Adaptive-Masking Policy with Deep Reinforcement Learning for Self-Supervised Medical Image Segmentation

Gang Xu[1], Shengxin Wang[2], Thomas Lukasiewicz[3,4], Zhenghua Xu[1,2,*]

[1]School of Artificial Intelligence, Hebei University of Technology, Tianjin, China
[2] State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Health
Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, China
[3]Institute of Logic and Computation, TU Wien, Vienna, Austria
[4] Department of Computer Science, University of Oxford, Oxford, UK
[*]Corresponding Author, Email: zhenghua.xu@hebut.edu.cn

***Abstract*—Although self-supervised learning methods based on masked image modeling have achieved some success in improving the performance of deep learning models, these methods have difficulty in ensuring that the masked region is the most appropriate for each image, resulting in segmentation networks that do not get the best weights in pre-training. Therefore, we propose a new adaptive-masking policy self-supervised learning method. Specifically, we model the process of masking images as a reinforcement learning problem and use the results of the reconstruction model as a feedback signal to guide the agent to learn the masking policy to select a more appropriate mask position and size for each image, helping the reconstruction network to learn more fine-grained image representation information and thus improve the downstream segmentation model performance. We conduct extensive experiments on two datasets, Cardiac and TCIA, and the results show that our approach outperforms current state-of-the-art self-supervised learning methods.**

***Index Terms*—Masked image modeling, Self-supervised learning, Deep reinforcement learning, Medical image segmentation**

I. Introduction

Deep learning has achieved remarkable success in medical image-assisted analysis [1]–[3], but this success relies heavily on a large amount of accurate annotation data [4]. However, since (i) building a sufficiently large and high-quality annotated medical imaging dataset is costly and time-consuming, (ii) medical imaging datasets require experts with specialized knowledge for annotation, and (iii) the annotation process is prone to patient privacy issues, so the scarcity of annotation data has become the main obstacle to the application of deep learning in the medical field [5], limiting the performance of deep learning models.

In recent years, self-supervised learning (SSL) is emerging as a new paradigm to address the problem of annotating data scarcity [6]. SSL is capable of learning visual representations without using manual annotations. Specifically, to obtain general features, SSL first pre-trains deep learning models using unlabeled data; then, the pre-trained resultant model is fine-tuned on a small amount of labeled data. Currently, most self-supervised methods are based on contrastive learning [7]–[9]. However, contrastive learning prefers to learn global semantic features and does not learn image details well, which leads to limitations in the accuracy of pre-trained models in downstream tasks. For this reason, masked image modeling (MIM) for self-supervised pre-training has been widely studied [10]–[13]. MIM helps the model to learn finer-grained visual representations from unlabeled data by reconstructing the masked image to adapt to different downstream tasks and to better improve accuracy.

However, in MIM, the results of the reconstruction model cannot be fed back to adjust the masking policy, so the selected mask position and size are fixed or random, which makes it difficult to ensure that the masked region is the most appropriate for each image. Selecting an appropriate mask region for the reconstruction task is necessary for mining information-rich high-dimensional semantic features, and thus the need for a reasonable selection of mask positions and sizes in medical images is urgent. Since the selection of the mask region is a sequential decision, this selection process can be formulated as a Markov decision process (MDP) [14] and is implemented by deep reinforcement learning (DRL) [15].

Therefore, in this paper, we propose a new self-supervised method of MIM for medical image segmentation based on DRL (AMP-DRL). We apply DRL to the self-supervised domain to effectively utilize the unlabeled data and reduce the labeling cost while ensuring segmentation accuracy. Specifically, we develop a reinforcement learning agent based on the dueling deep q-learning network (Dueling DQN) [16], which uses the feedback signals provided by the reconstruction module to learn image-specific masking policy and select the appropriate mask position and size for each image to help the segmentation network to obtain better weights in self-

This work was supported by the National Natural Science Foundation of China under the grants 62276089 and 61906063, by the Natural Science Foundation of Hebei Province, China, under the grant F2021202064, by the "100 Talents Plan" of Hebei Province, China, under the grant E2019050017.

