0.7251	0.3659	4.8330	0.7770	0.8375	0.7202	0.6758	0.2430	10.2233	10.82M
Our	**0.8346**	**0.8678**	**0.8469**	**0.7510**	**0.4081**	**4.0487**	**0.8147**	**0.8877**	**0.7659**	**0.7152**	**0.2733**	**9.0897**	**7.14M**

Table 2. Results of applying the proposed model and baselines on two public datasets, where the best results are in bold, the second best one on deep-learning-based and deep-reinforcement-learning-based methods are underlined.

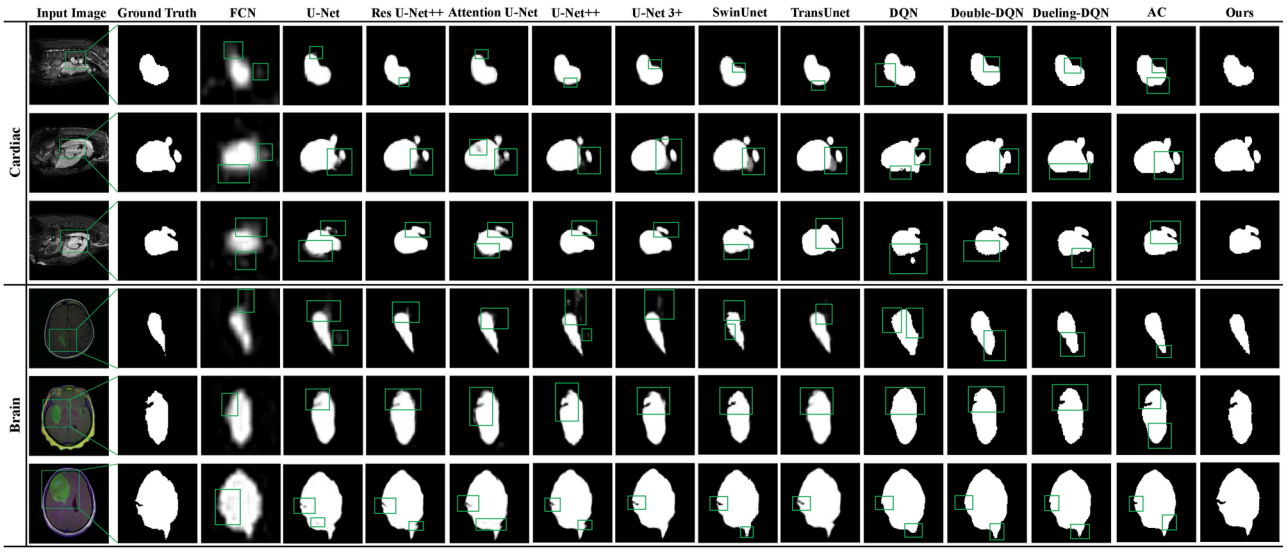

Fig. 3. Examples of visualized segmentation results of our proposed PixelDRL-MG and the baselines on two public datasets..

To prove that our model is more effective than common deep reinforcement learning, we choose five deep reinforcement learning methods as baselines. (i) **AMP-DRL**[24] applies reinforcement learning strategies to self-supervised medical image segmentation. (ii) **DQN**[30] first combines Q-Learning and deep learning to solve the instability problem, (iii) **Double DQN**[31] based on DQN uses different value functions to select and evaluate actions to solve the problem of overestimation, (iv) **Dueling DQN**[32] divides the last layer of DQN into two parts to obtain a more robust learning effect, and (v) **AC**[33] uses Actor and Critic to reduce the variance of gradient estimation. Note that we have not chosen A3C[15] as a baseline, because PixelRL essentially uses A3C to do pixel-level tasks, we mainly compare it with our pixel-by-pixel mask generation in ablation studies.

Main results

To illustrate the effectiveness of our PixelDRL-MG approach, we conducted experiments on two public medical image datasets, comparing the performance of PixelDRL-MG with common deep-learning-based and deep-reinforcement-learning-based models. The quantitative experimental results are shown in Table 2, while Fig. 3 presents examples of segmentation results for PixelDRL-MG and the baseline models on the two datasets.

