Computers in Biology and Medicine 169 (2024) 107877

Contents lists available at ScienceDirect

Computers in Biology and Medicine

journal homepage: www.elsevier.com/locate/compbiomed

Automatic data augmentation for medical image segmentation using Adaptive Sequence-length based Deep Reinforcement Learning

Zhenghua Xu [a], Shengxin Wang [a], Gang Xu [b], Yunxin Liu [a], Miao Yu [a,*], Hongwei Zhang [c], Thomas Lukasiewicz [d,e], Junhua Gu [b]

[a] State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Health Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, China
[b] School of Artificial Intelligence, Hebei University of Technology, Tianjin, China
[c] School of Computer Science and Engineering, Tianjin University of Technology, Tianjin, China
[d] Institute of Logic and Computation, Vienna University of Technology, Vienna, Austria
[e] Department of Computer Science, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:
Adaptive Sequence-length Deep Reinforcement Learning
Automatic Data Augmentation
Reward function
Medical image segmentation

ABSTRACT

Although existing deep reinforcement learning-based approaches have achieved some success in image augmentation tasks, their effectiveness and adequacy for data augmentation in intelligent medical image analysis are still unsatisfactory. Therefore, we propose a novel Adaptive Sequence-length based Deep Reinforcement Learning (ASDRL) model for Automatic Data Augmentation (AutoAug) in intelligent medical image analysis. The improvements of ASDRL-AutoAug are two-fold: (i) To remedy the problem of some augmented images being invalid, we construct a more accurate reward function based on different variations of the augmentation trajectories. This reward function assesses the validity of each augmentation transformation more accurately by introducing different information about the validity of the augmented images. (ii) Then, to alleviate the problem of insufficient augmentation, we further propose a more intelligent automatic stopping mechanism (ASM). ASM feeds a stop signal to the agent automatically by judging the adequacy of image augmentation. This ensures that each transformation before stopping the augmentation can smoothly improve the model performance. Extensive experimental results on three medical image segmentation datasets show that (i) ASDRL-AutoAug greatly outperforms the state-of-the-art data augmentation methods in medical image segmentation tasks, (ii) the proposed improvements are both effective and essential for ASDRL-AutoAug to achieve superior performance, and the new reward evaluates the transformations more accurately than existing reward functions, and (iii) we also demonstrate that ASDRL-AutoAug is adaptive for different images in terms of sequence length, as well as generalizable across different segmentation models.

1. Introduction

In recent years, deep learning models have achieved remarkable success in various computer vision tasks, such as image classification [1], object detection [2], segmentation [3–5], super-resolution and denoising [6], which have contributed to the booming development of automated medical image analysis. Automated medical image analysis is beneficial to help clinicians save time and effort in disease diagnosis, treatment plan design, and prognosis assessment, and reduce the risk of misdiagnosis and missed diagnosis. However, limited by the volume of medical image data, deep learning models for automated medical image analysis usually suffer from overfitting [7], which reduces the generalization ability of the model and prevents reliable results.

Data augmentation increases the amount and diversity of training data by applying semantically invariant image transformations and is the most commonly used regularization technique to combat overfitting, due to its simplicity and effectiveness [8–10]. Traditional data augmentation methods generate training images by applying basic operations such as geometric transformations (e.g., rotation, cropping) and pixel-level transformations (e.g., noise addition, sharpening) on the training images [11–14], but these methods usually randomly select a sequence of transformations for each image or rely on human experts with a priori knowledge of the dataset to craft a sequence of transformations to be applied in training, resulting in training samples with insufficient diversity. Therefore, some advanced data augmentation operations, such as Random Erasing [15], KeepAugment [16], and

* Corresponding author.
E-mail address: 202011402001@stu.hebut.edu.cn (M. Yu).

https://doi.org/10.1016/j.compbiomed.2023.107877
Received 28 June 2023; Received in revised form 3 December 2023; Accepted 18 December 2023
Available online 21 December 2023

SaliencyMix [17] have been proposed, which generate richer images by region erasing or sample mixing. In addition, data augmentation based on generative adversarial networks [18] improves the performance of medical image analysis tasks by generating virtual data with the same distribution as the real data. All these methods improve the accuracy of the model to some extent, but a common limitation is that the change in the accuracy of the model cannot be fed back to adjust the data augmentation network. So, it is difficult to determine the data augmentation operation that improves the model performance best.