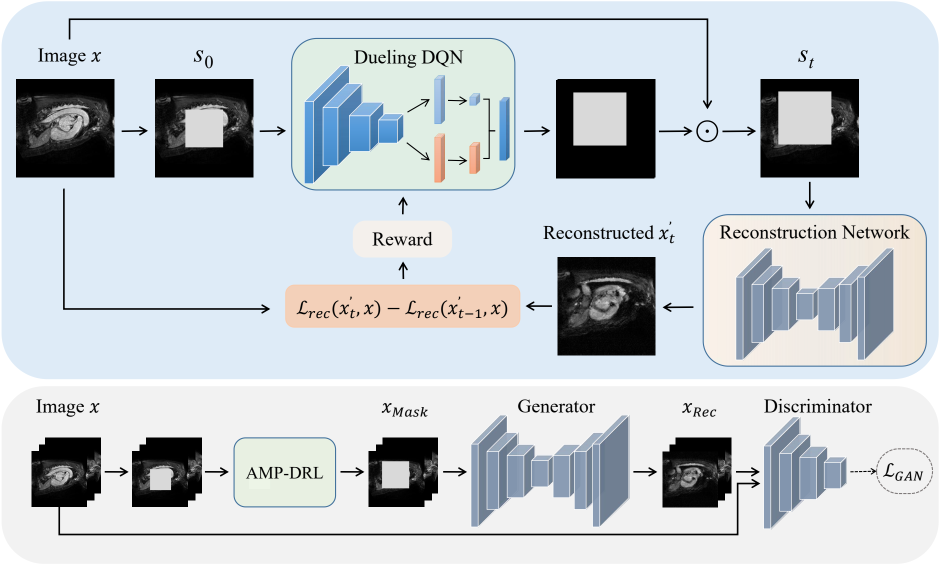

Fig. 1. Overview of our proposed mask-based self-supervised medical image segmentation method.

supervised pre-training, thus obtaining better performance with a small amount of finely labeled data for training.

The main contributions of this paper can be summarized as the following three points: (i) We combine DRL and SSL to propose AMP-DRL, which is applied to medical image segmentation tasks to reduce annotation cost while ensuring segmentation accuracy and effectively solve the problem of scarcity of medical image annotation data. (ii) We model the process of masking images as a reinforcement learning problem, and use the reconstruction module to provide feedback signals to guide the Dueling DQN to learn image-specific mask policy to select the appropriate mask position and size for each image. (iii) Extensive experiments are conducted on two public medical image datasets; the experimental results show that AMP-DRL outperforms state-of-the-art (SOTA) self-supervised methods in medical image segmentation tasks.

II. Related Work

A. *Self-Supervised Learning*

The main goal of SSL is to learn transferable knowledge from unlabeled data through well-designed pretext tasks and then transfer the learned knowledge to downstream tasks [5].

Contrastive learning. It mainly learns by constructing positive and negative samples and then comparing the distance between them. SimCLR [7] performs random data augmentations on an input batch image, maximizing the similarity between positive samples while minimizing the similarity between negative samples. BYOL [8] does not require negative samples and constrains the online network to predict the target network representation of the same image under different enhanced images by MSE loss. SwAV [9] introduces clustering algorithms that reinforce the consistency between cluster assignments for different views of the same image while clustering the data to learn useful information representations. However, these designs are more biased to learn global semantic features of images, and thus have shortcomings in learning fine-grained representations, which is detrimental to downstream segmentation tasks. Unlike previous approaches, AMP-DRL can help pre-trained models learn finer-grained image features, thus improving the accuracy of downstream segmentation models more effectively.

Masked image modeling. It learns fine-grained visual representations by reconstructing the masked region of an image. Deepak et al. [10] mask a fixed central area of an image and let the network use the surrounding image information to infer the missing region. MAE [11] uses an asymmetric encoder and decoder structure to divide the image into a number of the same size patches, and the masked patches are predicted directly based on the unmasked image patches. SimMIM [12] lightens the weight of the decoder based on MAE and takes all visible and masked patches as input, which allows it to achieve similar results as MAE while speeding up the pre-training process. ConvMAE [13] performs multi-scale coding operations based on MAE, which makes the model learn richer semantic information. However, the positions and sizes of masks in these tasks are fixed or random, which cannot be guaranteed to be the most appropriate for each image, which leads to the weights of the pre-trained model are not optimal. AMP-DRL learns the masking policy based on the deep reinforcement learning framework so that the reconstruction model uses the most effective part of the image to learn to get better pre-training weights.

B. *Deep Reinforcement Learning*

DRL is used to describe and solve the problem of maximizing the reward or achieving a specific goal through learning policies during the interaction between the agent and the environment [17]. Yunze et al. [18] apply deep Q-learning (DQN) to identify each pancreas's bounding box and then use a modified U-Net to segment the pancreas in cropped CT images. Qin et al. [19] propose to train both augmentation and segmentation modules simultaneously and use the errors during segmentation as feedback to adjust the augmentation module. DRL-LNS [20] proposes a DRL method for weakly supervised lesion segmentation. However, these DRL methods are based on fully supervised or weak-supervised learning, and thus they cannot utilize unlabeled data. Our proposed AMP-DRL is based on SSL, which effectively uses unlabeled data to pre-train the segmentation network, reducing the labeling cost while ensuring segmentation accuracy, and effectively solving the problem of scarcity of medical image labeled data.