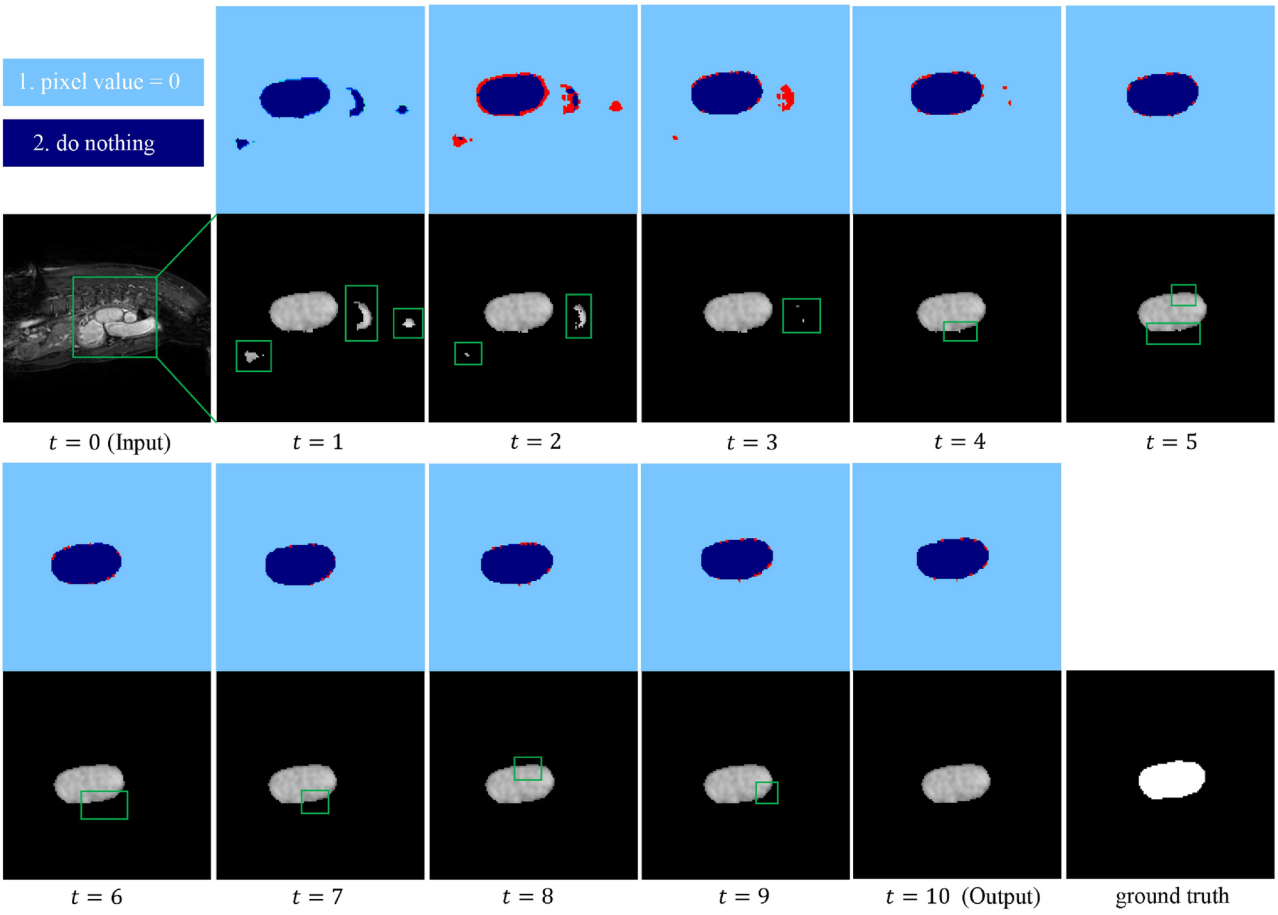

Fig. 4. Segmentation process of PixelDRL-MG and the action map at each time step. Red indicates the change of action between the current step and the previous step..

Moreover, Fig. 4 displays the segmentation results of our proposed PixelDRL-MG at each time step. Through both quantitative and qualitative analyses, we obtain the following conclusions.

First, as shown in Table 2, the proposed PixelDRL-MG can achieve more obvious improvements than all baselines, which proves that PixelDRL-MG is more effective than the most common image segmentation methods. Specifically, we first observe that compared with the deep learning methods, PixelDRL-MG is 1.51%, 1.46%, 1.94%, 0.88%, and 3.20% higher than the second best result on the Brain dataset for DICE, PPV, SEN, IoU, and BIoU, respectively, and HD95 is 1.089 lower. Then, compared with the common deep reinforcement learning methods, PixelDRL-MG also achieves the best segmentation performance. For example, PixelDRL-MG is 2.24%, 5.74%, 2.88%, 2.59%, and 4.22% higher than the second best result on the Cardiac dataset for DICE, PPV, SEN, IoU, and BIoU, respectively, and HD95 is 0.7843 lower; while on the Brain dataset, DICE, PPV, SEN, IoU, and BIoU are 3.77%, 5.02%, 4.57%, 3.94%, and 3.03% higher than the second best result, respectively, and HD95 is 1.1336 lower. Besides, we observe that PixelDRL-MG improves more on the Brain dataset than the Cardiac dataset, such as DICE, IoU, and HD95. This observation shows that our model is still effective or even improved significantly on datasets where the edge information of the segmentation objects is particularly complex. This is because (i) our model introduces SAM to extract global information, (ii) further introduces DC to learn the information of surrounding pixels, and (iii) our policy network directly generates the segmentation masks instead of modifying the probability maps, which further alleviates the quantization error and precision loss. Finally, we have observed that PixelDRL-MG achieves the best segmentation performance while having the smallest number of parameters compared to other models. This observation shows the efficacy of our method in achieving a superior segmentation performance without relying on complex model architectures.

To visually demonstrate the superior performance of PixelDRL-MG on two public datasets. Figure 3 displays the outcomes of both baseline models and PixelDRL-MG across twelve examples from two datasets. In particular, the segmentation results for heart images in the first three rows reveal that: (i) while the variants of U-Net and deep reinforcement learning methods exhibit significant improvements in clarity compared to the segmentation results of FCN and U-Net, the segmentation edges still fall short of the desired quality (as seen in the green box), and (ii) our method not only delivers high-definition results but also presents segmentation edges that align more closely with the ground truths. Similar observations can be made from the segmentation results of brain images in the last three rows, demonstrating that our approach excels in segmenting objects with intricate details

and complex information, yielding edges that are closer to the ground truths. Thus, these visual comparisons serve as compelling evidence of our model's ability to produce segmentation results that closely approximate the ground truths.

Finally, we visualize the segmentation process of PixelDRL-MG and the action map at each time step in Fig. 4. Light blue represents the background, navy blue represents the segmentation object, and red represents the change of the action at the current time step compared to the previous time step. The action of "pixel value = 0" is chosen in almost all the other regions, which is because the proportion of the segmentation object is small. The first and third lines of Fig. 4 represent the action map at each time step of our model, and the second and fourth lines represent the process of segmentation. It can be seen from Fig. 4 that from step $t = 1$ to step $t = 4$, the improvement of the segmentation performance is the most obvious, and after the fourth step, the segmentation performances are mainly fine-tuned. Specifically, from step $t = 1$ to step $t = 4$, there are many actions to solve the over-segmentation problem, after step $t = 4$, most of the actions are to fine-tune the segmentation boundary.