To address the above problem, Cubuk et al. [19] pioneered an automatic data augmentation method called AutoAugment, which is based on deep reinforcement learning (DRL). AutoAugment automatically searches for a sequence of transformations that can maximize the performance of the classifier. Although this approach reduces the classification error rate to some extent, this search process requires repeatedly training the sub-model sufficiently with different augmentation policies to obtain feedback on the agent. So, AutoAugment is not only computationally intensive but also time-consuming. In addition, AutoAugment aims to find the best augmentation sequence for the whole dataset and then apply it to each training image. However, the optimal combination of transformations is different for each image, so some augmented images may not be useful or even harmful for model training.

Aware of these limitations, researchers have recently started to focus on improved approaches to AutoAugment through grid search [20], approximation algorithms for bi-level problems [21,22], and learning sample-specific augmentation sequences [23–25] to improve the speed and effectiveness of image augmentation. However, these existing methods suffer from the following two problems. (i) **Some of the augmented images are invalid.** Since the reward functions of these methods only consider the relative change in model accuracy between two adjacent augmentations, their assessment of the validity and necessity of each augmentation transformation is inaccurate. As shown in the rising stage of the red line in Fig. 1, the performance of the model at this point is lower than the initial performance. This case indicates that the transformations selected by the agent for the current image are invalid. However, the environment of existing methods gives a positive reward value back to the agent in this case, which encourages the agent to select these invalid transformations and then produce invalid images. (ii) **The augmentation of some images is insufficient.** These methods set the maximum length of the transformation sequence, i.e., the maximum number of transformations, based on human priori knowledge. For images from different datasets or different images from the same dataset, the limitation of a uniform length will result in the augmentation of some images being inadequate. As shown by the blue line in Fig. 1, the model performance can still be improved if the augmentation of images continues after reaching the maximum number of times that has been artificially set.

Therefore, in this work, we propose a novel **A**daptive **S**equence-length based **D**eep **R**einforcement **L**earning (ASDRL) model for **Auto**matic data **Aug**mentation (AutoAug) in intelligent medical image analysis, denoted ASDRL-AutoAug, where the problem of partially augmented images being ineffective and the problem of augmentation of partial images being insufficient are remedied using the reward function with dual constraints (R_{dc}) and the automatic stopping mechanism (ASM), respectively. Specifically, in deep reinforcement learning, the agent learns the ASDRL-AutoAug policy by interactively optimizing training with the task model using R_{dc} and then is gradually able to sequentially select effective transformations for each input image. By using ASM, this sequential decision process is not limited by the maximum number of times, which makes the sequence of transformations for this image have a suitable length. Overall, the proposed ASDRL-AutoAug has the following two main improvements. First, to remedy the problem of invalid augmented images due to inaccurate evaluation, we propose a more accurate reward function R_{dc} based on the different variations of the augmentation trajectories. In R_{dc}, we introduce an absolute difference between the model accuracy after each augmentation and the initial accuracy while preserving the relative difference of the model accuracy between two adjacent augmentations. These two differences contain different information about the validity of each transformation, so we use these two differences to provide dual constraints for R_{dc}, which makes its assessment of the validity of each transformation more accurate.