III. METHODS

We propose a mask-based self-supervised medical image segmentation method. The method mainly consists of three stages: pre-training of the reconstruction network, adaptive-masking policy network training, and self-supervised medical image segmentation. First, we generate a random mask on the images to go for pre-training of the randomly initialized reconstruction network. This stage aims to obtain a reconstruction model that can recover the masked images, as the environment to provide feedback signals in the second stage. Second, we use the feedback signal to train the policy network to learn better mask policy and find more effective mask regions for each image. Then, we use the learned policy to select the appropriate mask for each image to complete the self-supervised pre-training of the reconstruction network with the random initialization, aiming to obtain a better encoder to improve the accuracy of the downstream segmentation model.

Specifically, in the first stage, we generate a mask of random size at random locations on the image, and then we use the masked image as the input to the reconstruction network to recover the masked region. Our reconstruction network uses a similar structure as Context Encoders [10], but Context Encoders can only recover mask areas of fixed location and size. We have improved on this to apply the reconstruction task to mask areas with different locations and sizes.

In the second stage, the overall structure of the adaptive-masking policy network is illustrated at the top of Figure 1, which mainly consists of two parts: Dueling DQN and a reconstruction module, representing the agent and the environment in the reinforcement learning framework, respectively. The agent is trained to learn adaptive-masking policy using signals received from the environment. This policy aims to search for a more appropriate location and size of the mask region for each image and thus train the reconstruction network more efficiently. The process of learning this policy is an MDP, and its states, actions, and rewards are defined in detail in Section 3.1. We show its specific training process in Section 3.2.

In the third stage, we use the learned adaptive-masking policy to automatically mask more appropriate regions for the images to complete the self-supervised pre-training of the randomly initialized reconstruction network so that the encoder of the reconstruction network can better learn the fine-grained semantic features, as shown in the bottom of Figure 1. The loss of the self-supervised pre-training is defined as:

$$\begin{aligned} min_G max_D V(D,G) = E_{x\sim p_{data}(x)}[logD(x)]+ \\ E_{x_{mask}\sim p_{data}(x_{mask})}[log(1-D(G(x_{mask})))]. \end{aligned} \tag{1}$$

Then, we migrate the obtained encoder to the downstream segmentation task, fine-tune the segmentation network with a small amount of labeled data, and test the segmentation accuracy in the test set.

TABLE I
THE ACTIONS OF AMP-DRL.

Action	Operation	Amplitude
Up	Move Up	±1 ±3 ±5
Down	Move Down	±1 ±3 ±5
Left	Shift Left	±1 ±3 ±5
Right	Shift Right	±1 ±3 ±5
Zoom	Scale Up or Down	±1 ±2

A. Network Components

Action. We define the operations to adjust the position and size of the mask area as actions. Table 1 shows 16 transformations containing 12 actions to control the direction to adjust the position of the mask and 4 scaling actions to control the size to adjust the range of the mask. In order to mask the images more comprehensively and efficiently, we set different magnitudes for each action, which allows for finer adjustments to the position and size of the mask.

State. The state is the input mask image. We use a binary matrix with a fixed position and size of the mask as the initial mask matrix to cascade with the pre-processed input image to obtain the masked image. We take this mask image as the initial state S_0. The subsequent state S_t is the cascade of the input image and the mask matrix that records the past t actions output by the agent.

Reward. The reward is a scalar value fed to the agent by the environment to evaluate the goodness of the action. To guide the agent in selecting more suitable regions to mask, we use the loss function of the reconstruction module to calculate the reward. Specifically, we use the difference between the current reconstruction loss and the loss of the previous time step as the reward of the current time step. The goal of reinforcement learning is to maximize the cumulative reward, so the reward guides the agent to learn actions that increase the difference between the losses of the two steps. The reward for step t is formally written as:

$$Reward^t = \mathcal{L}_{rec}(x_t^{'}, x) - \mathcal{L}_{rec}(x_{t-1}^{'}, x) \tag{2}$$

where x is the original image and $x^{'}$ is the image output by the reconstruction module. $\mathcal{L}_{rec}$ is defined as follows:

$$\mathcal{L}_{rec} = \frac{k_{t-1}}{k_t}(x_t^{'}, x)^2 \tag{3}$$

where k is the size of the masked region.