Ablation studies

To further study the effectiveness and necessity of the self-attention module (SAM) and dilation convolution (DC) in our PixelDRL-MG, ablation studies were conducted with four models based on PA3C. The corresponding experimental results are shown in Table 3. To further demonstrate that our pixel-by-pixel mask generation is more efficient than the existing method of using a threshold to choose actions, we also compared PixelDRL-MG to PixelRL with the same structure and settings in medical image segmentation. The PixelRL here uses the original PixelRL[29] structure with the same SAM and DC added, but the original PixelRL uses a threshold method for many image processing tasks and is not applied to medical image segmentation. Therefore, we refer to the settings of the iteratively-refined deep reinforcement learning method[13] in medical image segmentation and apply them to PixelRL for medical image segmentation, that is, using a coarse segmentation output as an input and refining the input by fixing the number of actions and varying the action values (we use ± 1.0, ± 0.4, ± 0.2 , and ± 0.1 following the work in[13]). In addition, referring to[13] using V-Net to provide 3D coarse segmentation, we use the outputs of U-Net and U-Net3+ as the coarse segmentation of PixelRL, denoted as PixelRL-U-Net and PixelRL-U-Net3+, respectively.

Effectiveness and necessity of each module

Table 3 demonstrates that all models, for both datasets, surpass PA3C across all metrics. This confirms the effectiveness of the proposed modules in enhancing the performance of PA3C in medical image segmentation. Specifically, it can first be observed that the results using only PA3C are suboptimal, with a 5.4% decrease in the Dice score compared to the full model. This is attributed to the model's lack of global and local information, thereby demonstrating that integrating both global and local information can assist each agent in making better decisions and achieving more accurate segmentation masks. Subsequently, when comparing the results of PA3C+SAM with those of PA3C, it is evident that PA3C+SAM outperforms PA3C across all metrics on all datasets (e.g., leading by 1.73% in the Dice score). This advantage is attributed to SAM's ability to more effectively capture global information and enrich the feature maps. Furthermore, it is observed that PA3C+DC consistently outperforms PA3C as well (e.g., leading by 4.1% in the Dice score). This is because the former can capture information not only from the current pixel but also from its neighboring pixels, thereby preserving more comprehensive details. Additionally, PA3C+DC achieves better results than PA3C+SAM (e.g., leading by 2.37% in the Dice score), which also indicates that the local information from surrounding pixels plays a more critical role than global information and serves as the primary reference for the agents. Finally, we find that our proposed model consistently outperforms all other models, as it combines SAM and DC to incorporate more diverse and richer information, thereby enhancing segmentation performance. This also demonstrates that both SAM and DC are indispensable for our deep reinforcement learning strategy. Moreover, since SAM is a lightweight and plug-and-play module, and DC is merely a simple convolutional operation, they hardly affect the computational efficiency of the model. This further underscores that PixelDRL-MG has low deployment costs.

Effectiveness and necessity of pixel-by-pixel mask generation

In Figs. 5 and 6, PixelRL-MG outperforms PixelRL on all metrics across all datasets, and its convergence speed is not slow. Specifically, we first compare the segmentation results of PixelRL-based methods (i.e., PixelRL-U-Net and PixelRL-U-Net3+), and find that PixelRL can optimize the coarse segmentation of U-Net and U-Net3+ through multi-step iterations, thereby getting a better segmentation performance. In Fig. 5, PixelRL-MG outperforms PixelRL on all metrics across all datasets. Specifically, we first compare the segmentation performances of PixelRL-based methods, and find that the segmentation performance largely depends on the performance of coarse segmentation, i.e., the segmentation performance of PixelRL-U-Net3+ is better than

Model			Cardiac						Brain					
PA3C	SAM	DC	DICE ↑	PPV ↑	SEN ↑	IoU ↑	BIoU ↑	HD95 ↓	DICE ↑	PPV ↑	SEN ↑	IoU ↑	BIoU ↑	HD95 ↓
✓	–	–	0.7806	0.8038	0.7925	0.6933	0.3403	5.5938	0.7662	0.8380	0.7059	0.6637	0.2255	10.3503
✓	✓	–	0.7979	0.8242	0.8051	0.7139	0.3614	5.0289	0.7853	0.8531	0.7165	0.6862	0.2425	10.1482
✓	–	✓	0.8216	0.8430	0.8299	0.7331	0.3762	4.5256	0.8080	0.8715	0.7581	0.7075	0.2621	9.3517
✓	✓	✓	**0.8346**	**0.8678**	**0.8469**	**0.7510**	**0.4081**	**4.0487**	**0.8147**	**0.8877**	**0.7659**	**0.7152**	**0.2733**	**9.0897**

Table 3. Results of ablation experiments on two public datasets, where the best results are in bold.

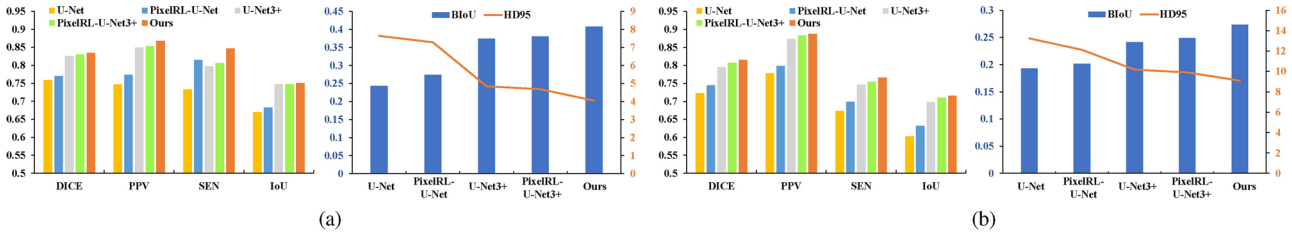

Fig. 5. Comparison of the segmentation results of the proposed PixelDRL-MG and the existing PixelRL on public datasets.