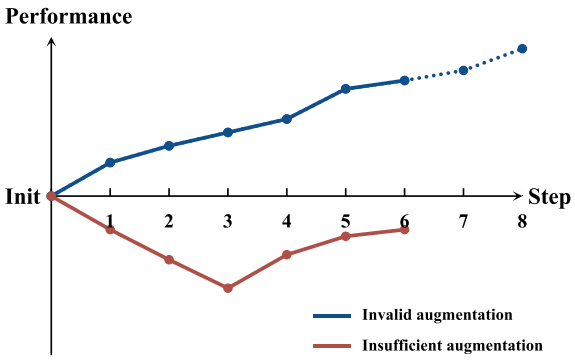

Fig. 1. Augmentation trajectory, representing the change of segmentation model accuracy caused by each augmentation of an image. The horizontal axis indicates the number of augmentations, the vertical axis indicates the model accuracy, and the origin indicates the initial model accuracy, which is a non-zero value. The red line is an example of augmentation trajectories when some of the augmented images are invalid while the blue line indicates an example when the augmentation of some images is insufficient.

In addition, to alleviate the problem of insufficient augmentation of some images, we propose an intelligent automatic stopping mechanism ASM, which first judges the adequacy of image augmentation by the change of the augmentation trajectory of each image. Then, ASM makes the environment feed a stop signal for the agent when the augmentation is sufficient, which makes the agent automatically stop the augmentation process of that image. This automatic stopping method enables the agent to learn the ability to make an increasing trend of the augmentation trajectory in a nearly infinite search space, i.e., each transformation before stopping the augmentation can smoothly improve the model performance. So, the agent decides on the sequence of transformations with the appropriate length for each image, which makes the image augmentation more sufficient.

The main contributions of this paper can be summarized as the following three points:

- We identify two problems in the existing AutoAugment family and then propose an Adaptive Sequence-length based Deep Reinforcement Learning (ASDRL) model for Automatic data Augmentation (AutoAug) in intelligent medical image analysis, denoted ASDRL-AutoAug, to remedy these problems thereby improving the performance of the existing AutoAugment family.
- In ASDRL-AutoAug, we innovatively propose a reward function with double constraints R_{dc}. R_{dc} remedies the problem of invalid augmented images by more information about the validity of the transformation to more accurately evaluate the validity of each transformation. Then, in order to solve the problem of insufficient augmentation of some images due to human constraints in existing methods, we further innovatively propose an automatic stopping mechanism (ASM). ASM automatically feeds the stopping signal to the agent by assessing the adequacy of image augmentation, ensuring that each transformation before stopping the augmentation can steadily improve the model performance, and thus improve the adequacy of image augmentation.
- Extensive experimental studies on three public medical image datasets show that (i) the proposed ASDRL-AutoAug achieves better performance than state-of-the-art data augmentation methods in downstream segmentation tasks, and (ii) both the proposed

R_{dc} and ASM are effective and essential for ASDRL-AutoAug to achieve superior performance, and R_{dc} evaluates the transformation more accurately than existing reward functions. (iii) we also demonstrate the adaptiveness of ASDRL-AutoAug to different images in terms of sequence-length, and its universality to different segmentation models.

The rest of this paper is organized as follows. Section 2 presents the related work and clarifies the differences between our work and the references. The detailed methodology of the proposed ASDRL-AutoAug is given in Section 3, which is followed by the experimental studies in Section 4. Section 5 discusses the social impact, limitations, and future works of the proposed work, and the conclusion is given in Section 6.

2. Related work

Basic image augmentation. Basic image augmentation transformations include geometric and non-geometric transformations, such as rotation, crop, flip, zoom, noise injection, contrast, and kernel filtering. Such transformations help improve the generalization performance of the model by modifying the geometry and visual features of the image to expose the model to a wider range of image variations during training. Being relatively efficient and easy to implement, basic image augmentation transformations have been widely used for a variety of image-related tasks in recent years. For example, to improve the performance of classification models, 3-Augment [13] randomly selects one of the three data transformations grayscale, Gaussian blur, and solarization with equal probability to augment each image. Han et al. [11] apply horizontal flipping and translation by 4 pixels to the images in CIFAR-10 and CIFAR-100 for augmentation during training. In detection, Liu et al. [12] improve the detection model accuracy by resizing each sample block to a fixed size and flipping it horizontally with a probability of 0.5. Yang et al. [14] use cropping and flipping for semi-supervised semantic segmentation for weak augmentation to improve the performance of segmentation on medical image datasets through the weak-to-strong consistency framework. Furthermore, for biomedical images, Bloice et al. [26] propose the Augmentor package, which allows the user to achieve stochastic, pipeline-based data augmentation by controlling the probability parameter of each transformation.