B. Agent Learning

AMP-DRL uses Dueling DQN as an agent to learn the probability distribution from states to actions according to the Q-function. Specifically, the outputs of Dueling DQN are the state valuation function $V(S_t)$ and the state-independent action dominance valuation function $A(S_t)$, respectively, which decouples the action-independent state values from the Q-values in such a way to obtain a more robust learning effect. Its Q-function is defined as follows:

$$\begin{aligned} Q^\pi(s^t, a^t; \theta, \theta_v, \theta_a) = V(s^t; \theta, \theta_v) + (A(s^t, a^t; \theta, \theta_a) \\ - \frac{1}{|A|}\sum_{a^{t+1}} A(s^t, a^{t+1}; \theta, \theta_a)). \end{aligned} \tag{4}$$

TABLE II

RESULTS OF THE PROPOSED AMP-DRL AND THE BASELINES ON CARDIAC AND TCIA DATASETS.

Methods		**Cardiac**						**TCIA**					
		DSC	MIoU	PPV	Sen	BIoU	HD95	DSC	MIoU	PPV	Sen	BIoU	HD95
5%	U-Net [1]	0.4836	0.3612	0.5873	0.5299	0.1505	12.1369	0.5836	0.5546	0.7416	0.7881	0.1443	6.8566
	SimCLR [7]	0.5629	0.4410	0.6822	0.5823	0.2445	11.3345	0.6217	0.5966	0.8088	0.7618	0.1347	6.3570
	BYOL [8]	0.5819	0.4547	0.6347	0.6644	0.2167	14.0389	0.6291	0.6014	0.8002	0.7803	0.1511	5.9131
	SwAV [9]	0.5993	0.4623	0.6115	0.7142	0.1978	9.9339	0.6346	0.6056	0.8086	0.7795	0.1397	6.7683
	Context [10]	0.6357	0.5081	0.6791	0.6729	0.2504	12.3797	0.6506	0.6244	0.8629	0.7483	0.1475	5.6945
	SimMIM [12]	0.6422	0.5153	0.6511	0.6857	0.2704	12.4581	0.6665	0.6380	0.8743	0.7482	0.1583	5.9744
	MAE [11]	0.6436	0.5153	0.6444	0.7440	0.2655	11.4665	0.6761	0.6469	0.8793	0.7485	0.1586	5.6218
	ConvMAE [13]	0.6537	0.5333	0.6614	0.7007	0.2802	12.6538	0.6793	0.6516	**0.8826**	0.7580	0.1353	5.5974
	Ours	**0.6646**	**0.5500**	**0.6890**	**0.7447**	**0.2839**	**9.0345**	**0.6895**	**0.6599**	0.8543	**0.7931**	**0.1623**	**5.3198**
10%	U-Net [1]	0.6493	0.5163	0.6591	0.7227	0.2564	10.9397	0.6777	0.6456	0.8624	0.7643	0.1729	5.3267
	SimCLR [7]	0.6773	0.5663	0.6806	0.7631	0.3026	14.4016	0.6920	0.6595	0.8668	0.7650	0.1922	6.2501
	BYOL [8]	0.6839	0.5765	0.7121	0.7372	0.3193	13.7063	0.7089	0.6784	0.8878	0.7636	0.1722	6.4104
	SwAV [9]	0.6891	0.5773	0.7321	0.7326	0.3180	7.6221	0.7131	0.6791	0.8734	0.7822	0.1832	5.7097
	Context [10]	0.7101	0.6007	0.7342	0.7595	0.3222	7.6234	0.7299	0.6966	0.8601	**0.8156**	0.1753	5.4416
	SimMIM [12]	0.7124	0.5979	0.7074	0.7767	0.3257	8.8852	0.7327	0.6939	0.8608	0.8081	0.1809	6.1551
	MAE [11]	0.7147	0.5988	0.7104	0.7771	0.3180	7.4786	0.7395	0.7075	0.9169	0.7742	0.2052	**5.0874**
	ConvMAE [13]	0.7174	0.6109	0.7245	0.7919	**0.3492**	9.8764	0.7414	0.7086	0.9102	0.7750	0.2056	5.6757
	Ours	**0.7408**	**0.6297**	**0.7373**	**0.7988**	0.3260	**4.5542**	**0.7534**	**0.7229**	**0.9191**	0.7883	**0.2165**	5.2464
50%	U-Net [1]	0.7222	0.6349	0.7730	0.7279	0.3975	5.8154	0.7407	0.7091	0.8545	0.8251	0.2822	4.7658
100%	U-Net [1]	0.7944	0.6941	0.8184	0.8200	0.4547	3.9390	0.8316	0.7946	0.8815	0.9002	0.2958	3.6435

where θ denotes the parameters of the convolutional layer shared by the two branches and θ_v and θ_a are the parameters owned by each independently.