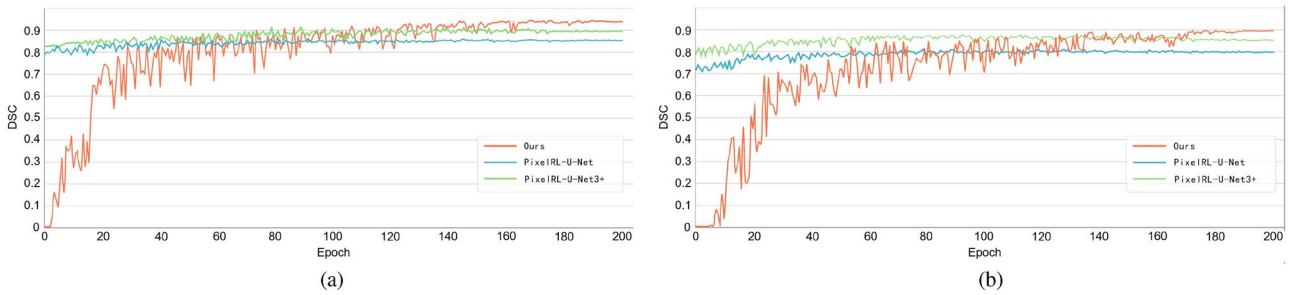

Fig. 6. Comparison of the training process of the proposed PixelDRL-MG and the existing PixelRL on public datasets. We take the DICE metric as an example.

that of Pixel-U-Net; the improvements of segmentation performance are also limited, especially the better the coarse segmentation performance, the more limited the improvement, i.e., the improvement of PixelRL-U-Net3+ for U-Net3+ is significantly smaller than that of PixelRL-U-Net for U-Net. These phenomena prove that improving the segmentation performance of PixelRL based on the coarse segmentation is not worth the loss compared with the high computational cost (both to obtain coarse segmentation and to use PixelRL for iterative optimization). Then, we observe that PixelDRL-MG is optimal for all metrics on both datasets, which proves that PixelDRL-MG can generate an optimal segmentation performance directly from the original image without providing coarse segmentation. In theory, if the feature extraction module of PixelDRL-MG is replaced with a module such as U-Net3+ with a better segmentation performance, a better segmentation performance will be achieved. Finally, comparing the training process of PixelRL-U-Net, PixelRL-U-Net3+, and the proposed PixelDRL-MG, we observe that the methods based on PixelRL in the Cardiac dataset converge at about 100 epochs; PixelDRL-MG achieves convergence at about 150 epochs from the original images, but it can achieve a segmentation performance similar to PixelRL-based methods in about 100 epochs. In the Brain dataset, PixelDRL-MG can achieve a segmentation effect similar to PixelRL in about 150 epochs. The above observations prove that PxielDRL-MG's direct selection action based on a pixel-by-pixel mask generation is better than PixelRL's threshold update state based on the coarse segmentation in terms of segmentation performance, and the convergence speed of PixelDRL-MG to the most segmented performance has not decreased.

Effects on extreme data

We conducted supplementary experiments on extreme datasets to verify the effectiveness and robustness of PixelDRL-MG. The extreme datasets here are randomly selected 50 and 100 heart images of the Cardiac dataset and 50 and 100 brain images of the Brain dataset as a new training set to show the effectiveness of our method in extremely small datasets. Table 4 presents the experimental results. It's worth noting that Transformer-based methods (i.e., SwinUnet and TransUnet) are not included here because using Transformer architectures on extremely small datasets fails to converge. Therefore, comparing them under such extreme data conditions would be unfair.

Generally, in Table 4, PixelDRL-MG can achieve a better segmentation performance than all baselines in terms of all metrics for two extreme datasets. Specifically, the segmentation results of the commonly deep-learning-based and deep-reinforcement-learning-based segmentation methods in the 50-shot and 100-shot datasets are significantly worse than that of the whole datasets. For example, the DICE of the deep supervised segmentation model in the 50-shot and 100-shot Cardiac datasets are at least 6.42% and 4.23% lower than that of the Cardiac dataset, respectively; and the DICE of the deep reinforcement learning model in the 50-shot and 100-shot Cardiac datasets are at least 5.98% and 3.93% lower than that of the Cardiac dataset, respectively. This demonstrates that extreme datasets can severely degrade the segmentation performance. Furthermore, we note that our model performs slightly inferior on the extreme dataset than on the Cardiac dataset. For example, the DICE of our model in the 50-shot and 100-shot Cardiac datasets are only 2.72% and 1.85% lower than that of the whole Cardiac dataset, respectively; and the HD95 of our model in the 50-shot and 100-shot Brain datasets