The above works use traditional data augmentation methods, i.e., randomly selecting a sequence of basic transformations for each image or relying on a human expert with prior knowledge of the dataset to craft a sequence of transformations to be applied in the training. Since randomness tends to introduce significant noise into the data and manually selected transformations based on the nature of the dataset are not applicable to each image, random and manual augmentation strategies are not optimal for improving model performance. However, our proposed ASDRL-AutoAug can automatically learn the appropriate combination of basic transformations for each image based on the task model to better improve the model performance.

Advanced image augmentation. To force the model to better learn the representational information of the image, two advanced image augmentation transformations, image erasing and image mixing, have been proposed. Image erasing involves removing one or more sub-regions of an image and replacing the pixel values of these sub-regions with constant values, random values, or the average of the entire dataset. For example, DeVries and Taylor [27] propose the Cutout, which randomly masks a square region on the training image during the training of the convolutional neural network and can improve the robustness and overall performance of the convolutional neural network. Kumar Singh and Jae Lee [28] propose Hide-and-Seek, which divides the training image into uniform squares of random size and then randomly hides a specified number of these squares, aiming to force the neural network to learn relevant features by hiding the most discriminative content. Zhong et al. [15] propose Random Erasing, which selects a rectangular region in an image by randomly deciding whether to mask the region and also randomly deciding the aspect ratio and size of the masked region. Chen et al. [29] propose a structured method, called GridMask, to apply multi-scale grid masks to images to simulate occlusion, aiming to trade off complete object erasing with stripping contextual information. While image erasing is effective in some applications, such methods will lead to the removal of important information, negatively impacting the performance of the model.

Image mixing is divided into two categories based on the number of images used in each mixing. (i) Single-image mixing techniques: a single image sub-region is mixed from different mixing perspectives. For example, Kim et al. [30] propose LocalAugment, which increases the diversity of local features to improve the generalization performance of neural networks by dividing the image into smaller blocks and applying different data augmentations to each block; Seo et al. [31] propose self-augmentation, which improves the generalization performance by cropping random regions of the image and then pasting them to random locations in the image generalization ability of few-shot learning; Choi et al. [32] propose SalfMix, which crops most salient regions and places them into non-salient regions based on saliency map; Gong et al. [16] propose KeepAugment, which increases fidelity by preserving salient features of images and increasing non-salient regions to increase diversity without changing the distribution. (ii) Multi-image mixing techniques: applying different mixing strategies to mix two or more images into one image. For example, the mixup proposed by Zhang et al. [33] improves the model accuracy and robustness by mixing two images and corresponding labels based on a mixing factor (alpha). Yun et al. [34] propose CutMix to solve the problem of information loss and region loss by randomly selecting a square region on the training image and filling it with patches from another image. SaliencyMix proposed by Uddin et al. [17] avoids the model from learning unnecessary information by selecting a salient part of an image and pasting it into another image.

While these methods achieve promising improvements on their corresponding tasks by generating richer images, the challenge lies in the inability to feed back variations in model accuracy to adjust the data augmentation strategy. Consequently, the stochastic strategy they employ becomes challenging to balance increasing the amount of data while maintaining the integrity of information in the images. In contrast, ASDRL-AutoAug evaluates each augmentation transformation based on our newly proposed reward function. It learns through this feedback, enabling the obtention of sequences of transformations that result in improved model performance.