To make the training process of AMP-DRL more stable, we use the ϵ-greedy algorithm [21] to balance the exploitation and exploration of the agent:

$$\epsilon = 0.05 + (0.95 - 0.05) \times e^{-1 \times \frac{epoch}{15}} \tag{5}$$

In the early stage of training, the agent is mainly exploring. As the number of training rounds increases, ϵ shows a decreasing trend, thus increasing the probability of the agent choosing the action by learning from previous experience.

When reward>0, it means that the mask region adjusted by the current action is beneficial to the training of the reconstruction network, otherwise it means that the action is harmful to the training. Therefore, for an image, when reward<0 occurs n times in a row, it indicates that the policy continues to harm the adjustment of the mask region, so we end the adjustment process.

IV. EXPERIMENTS

A. Dataset

To evaluate the performance of our proposed AMP-DRL in the segmentation task, we conduct extensive experiments on two publicly available magnetic resonance imaging (MRI) datasets (Cardiac [22], [23] and TCIA [24]). We divide the dataset into training, validation, and test sets by case with a ratio of 7:1:2.

Cardiac: The Cardiac dataset contains 20 cases with 1350 MRI images, each of which is 320×320 mm^2 in size. It has a small number of data and large anatomical variability, which is more challenging for the segmentation task.

TCIA: The TCIA dataset contains 110 LGG cases with a total of 3929 MRI images, each with a size of 256×256 mm^2.

We resize all images of the Cardiac to 256×256 mm^2 as network input. For network training, we augment the training set using standard data augmentation including random rotation, random flipping, scaling, and adding Gaussian noise.

B. Implementation Details

Our experiments are implemented using the Pytorch framework and run on an NVIDIA GeForce GTX2080 GPU. To evaluate the performance of AMP-DRL, we perform fully supervised training on a randomly initialized U-Net using 5% and 10% of the data and use it as our original baseline. In our experiments, we select several state-of-the-art contrastive learning and MIM methods as self-supervised baselines.

For U-Net, we use Adam [25] optimizer with an initial learning rate set to 0.0003, 10% decay every 3 epochs, and batch size of 4. For Dueling DQN, we use Adam optimizer with a learning rate set to 0.001 and Replay Memory D of size 100000. The discount factor β of the reward is set to 0.1. In the training, n is set to 2. In the pre-training of the reconstructed network, we use the Adam optimizer with the initial learning rate set to 0.0002 and the batch size of 12.

We use six widely used segmentation evaluation metrics including Mean Intersection over Union (MIoU), Dice Similarity Coefficient (DSC), Positive Predict Value (PPV), Sensitivity (Sen), Boundary Intersection over Union (BIoU), and 95% Hausdorff Distance (HD95).

C. Results and Discussion

We show the quantitative and qualitative results of AMP-DRL in the segmentation task. The experimental results of AMP-DRL and seven self-supervised learning baselines in the segmentation task are shown in Table 2, four examples of visual segmentation results are shown in Figure 2, Figure 3 shows the performance comparison of the segmentation model when different conditions are set for the location and size of the mask region, and Figure 4 visualizes the activation maps of the objects of interest on the Cardiac dataset.

Compare with Fully Supervised Learning from Scratch. As shown in Table 2, for all metrics, the SSL (including the proposed AMP-DRL) generally outperforms the fully supervised baseline with the same proportion of labeled data. This is because the self-supervised approach learns valuable