Datasets	Model	50-shot						100-shot					
		DICE ↑	PPV ↑	SEN ↑	IoU ↑	BIoU ↑	HD95 ↓	DICE ↑	PPV ↑	SEN ↑	IoU ↑	BIoU ↑	HD95 ↓
Cardiac	FCN	0.5238	0.4873	0.5012	0.4047	0.0979	12.0909	0.5509	0.4800	0.5155	0.4245	0.1726	10.4666
	U-Net	0.6843	0.6873	0.6726	0.5718	0.1273	11.1799	0.7072	0.7185	0.6979	0.6011	0.2544	9.6988
	Attention U-Net	0.7194	0.6389	0.7007	0.6001	0.1494	9.3551	0.7339	0.7334	0.7395	0.6188	0.2734	8.4078
	ResUNet++	0.7072	0.6352	0.7348	0.5967	0.1356	10.0048	0.7282	0.7328	0.7559	0.6167	0.2632	9.1119
	U-Net++	0.7127	0.7254	0.6881	0.6047	0.1784	8.5247	0.7326	0.7460	0.7065	0.6226	0.3296	7.1666
	U-Net3+	0.7454	0.7156	0.6998	0.6304	0.1850	8.6080	0.7570	0.7475	0.6754	0.6633	0.3344	6.9106
	DQN	0.7148	0.6569	0.7590	0.6042	0.1674	9.8448	0.7324	0.7482	0.7363	0.6277	0.2705	8.6223
	Double-DQN	0.7237	0.6871	0.7244	0.6154	0.1704	9.0577	0.7442	0.7586	0.7484	0.6508	0.2857	8.1778
	Dueling-DQN	0.7322	0.6880	0.7381	0.6289	0.1816	8.7631	0.7488	0.7520	0.7501	0.6561	0.3020	7.5715
	AC	0.7473	0.7319	0.7566	0.6393	0.2072	8.0565	0.7578	0.7614	0.7541	0.6627	0.3378	7.0192
	Our	**0.8074**	**0.8177**	**0.8133**	**0.7110**	**0.2582**	**6.7962**	**0.8161**	**0.8336**	**0.8274**	**0.7277**	**0.3567**	**5.2373**
Brain	FCN	0.5011	0.5215	0.4927	0.4073	0.1100	22.7416	0.5417	0.5773	0.5235	0.4370	0.1203	18.9221
	U-Net	0.6807	0.7518	0.6517	0.5671	0.1604	19.0654	0.7001	0.7651	0.6710	0.5883	0.1826	17.6037
	Attention U-Net	0.7046	0.7687	0.6730	0.5962	0.1876	16.5067	0.7180	0.7707	0.6805	0.6157	0.1977	14.7960
	ResUNet++	0.7149	0.7776	0.6781	0.6014	0.1824	18.0060	0.7221	0.8092	0.6888	0.6127	0.1932	15.1923
	U-Net++	0.7241	0.8041	0.6934	0.6114	0.1919	15.7402	0.7345	0.8241	0.7034	0.6230	0.2164	13.9584
	U-Net3+	0.7305	0.8101	0.7047	0.6155	0.2048	14.6539	0.7456	0.8220	0.7227	0.6381	0.2240	12.9880
	DQN	0.6848	0.7558	0.6311	0.5729	0.1814	17.5838	0.7022	0.7613	0.6386	0.5931	0.1952	15.8615
	Double-DQN	0.7049	0.7711	0.6707	0.5982	0.1901	16.2542	0.7176	0.7711	0.6861	0.6080	0.2022	14.7701
	Dueling-DQN	0.7262	0.7847	0.6893	0.6155	0.1984	15.6062	0.7336	0.7882	0.6879	0.6278	0.2086	14.1517
	AC	0.7327	0.7936	0.6951	0.6284	0.2070	14.2097	0.7455	0.8093	0.7045	0.6374	0.2218	12.7538
	Our	**0.7730**	**0.8418**	**0.7347**	**0.6651**	**0.2366**	**11.2409**	**0.7871**	**0.8572**	**0.7442**	**0.6888**	**0.2441**	**10.6495**

Table 4. Performance under extreme data constraints on public datasets, where the best results are in bold. TransUnet and SwinUnet, which use Transformer, are not included in the table as it would be unfair to compare them under extreme data.

Model	Hecktor						Params ↓
	DICE ↑	PPV ↑	SEN ↑	IoU ↑	BIoU ↑	HD95 ↓	
U-Net	0.5976	0.6541	0.6449	0.4839	0.0966	7.3645	31.03M
Attention U-Net	0.6351	0.6661	<u>0.6982</u>	0.5224	0.1053	6.5005	32.43M
nnU-Net	0.6513	0.6575	0.6876	0.5416	0.1075	6.2755	30.50M
TransUnet	<u>0.6547</u>	<u>0.6667</u>	0.6546	<u>0.5474</u>	0.1091	<u>5.9768</u>	105.32M
AMP-DRL	0.6156	0.6573	0.6663	0.5191	0.1020	7.1674	31.03M
AC	0.6440	0.6652	0.6772	0.5328	<u>0.1108</u>	6.2903	10.82M
Our	**0.6585**	**0.6694**	**0.6888**	**0.5501**	**0.1176**	**5.3583**	**7.14M**

Table 5. Results of applying the proposed model and baselines on the Hecktor datasets, where the best results are in bold, the second best methods are underlined.

is only 2.1512 and 1.5598 higher than that of the whole Brain dataset, respectively. The above phenomena prove the advantages of our method even in extreme datasets: exploiting our method's full use of data information, our model has a good robustness.

Scalability validation

We introduced the Hecktor dataset[46] to demonstrate the effectiveness and scalability of our method. This dataset was released by the Hecktor Challenge held at MICCAI 2020 for head and neck tumor segmentation. This dataset comprises 201 3D PET-CT scans of the head and neck region. In this work, we utilized only the PET modality, which is more intuitive for human perception, containing a total of 28,949 2D slices. The segmentation challenge of this dataset lies in the relatively blurred object boundaries and the presence of significant misleading features. We validated the effectiveness of our method on the Hecktor dataset. The segmentation performance results, as shown in Table 5, demonstrate that our method outperforms other baselines on this new dataset. Specifically, PixelDRL-MG is 0.38%, 0.27%, 0.94%, 0.27%, and 0.64% higher than the second best result on the Hecktor dataset for DICE, PPV, SEN, IoU, and BIoU, respectively, and HD95 is 0.6185 lower. This superior performance comes from our model's ability to leverage deep reinforcement learning strategies to individually optimize each pixel while maintaining relevant connections between them.

Model	Cardiac		Brain	
	Mean DICE ± SD	Mean BIoU ± SD	Mean DICE ± SD	Mean BIoU ± SD
U-Net	0.7572 ± 0.0087	0.2468 ± 0.0124	0.7164 ±0.0091	0.1957 ± 0.0153
TransUnet	0.8330 ± 0.0039	0.3602 ± 0.0082	0.7951 ± 0.0067	0.2328 ± 0.0120
AC	0.8147 ± 0.0047	0.3674 ± 0.0077	0.7702 ± 0.0084	0.2382 ± 0.0114
Our	**0.8382 ± 0.0033**	**0.4122 ± 0.0053**	**0.8127± 0.0042**	**0.2703 ± 0.0062**

Table 6. The 5-Fold Cross-Validation results on two public datasets, where the best results are in bold.