Automatic data augmentation. In recent years, to solve the problems caused by manual or random augmentation strategies, automatic data augmentation methods are proposed to search for the optimal augmentation policy for a given dataset. After validating the effectiveness of AutoAugment in classification [19] and detection [35] tasks, RandAugment [20] is further proposed to select the transformation with uniform probability 1/K and automatically find the number and amplitude of transformations by grid search. To reduce the search time, the subsequent work has two main directions of improvement. On the one hand, both Fast AutoAugment [36] and Faster AutoAugment [37] use an efficient search strategy based on density matching to find more optimal sequences of transformations. Addressing optimization speed challenges in works like AutoAugment and Fast AutoAugment, [38] propose Differentiable Automatic Data Augmentation (DADA), reducing costs by relaxing discrete data augmentation strategies to differentiable optimization problems through Gumbel-Softmax. Subsequently, Tang et al. [39] propose AutoPedestrian, which models the data augmentation strategy and the loss function strategy as two distributions with different hyper-parameters. The method then finds the data augmentation strategy and the loss function simultaneously via an automatic pedestrian scheme with significant sampling. Zhou et al. [40] propose IDADA to solve the error problem of discrete-continuous coding in DADA and the gradient estimation inaccuracy

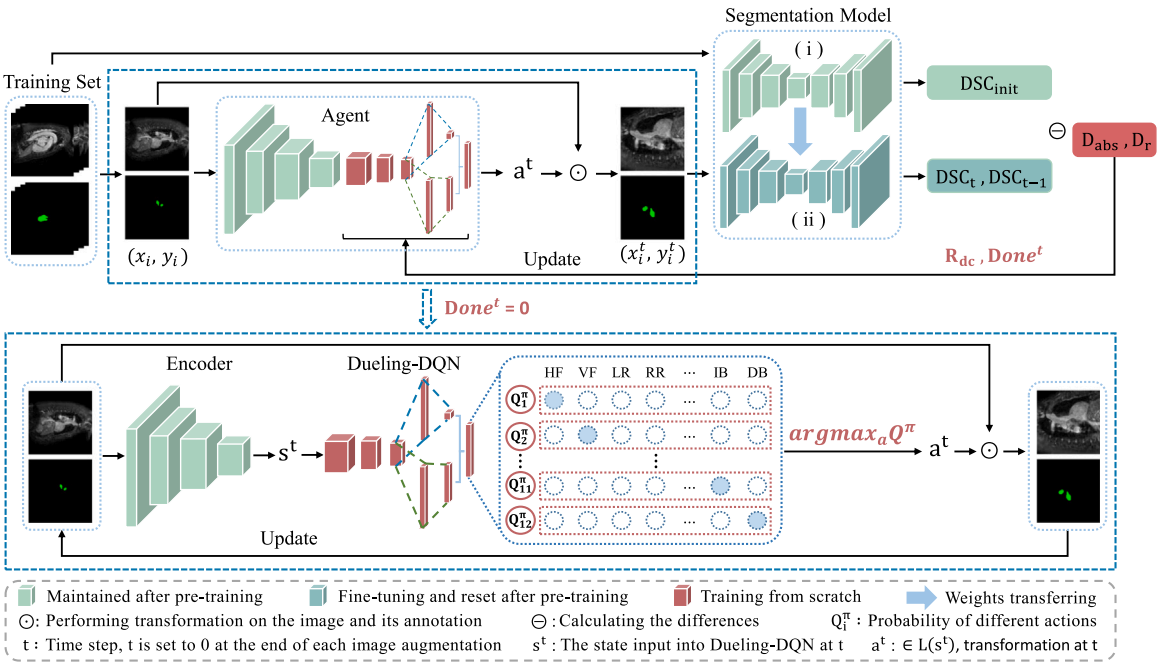

Fig. 2. Overview of our adaptive sequence-length based deep reinforcement learning (ASDRL) model for automatic data augmentation (AutoAug) in intelligent medical image analysis (ASDRL-AutoAug). We demonstrate ASDRL-AutoAug using the Cardiac dataset as an example, where (x_i, y_i) denotes the original image and its annotation, and (x_i^t, y_i^t) denotes the image and its annotation after the t-th augmentation. Specifically, ASDRL-AutoAug mainly consists of two parts: the agent and the segmentation model, where the agent includes the encoder and the Dueling-DQN, which are used for extracting the image features and deciding the augmentation transformation, respectively; there are two segmentation models: (i) and (ii) have the same structure, with (i) pre-trained by the training set, and the initial weights of (ii) migrated from (i), and at each time step (ii) is fine-tuned by (x_i^t, y_i^t) and reset to the initial weights at the end of (x_i, y_i) augmentation. The agent updates Dueling-DQN according to the reward with dual constraints (R_{dc}) and performs adaptive sequence-length augmentation on the current image according to $Done^t$ in the Automatic Stopping Mechanism(ASM). In particular, the processed image (x_i^t, y_i^t) is used as input of the agent at t+1 until the end of the augmentation of (x_i, y_i).

problem so that the selection probability of the augmentation operation is tilted towards its extreme value, which achieves better results in the pedestrian detection task. However, the augmentation strategies in the above methods are applied to the entire dataset and are not image-specific, so the augmentation may not be suitable for each image. Furthermore, in these methods, the length of the sub-strategy is set to 2, so each image can only be transformed twice per augmentation. This restriction on the number of transformations makes the augmentation insufficient for some images.