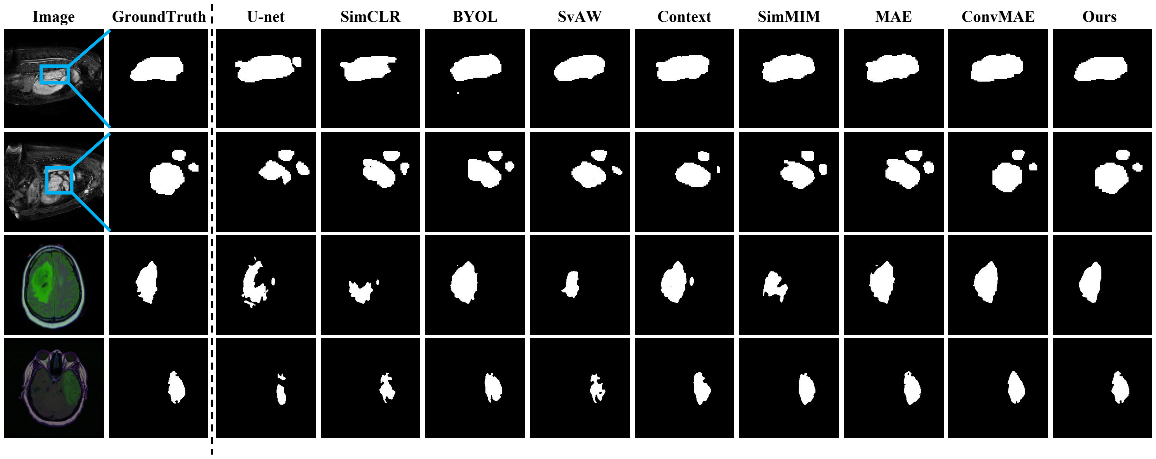

Fig. 2. Visualization of segmentation results.

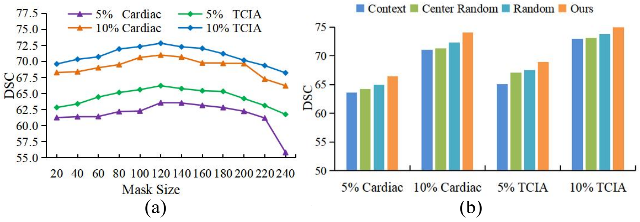

Fig. 3. (*a*) is the result of segmentation when a mask region of a different size is used at the central position, and (*b*) is the result of segmentation under four different mask policies.

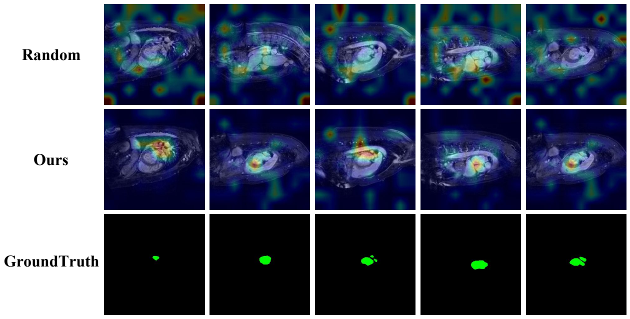

Fig. 4. Visualization of the activation maps of objects of interest on Cardiac.

representations for the downstream task from a large amount of unlabeled data, thus improving the performance of the downstream segmentation model. In addition, AMP-DRL greatly outperforms the baseline model trained from scratch in the 5% and 10% annotation cases. When using 10% annotation, we can outperform the fully supervised approach with 50% annotation on several metrics.

Compare with Self-Supervised Learning Baselines. Then, we further compare our AMP-DRL with SOTA self-supervised methods on 5% and 10% labeled data. Specifically, Context, SimMIM, MAE, and ConvMAE generally outperform SimCLR, BYOL, and SvAW on both datasets. This is because the MIM approach can learn more fine-grained image features than contrastive learning, thus improving the performance of the downstream segmentation model more effectively. Finally, we find that our proposed AMP-DRL generally outperforms all baselines on two datasets: i) at 10% annotation ratio on the Cardiac dataset, AMP-DRL improves DSC by 2.34%, MIoU has a 1.88% improvement and HD95 is reduced by 2.9244mm compared to the SOTA self-supervised approach; ii) on the TCIA dataset of 10% annotation ratio, AMP-DRL has 1.20% improvement in DSC and 1.43% improvement in MIoU compared to SOTA self-supervised method. This demonstrates that AMP-DRL achieves better performance than the SOTA self-supervised methods in medical image segmentation tasks. The reason for the superior performance of AMP-DRL is that AMP-DRL finds a more suitable mask region for each image, which helps the reconstruction network to learn more fine-grained image representation information during training, resulting in better pre-training parameters.

Analysis of Visualized Segmentation Results. Furthermore, the visualization results in Figure 2 support the above conclusions, where AMP-DRL is significantly better than all the self-supervised methods. Specifically, i) the segmentation results of the contrastive learning methods SimCLR, BYOL, and SvAW are highly inaccurate and even over-segmented; ii) Context, SimMIM, MAE, and ConvMAE have better results but unsatisfactory segmentation performance in the edge region; and iii) the proposed AMP-DRL segmentation results are closer to the ground truth and retain more details in the foreground region. Thus, these visualization examples demonstrate again that AMP-DRL compensates for the shortcomings of existing self-supervised medical image segmentation methods and achieves better performance in medical image segmentation tasks with a small amount of annotation.