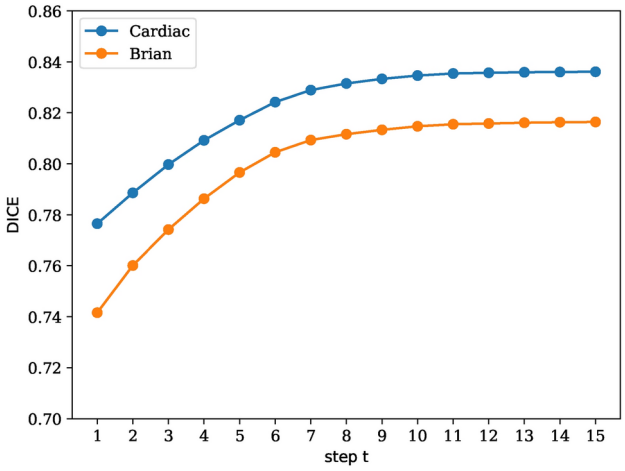

Fig. 7. Step-varying performance on public datasets.

Statistical validation

By employing 5-fold cross-validation on the Cardiac and Brain datasets, we conducted an in-depth investigation into the robustness and generalizability of the model. As shown in Table 6, the results reaffirm the superior performance of PixelDRL-MG, which not only achieves higher average performance but also exhibits more consistent outcomes. Specifically, on the Cardiac dataset, our model outperformed U-Net, TransUnet, and AC by an average margin of 8.1%, 0.52%, and 2.35%, respectively, in terms of the Mean Dice score. On the Brain dataset, it led by 9.63%, 1.76%, and 4.25%, respectively. These results highlight the resilience and adaptability of PixelDRL-MG, underscoring its statistical significance in the evaluation.

Effects on step-varying

We further investigated the impact of the iteration step size (i.e., t_{max} in Algorithm 1) on the segmentation performance of PixelDRL-MG on public datasets, using the Dice score as an example. The results are illustrated in Fig. 7. From Fig. 7, it can be observed that when the iteration step size increases from 1 to 6, the segmentation performance of PixelDRL-MG improves significantly. As the iteration step size further increases from 7 to 10, the performance growth slows down but still shows noticeable improvement. However, when the iteration step size exceeds 10, the performance improvement becomes marginal. This indicates that appropriately increasing the iteration step size can enhance the model's segmentation performance, but excessively large step sizes not only fail to provide substantial gains but also lead to unnecessary computational resource consumption. Therefore, to better balance computational efficiency and performance, we recommend setting the iteration step size to 10 in this work.

Discussion

We now briefly summarize the social impact of our approach, as well as the limitations of our work and future works.

Social impact of proposed approach

The purpose of medical image segmentation is to accurately outline organs and lesions, which greatly impacts subsequent diagnosis and clinical tasks. However, obtaining pixel-wise segmentation masks in practice is a task with high labor and cost requirements, which greatly limits the deployment efficiency and performances of intelligent medical image segmentation models in clinical practice. Therefore, most of the commonly used pixel-by-pixel segmentation methods based on deep reinforcement learning use user interaction or coarse-to-fine iterative optimization to obtain segmentation masks. However, these methods are not automatic models, as they still require expert interaction or a rough segmentation mask in advance, and these methods ignore the importance of global information and adjacent pixel information. Furthermore, these methods may lead

to quantization errors and precision losses by changing the feature map according to the threshold to obtain the segmentation mask. These issues hinder the rapid deployment and efficiency of the corresponding deep-reinforcement-learning-based medical image segmentation models. Therefore, we propose PixelDRL-MG to automatically generate pixel-by-pixel masks for the original medical images, which uses the policy network to directly select actions according to the current state of the pixels (set zero to represent the backgrounds, set one to represent the objects), and then makes the segmentation masks gradually approach the ground truths through multiple iterations. Therefore, our work can alleviate the problems of existing pixel-by-pixel segmentation methods based on deep reinforcement learning even in extremely small datasets, and help to achieve a high-accuracy segmentation on datasets with small objects and objects with complex boundary details. In addition, our automatic mask generation is a general concept that can be used in any deep reinforcement learning, it can use any form of the model to extract features, and it can also easily integrate various modules such as attention mechanisms. Hence, besides the technical contributions, this work also brings great social benefits in the related research and clinical areas, e.g., accelerating the application of intelligent computer-aided diagnosis systems in clinical practice to significantly reduce the workload of doctors, and saves both time and money for patients.

Limitations and future work

Although the proposed model of PixelDRL-MG achieves a better segmentation performance, larger and deeper models cannot be tried due to resource constraints. Besides, our method relies on the dilation convolution and self-attention module to provide a large amount of local and global information. Therefore, an interesting research direction is to introduce Transformers[9] (such as Vision Transformers[47]) to use the multi-head attention module to better obtain the global information. However, there are two more issues that require attention when using Transformers. First, due to too many parameters that need to be saved during training, it requires a lot of memory. Although we design PixelDRL-MG with small parameters, using Transformers directly will lead to a high memory overhead, which will lead to high storage and computation costs of the model, which is clearly undesirable. Second, if we train Transformers with a small dataset like the two in this paper, the model will struggle to converge. Therefore, our future work will focus on how to better combine Transformers with deep reinforcement learning to extract richer information to generate segmentation masks pixel by pixel, thereby improving the segmentation performance of medical images even further. Furthermore, the proposed PixelDRL-MG model in this work was only experimented with in a binary action space, as most medical segmentation tasks typically involve binary spaces. However, this does not mean that the method is limited to binary segmentation tasks. For multi-class segmentation in medical tasks, additional actions need to be incorporated into the action space (i.e., increasing the number of channels in the policy network's output). However, increasing the number of actions would also increase both the training cost and optimization complexity. Furthermore, in medical image segmentation tasks, multiple segmentation targets are not mutually exclusive. A single target may have multiple labels (e.g., an organ label and a tumor label), which means that the optimization strategy for binary segmentation cannot be directly applied. This requires a new multi-class optimization strategy or a new reward function to coordinate targets with multiple labels. Therefore, this represents an intriguing research direction that we plan to explore in the future.