On the other hand, in order to reduce the search time, AWS [21] improves the bi-level optimization process of automatic data augmentation by using a two-stage approximation algorithm to complete the evaluation of the image transformations without training the classifier from scratch. Subsequently, based on the approximation algorithm of the bi-level optimization process, Minh et al. [24] and Li et al. [23] improve the downstream classification and segmentation accuracy by jointly training the task network and the agent, so that the agent automatically makes decisions on the transformations sequence used for each image. However, the reward functions of these methods contain less information about the effectiveness of augmentation, which prevents the reward function from accurately assessing the goodness of the augmentation transformations, thus rendering some of the transformed data invalid. In addition, the above methods impose a limit on the number of transformations per image, which makes the augmentation of some images inadequate.

In particular, several works migrate automatic data augmentation methods to medical image augmentation. ASNG [22] defines the problem of finding an optimal augmentation strategy as a discrete search problem. By designing a search space and a search algorithm, ASNG automatically searches for the augmentation strategy for the medical image segmentation task and determines the magnitude of the augmentation operation and the probability of applying the operation. SLSRL [41] employs reinforcement learning to search for the probability of applying each augmentation operation, shaping the augmentation strategy for a specific dataset. He et al. [42] migrate DADA to medical image segmentation and design a dedicated search space and optimization strategy for the medical image segmentation task. ADA-DRL [25] enables the agent to find a specific data augmentation sequence for each kidney tumor image by interacting with a downstream segmentation model. MedAugment [43] designs a new augmentation space and an operation sampling strategy to automatically select the magnitude and probability of an augmentation operation for a medical image dataset. While automatic data augmentation methods designed for medical images share similarities with those for natural images in terms of the augmentation process, it is crucial to note that medical images exhibit greater sensitivity to the selection of augmentation operations. In addition, the above works also set the maximum number of augmentations, which makes some of the augmented medical images invalid and some of the medical images insufficiently augmented.

In conclusion, the existing methods suffer from the invalidation of some transformed images and insufficient augmentation of some images, which results in non-optimal basic transformation sequences for some images. Our work aims to address the above issues and use deep reinforcement learning to find more valid image augmentation sequences for downstream tasks, and thus improve the accuracy of the model more sufficiently.