D. Additional Experiments

We conduct additional self-supervised experiments to investigate the effect of both on the downstream segmentation performance by setting the position and size of the mask region and thus validate the feasibility of our proposed AMP-DRL. The results are shown in Figure 3, where (a) shows the segmentation accuracy downstream when using different sizes of mask regions at the image center. We can observe that the segmentation performance shows an increasing and then decreasing curve with increasing mask size, and the downstream segmentation accuracy is relatively high in the interval of mask size from $80{\times}80\ mm^2$ to $180{\times}180\ mm^2$, which proves that different mask sizes have an impact on the quality of self-supervised pre-training of the reconstruction network.

(*b*) shows the effect of the mask region on the segmentation accuracy with fixed size at the center position, random size at the center position, random both position and size and our method, respectively. We see that the segmentation results when the mask region is set with a random position and random size outperform the segmentation results with random size at the center position, and our method obtains the highest DSC. This indicates that the location and size of suitable masks are different for different images. Therefore, our AMP-DRL to select more appropriate mask regions for each image can more effectively help the reconstruction network learn more fine-grained image representation information and thus improve the downstream segmentation model performance.

In addition, we conducted experiments on the Cardiac dataset, comparing the visual activation maps of the objects of interest of the random masking policy and our policy. The results are shown in Figure 4. We found that the random masking policy is focused on random areas, while ours is more focused on the tumor regions. This enables our policy to more likely mask complex areas in the image, increase the difficulty of reconstruction, and thus obtain better pre-training weights.

V. Conclusion

In this work, we propose a self-supervised method of adaptive-masking policy, called AMP-DRL. AMP-DRL uses the learned adaptive-masking policy to select a more appropriate mask region for each image, which more effectively helps the reconstruction network to learn more fine-grained image representation information and thus improves the downstream segmentation model performance. We conducted extensive experiments on two datasets, Cardiac and TCIA, and the results showed that our approach outperformed the current SOTA self-supervised methods.

Given the abundance of mutual information present in multimodal medical image data and the issue of imbalanced data, it is worthwhile to conduct further research on the application of MIM methods in tasks related to the analysis of multimodal [26] and imbalanced [27] medical images. By doing so, it may be possible to extract more informative representations and improve the performance of deep models.

References

[1] Olaf Ronneberger, Philipp Fischer, and Thomas Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proceedings of MICCAI*, 2015, pp. 234–241.

[2] Zhenghua Xu, Tianrun Li, Yunxin Liu, Yuefu Zhan, Junyang Chen, and Thomas Lukasiewicz, "PAC-Net: Multi-pathway FPN with position attention guided connections and vertex distance IoU for 3D medical image detection," *Frontiers in Bioengineering and Biotechnology*, vol. 11, pp. 1049555, 2023.

[3] Zhenghua Xu, Shijie Liu, Di Yuan, Lei Wang, Junyang Chen, Thomas Lukasiewicz, Zhigang Fu, and Rui Zhang, "ω-net: Dual supervised medical image segmentation with multi-dimensional self-attention and diversely-connected multi-scale convolution," *Neurocomputing*, vol. 500, pp. 177–190, 2022.

[4] Zhenghua Xu, Chang Qi, and Guizhi Xu, "Semi-supervised attention-guided CycleGAN for data augmentation on medical images," in *Proceedings of IEEE BIBM*, 2019, pp. 563–568.

[5] Saeed Shurrab and Rehab Duwairi, "Self-supervised learning methods and applications in medical imaging analysis: A survey," *ArXiv preprint ArXiv:2109.08685*, 2021.

[6] Saeed Shurrab and Rehab Duwairi, "Self-supervised learning methods and applications in medical imaging analysis: A survey," *PeerJ Computer Science*, vol. 8, pp. e1045, 2022.

[7] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton, "A simple framework for contrastive learning of visual representations," in *Proceedings of ICML*, 2020, pp. 1597–1607.

[8] Jean-Bastien Grill, Florian Strub, Florent Altché, Corentin Tallec, Pierre Richemond, Elena Buchatskaya, et al., "Bootstrap your own latent: A new approach to self-supervised learning," *NeurIPS*, vol. 33, pp. 21271–21284, 2020.

[9] Mathilde Caron, Ishan Misra, Julien Mairal, Priya Goyal, Piotr Bojanowski, and Armand Joulin, "Unsupervised learning of visual features by contrasting cluster assignments," *NeurIPS*, vol. 33, pp. 9912–9924, 2020.