Conclusion

Due to the limitations of deep learning methods, we attempt to address the challenges of medical image segmentation from a deep reinforcement learning perspective and identified the problems of most existing deep reinforcement learning-based segmentation methods: (i) High training cost, (ii) independent iterative process, and (iii) high uncertainty of segmentation mask. To address these issues, we proposed a new pixel-level deep reinforcement learning model with pixel-by-pixel mask generation using a dynamic iterative update policy (PixelDRL-MG), which does not require user interaction or coarse segmentation masks, can take different actions according to the current state to obtain a new state after directly inputting the original images, and makes the segmentation outputs slowly approach the ground truths through multiple iterations. To make the selected actions and calculated rewards more accurate, we introduced the self-attention module and dilated convolution module to help each agent obtain more global information and surrounding pixel information, respectively. Extensive experiments were conducted on two public medical image datasets. The results of these experiments proved that our model outperforms conventional deep learning and deep reinforcement learning methods in medical image segmentation performance, and each module and pixel-by-pixel mask generation are effective and complementary. Furthermore, we also demonstrated the robustness of our model on very small datasets.

Data Availability

The dataset is available at: https://www.kaggle.com/datasets/adarshsng/heart-mri-image-dataset-left-atrial-segmentation, https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation and https://www.aicrowd.com/challenges/miccai-2020-hecktor.

Received: 14 January 2025; Accepted: 25 February 2025
Published online: 10 March 2025

References

1. Sun, G. et al. Da-transunet: Integrating spatial and channel dual attention with transformer u-net for medical image segmentation. *Front. Bioeng. Biotechnol.* **12**, 1398237 (2024).
2. Sun, G. et al. Fkd-med: Privacy-aware, communication-optimized medical image segmentation via federated learning and model lightweighting through knowledge distillation. *IEEE Access* **12**, 33687–33704 (2024).

3. Pan, Y. et al. A mutual inclusion mechanism for precise boundary segmentation in medical images. *Front. Bioeng. Biotechnol.* **12**, 1504249 (2024).
4. Muksimova, S., Umirzakova, S., Kang, S. & Im Cho, Y. Cervilearnnet: Advancing cervical cancer diagnosis with reinforcement learning-enhanced convolutional networks. *Heliyon* **10**, 29913 (2024).
5. Yuan, D. et al. μ-net: Medical image segmentation using efficient and effective deep supervision. *Comput. Biol. Med.* **160**, 106963 (2023).
6. Xu, Z. et al. ω-net: Dual supervised medical image segmentation with multi-dimensional self-attention and diversely-connected multi-scale convolution. *Neurocomputing* **500**, 177–190 (2022).
7. Zhang, S., Zhang, J., Tian, B., Lukasiewicz, T. & Xu, Z. Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation. *Med. Image Anal.* **83**, 102656 (2023).
8. Oktay, O. *et al.* Attention U-Net: Learning where to look for the pancreas. ArXiv Preprint. ArXiv:1804.03999 (2018).
9. Vaswani, A. *et al.* Attention is all you need. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems* 5998–6008 (2017).
10. Zhou, Z., Siddiquee, M.M.R., Tajbakhsh, N. & Liang, J. UNet++: A nested U-Net architecture for medical image segmentation. In *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support* 3–11 (2018).
11. Cao, H. *et al.* Swin-Unet: Unet-like pure Transformer for medical image segmentation. In *Proceedings of the European Conference on Computer Vision Workshops* 205–218 (2022).
12. Liu, Z. *et al.* Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF international conference on computer vision* 10012–10022 (2021).
13. Liao, X. *et al.* Iteratively-refined interactive 3D medical image segmentation with multi-agent reinforcement learning. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 9394–9402 (2020).
14. Tian, Z., Si, X., Zheng, Y., Chen, Z. & Li, X. Multi-step medical image segmentation based on reinforcement learning. *J. Ambient Intell. Hum. Comput.* **2020**, 1–12 (2020).
15. Mnih, V. *et al.* Asynchronous methods for deep reinforcement learning. In *Proceedings of the International Conference on Machine Learning* 1928–1937 (2016).
16. Long, J., Shelhamer, E. & Darrell, T. Fully convolutional networks for semantic segmentation. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 3431–3440 (2015).
17. Ronneberger, O., Fischer, P. & Brox, T. U-Net: Convolutional networks for biomedical image segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention* 234–241 (2015).
18. Jha, D. *et al.* ResUNet++: An advanced architecture for medical image segmentation. In *IEEE International Symposium on Multimedia* 225–2255 (2019).
19. Huang, H. *et al.* UNet 3+: A full-scale connected U-Net for medical image segmentation. In *Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing* 1055–1059 (2020).
20. Isensee, F., Jaeger, P. F., Kohl, S. A., Petersen, J. & Maier-Hein, K. H. nnu-net: A self-configuring method for deep learning-based biomedical image segmentation. *Nat. Methods* **18**, 203–211 (2021).
21. Chen, J. et al. Transunet: Rethinking the u-net architecture design for medical image segmentation through the lens of transformers. *Med. Image Anal.* **97**, 103280 (2024).
22. Song, Y. et al. Mega-reward: Achieving human-level play without extrinsic rewards. In *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 5826–5833 (2020).
23. Yuan, D. et al. Painless and accurate medical image analysis using deep reinforcement learning with task-oriented homogenized automatic pre-processing. *Comput. Biol. Med.* **153**, 106487 (2023).
24. Xu, G., Wang, S., Lukasiewicz, T. & Xu, Z. Adaptive-masking policy with deep reinforcement learning for self-supervised medical image segmentation. In *2023 IEEE International Conference on Multimedia and Expo (ICME)* 2285–2290 (2023).
25. Shokri, M. & Tizhoosh, H. R. Using reinforcement learning for image thresholding. In *Proceedings of the Canadian Conference on Electrical and Computer Engineering* **2**, 1231–1234 (2003).
26. Sahba, F., Tizhoosh, H.R. & Salama, M.M. A reinforcement learning framework for medical image segmentation. In *Proceedings of the 2006 IEEE International Joint Conference on Neural Networks* 511–517 (2006).
27. Wang, L., Lekadir, K., Lee, S.-L., Merrifield, R. & Yang, G.-Z. A general framework for context-specific image segmentation using reinforcement learning. *IEEE Trans. Med. Imaging* **32**, 943–956 (2013).
28. Song, Y. et al. Arena: A general evaluation platform and building toolkit for multi-agent intelligence. In *Proceedings of the AAAI Conference on Artificial Intelligence* **34**, 7253–7260 (2020).
29. Furuta, R., Inoue, N. & Yamasaki, T. PixelRL: Fully convolutional network with reinforcement learning for image processing. *IEEE Trans. Multimed.* **22**, 1704–1719 (2019).
30. Mnih, V. et al. Human-level control through deep reinforcement learning. *Nature* **518**, 529–533 (2015).
31. Van Hasselt, H., Guez, A. & Silver, D. Deep reinforcement learning with double Q-learning. In *Proceedings of the AAAI Conference on Artificial Intelligence*, vol. 30 (2016).
32. Wang, Z. *et al.* Dueling network architectures for deep reinforcement learning. In *Proceedings of the International Conference on Machine Learning* 1995–2003 (PMLR, 2016).
33. Konda, V. & Tsitsiklis, J. Actor-critic algorithms. In *Proceedings of the International Conference on Advances in Neural Information Processing Systems* 1008–1014 (2000).
34. Simonyan, K. & Zisserman, A. Very deep convolutional networks for large-scale image recognition. arXiv preprint arXiv:1409.1556 (2014).
35. Yu, F., Koltun, V. & Funkhouser, T. Dilated residual networks. In *Proceedings of the IEEE conference on computer vision and pattern recognition* 472–480 (2017).
36. Wang, X., Girshick, R., Gupta, A. & He, K. Non-local neural networks. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* 7794–7803 (2018).
37. Yu, F. & Koltun, V. Multi-scale context aggregation by dilated convolutions. ArXiv Preprint, ArXiv:1511.07122 (2015).
38. Duan, X., Liu, X., Gong, X. & Han, M. RL-CoSeg: A novel image co-segmentation algorithm with deep reinforcement learning. ArXiv Preprint. ArXiv:2204.05951 (2022).
39. Lowe, R. et al. Multi-agent actor-critic for mixed cooperative-competitive environments. *Adv. Neural Inf. Process. Syst.* **30**, 1–12 (2017).
40. Simpson, A.L. *et al.* A large annotated medical image dataset for the development and evaluation of segmentation algorithms. ArXiv Preprint. ArXiv:1902.09063 (2019).
41. Buda, M., Saha, A. & Mazurowski, M. A. Association of genomic subtypes of lower-grade gliomas with shape features automatically extracted by a deep learning algorithm. *Comput. Biol. Med.* **109**, 218–225 (2019).
42. Yu, Q., Xia, Y., Xie, L., Fishman, E.K. & Yuille, A. L. Thickened 2D networks for efficient 3D medical image segmentation. ArXiv Preprint. ArXiv:1904.01150 (2019).
43. Kingma, D.P. & Ba, J. Adam: A method for stochastic optimization. In *Proceedings of the International Conference for Learning Representations* 1–15 (2014).
44. Karimi, D. & Salcudean, S. E. Reducing the Hausdorff distance in medical image segmentation with convolutional neural networks. *IEEE Trans. Med. Imaging* **39**, 499–513 (2019).

45. Cheng, B., Girshick, R., Dollár, P., Berg, A.C. & Kirillov, A. Boundary IoU: Improving object-centric image segmentation evaluation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition* 15334–15342 (2021).
46. Andrearczyk, V. *et al.* Overview of the hecktor challenge at miccai 2021: Automatic head and neck tumor segmentation and outcome prediction in pet/ct images. In *Proceedings of the 3D Head and Neck Tumor Segmentation in PET/CT Challenge* 1–37 (2021).
47. Dosovitskiy, A. *et al.* An image is worth 16x16 words: Transformers for image recognition at scale. ArXiv Preprint. ArXiv:2010.11929 (2020).

Acknowledgements

This work was supported by the National Natural Science Foundation of China under the grant 62276089, by the Natural Science Foundation of Tianjin City, China, under the grant 24JCJQJC00200, by the Natural Science Foundation of Hebei Province, China, under the grant F2024202064, by the Ministry of Human Resources and Social Security, China, under the grant RSTH-2023-135-1, by the S&T Program of Hebei under the grant 225676163GH, by the Shenzhen Longgang District Innovation and Technology Special Fund (grant No. LGKCYLWS2024-27 and No.LGWJ2023-120), and by Natural Science Foundation of Chongqing, China, under the grant CSTB2024NSCQ-MSX0314.

Author contributions

Yunxin Liu: Writing - Original Draft, Writing - Review & Editing, Software, Methodology, Validation, Formal Analysis, Investigation, Visualization. Yuan Di: Writing - Original Draft, Writing - Review & Editing, Methodology, Validation, Formal Analysis, Investigation. Zhenghua Xu: Writing - Review & Editing, Methodology, Resources, Supervision, Funding Acquisition. Yuefu Zhan: Writing - Review & editing, Investigation, Funding Acquisition. Hongwei Zhang: Writing - Review & editing, Formal analysis. Jun Lu: Writing - Review & editing, Software. Thomas Lukasiewicz: Writing - Review & Editing, Supervision.

Declarations

Competing interests

The authors declare no competing interests.

Ethics approval

This article does not contain any studies with human participants performed by any of the authors.

Additional information

Correspondence and requests for materials should be addressed to Z.X. or Y.Z.