[10] Deepak Pathak, Philipp Krahenbuhl, Jeff Donahue, Trevor Darrell, and Alexei A Efros, "Context Encoders: Feature learning by inpainting," in *Proceedings of CVPR*, 2016, pp. 2536–2544.

[11] Kaiming He, Xinlei Chen, Saining Xie, Yanghao Li, Piotr Dollár, and Ross Girshick, "Masked autoencoders are scalable vision learners," in *Proceedings of CVPR*, 2022, pp. 16000–16009.

[12] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu, "SimMIM: A simple framework for masked image modeling," in *Proceedings of CVPR*, 2022, pp. 9653–9663.

[13] Peng Gao, Teli Ma, Hongsheng Li, Jifeng Dai, and Yu Qiao, "Convmae: Masked convolution meets masked autoencoders," *arXiv preprint arXiv:2205.03892*, 2022.

[14] Mehrnaz Abdollahian and Tapas K Das, "A MDP model for breast and ovarian cancer intervention strategies for BRCA1/2 mutation carriers," *IEEE J BIOMED HEALTH*, vol. 19, no. 2, pp. 720–727, 2014.

[15] Xin Duan, Xiabi Liu, Xiaopeng Gong, and Mengqiao Han, "RL-CoSeg: A novel image co-segmentation algorithm with deep reinforcement learning," *arXiv preprint arXiv:2204.05951*, 2022.

[16] Ziyu Wang, Tom Schaul, Matteo Hessel, Hado Hasselt, Marc Lanctot, and Nando Freitas, "Dueling network architectures for deep reinforcement learning," in *Proceedings of ICML*, 2016, pp. 1995–2003.

[17] Di Yuan, Yunxin Liu, Zhenghua Xu, Yuefu Zhan, Junyang Chen, and Thomas Lukasiewicz, "Painless and accurate medical image analysis using deep reinforcement learning with task-oriented homogenized automatic pre-processing," *Computers in Biology and Medicine*, vol. 153, pp. 106487, 2023.

[18] Yunze Man, Yangsibo Huang, Junyi Feng, Xi Li, and Fei Wu, "Deep Q learning driven CT pancreas segmentation with geometry-aware U-Net," *Transactions on Medical Imaging*, vol. 38, no. 8, pp. 1971–1980, 2019.

[19] Tiexin Qin, Ziyuan Wang, Kelei He, Yinghuan Shi, Yang Gao, and Dinggang Shen, "Automatic data augmentation via deep reinforcement learning for effective kidney tumor segmentation," in *Proceedings of ICASSP*, 2020, pp. 1419–1423.

[20] Zhe Li and Yong Xia, "Deep reinforcement learning for weakly-supervised lymph node segmentation in CT images," *IEEE JBHI*, vol. 25, no. 3, pp. 774–783, 2020.

[21] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, et al., "Human-level control through deep reinforcement learning," *Nature*, vol. 518, no. 7540, pp. 529–533, 2015.

[22] Michela Antonelli, Annika Reinke, Spyridon Bakas, Keyvan Farahani, Bennett A Landman, Geert Litjens, et al., "The medical segmentation decathlon," *arXiv preprint arXiv:2106.05735*, 2021.

[23] Amber L Simpson, Michela Antonelli, Spyridon Bakas, Michel Bilello, Keyvan Farahani, Bram Van Ginneken, et al., "A large annotated medical image dataset for the development and evaluation of segmentation algorithms," *arXiv preprint arXiv:1902.09063*, 2019.

[24] Kenneth Clark, Bruce Vendt, Kirk Smith, John Freymann, Justin Kirby, Paul Koppel, et al., "The cancer imaging archive (tcia): maintaining and operating a public information repository," *Journal of digital imaging*, vol. 26, pp. 1045–1057, 2013.

[25] Diederik P Kingma and Jimmy Ba, "Adam: A method for stochastic optimization," *arXiv preprint arXiv:1412.6980*, 2014.

[26] Shuo Zhang, Jiaojiao Zhang, Biao Tian, Thomas Lukasiewicz, and Zhenghua Xu, "Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation," *Medical Image Analysis*, vol. 83, pp. 102656, 2023.

[27] Jianfeng Wang, Thomas Lukasiewicz, Xiaolin Hu, Jianfei Cai, and Zhenghua Xu, "RSG: A simple but effective module for learning imbalanced datasets," in *Proceedings of CVPR*, 2021, pp. 3784–3793.