3. Adaptive sequence-length based deep reinforcement learning model for automatic data augmentation

Fig. 2 illustrates the overall flow of the proposed Adaptive Sequence-length based Deep Reinforcement Learning (ASDRL) model for Automatic data Augmentation (AutoAug) in intelligent medical image analysis (ASDRL-AutoAug). First, we conduct a fully supervised pre-training of the randomly initialized U-Net segmentation model [44]. The objective is to attain a model with good segmentation accuracy on the original training set, establishing an absolute constraint on the

Table 1
Transformation types and their amplitudes used in ASDRL-AutoAug. These transformations are defined as actions.

Action	Operation	Amplitude
HF	Horizontally Flip	–
VF	Vertically Flip	–
LR/RR	Left/Right Rotate	30°
CL/CR	Crop from Left/Right	20
CU/CD	Crop from Up/Down	20
ZM	ZooM	1.1
AN	Add Noise	Gaussian
IB	Increase Brightness	0.1
DB	Decrease Brightness	0.1

reward signal during data augmentation. Second, during data augmentation, we construct a new reward (see Section 3.2) using both the absolute constraint and another relative constraint, which aims to more accurately evaluate the impact of the augmentation transformation on the segmentation model performance. This reward facilitates the agent in gradually and automatically selecting the best augmentation transformation for each input image according to the segmentation task. Additionally, through our proposed stopping method, ASM (see Section 3.3), the agent can more intelligently find the sequence of augmentation transformations for each image that sufficiently improves the performance of the segmentation model.

Specifically, in the pre-training, we take the initial training set $\mathbb{N} = (x_1, y_1), \ldots, (x_n, y_n)$ as the input of U-Net and obtain a model $U(\sigma)$ with good segmentation accuracy on the validation set $\mathbb{V} = (z_1, w_1), \ldots, (z_m, w_m)$, where x_i/z_i and y_i/w_i are the ith images and its corresponding ground truth, respectively. In the data augmentation process, we consider the search for the best transformation sequence as a sequential decision process. First, the agent automatically selects the optimal augmentation operation for the input image. Second, the transformed image $a \odot x_i$ and its label $a \odot y_i$ are fed into the segmentation model $U(\sigma)$ for fine-tuning. Then, a reward is constructed by comparing the difference in segmentation accuracy before and after the augmentation from the perspective of two constraints. Finally, the reward is fed back to the agent to optimize the policy. In contrast to other methods where the maximum number of augmentations is uniformly set for all images based on manual experience, this augmentation process has no limit on the maximum number of augmentations. The agent automatically stops the augmentation process for each image when the augmentation is deemed sufficient, based on the change in the augmentation trajectory of that image.

In the above process, through interactive training with the segmentation model, the agent progressively selects for each image an augmentation transformation sequence (a^0, a^1, …, a^t). With this augmentation sequence, the input image x_i is transformed to x_i^t at step t, and the corresponding label y_i is transformed to y_i^t at the same time:

$$x_i^t = a^t \odot a^{t-1} \odot \cdots \odot a^1 \odot x_i, \tag{1}$$

$$y_i^t = a^t \odot a^{t-1} \odot \cdots \odot a^1 \odot y_i, \tag{2}$$

where $\odot$ denotes the operator of the augmentation transformation, and the transformed image of each step in the sequence is used as the input image for the next transformation. Finally, the augmented images form an augmentation training set for the downstream segmentation task.

3.1. Training process

Algorithm 1 summarizes the implementation of the ASDRL-AutoAug. In ASDRL-AutoAug, we use 12 classical data augmentation transformations as actions, as shown in Table 1. These 12 actions form the action space and are formally defined as:

$$L(s^t) = [HF, VF, LR, RR, CL, CR, CU, CD, ZM, AN, IB, DB], \tag{3}$$

where s^t denotes the state sent to the agent at time step t, which contains available information about the image content for the agent to make decisions. Since the speed of the agent is low when learning directly on large-size images, we use the encoder of the model $U(\sigma)$ as the feature extraction module $E(\sigma)$, so that the image x_i is represented stably on a high-dimensional latent space before being input to the agent. Thus, s^t is formalized as:

$$s^t = E(x_i; \sigma). \tag{4}$$

Since the actions are discrete, we use a Dueling Deep Q-learning Network (Dueling DQN) [45] as the agent to learn the data augmentation policy. Specifically, Dueling DQN takes the state s^t as input, then estimates the impact of all actions on the performance of the segmentation model according to the Q-function, and finally selects the actions with the maximum Q-value:

$$a^t = argmax_{a \in L(s^t)} Q^\pi(s^t, a; \theta, \alpha, \beta), \tag{5}$$

$$Q^\pi(s^t, a^t; \theta, \alpha, \beta) = V(s^t; \theta, \beta) + (A(s^t, a^t; \theta, \alpha) - \frac{1}{|A|} \sum_{a^{t+1}} A(s^t, a^{t+1}; \theta, \alpha)), \tag{6}$$

where $Q^\pi(s^t, a^t; \theta, \alpha, \beta)$ is the Q-function fitted by Dueling DQN, θ is a parameter of the convolutional neural network shared by the state-value function $V(\theta, \beta)$ and the action advantage value function $A(\theta, \alpha)$, and α, β are parameters owned independently by each. We use the mean square error loss to evaluate the error between the Q-value and the actual variation q_i of the segmentation model performance, which is formally written as:

$$Loss(\theta, \alpha, \beta) = (q_i - Q^\pi(s^t, a^t; \theta, \alpha, \beta))^2, \tag{7}$$

$$q_i = \begin{cases} r^t \\ r^t + \gamma max_{a^{t+1}} Q^\pi(s^{t+1}, a^{t+1}; \theta, \alpha, \beta), \end{cases} \tag{8}$$

where r^t denotes the immediate reward that the environment feeds to the agent at each time step t. Specifically, r is a scalar value that helps the agent evaluate whether the transformation is good or bad, and t denotes the time step. q_i has two kinds of assignments. If the augmentation ends, q_i is assigned the current reward value r^t; otherwise, it is assigned the weighted sum of the current reward value and the further reward value, where γ is a hyperparameter. We minimize this loss function so that Dueling DQN learns the policy to perform data augmentation for each image. Additionally, the process by which the agent acquires its capabilities is the process of trial-and-error exploration of the environment and an exploitation process of taking actions that are known to maximize rewards [46]. To balance the exploration and exploration of the agent, we use the *ϵ-decreasing* algorithm [47] for the training of Dueling DQN.

3.2. Reward with dual constraints

The design of the reward is critical to accurately assess the validity of each augmentation transformation, which facilitates guiding the agent to accurately select augmentation transformations. However, existing work only considers the relative changes in model accuracy between two adjacent transformations, which provides insufficient information about the validity of each transformation. This limitation makes the rewards inaccurate in assessing the validity of each augmentation transformation, leading to some of the augmented data being invalid. Therefore, we propose a new reward R_{dc} with dual constraints. We introduce the absolute difference in model accuracy to provide different information about the validity of each transformation to the reward function. This improvement allows R_{dc} to more accurately assess the validity of each transformation, thereby alleviating the problem of some augmented images being invalid.

The most intuitive performance evaluation metric in image segmentation is the Dice Similarity Coefficient (DSC), as shown in Eq. (14).

Algorithm 1 ASDRL-AutoAug

Input: Condition counter j_{max}; training epochs T; original training set $\mathbb{N} = (X, Y)$; validation set $\mathbb{V} = (Z, W)$; augmentation training set $\mathbb{G} = \emptyset$; segmentation loss function using dice loss $\mathcal{L}_s$ and Dueling-DQN loss function using MSE loss $\mathcal{L}_p$; segmentation network U parameterized with σ; the Dueling-DQN Q parameterized with θ, α, and β.

Initialize model parameter σ by the original training set $\mathbb{N}$
Initialize model parameter θ, α, and β randomly
Get the initial segmentation dice $DSC_{init} = Dice\ (U_\sigma(Z), W)$
for $epoch = 1, \ldots, T$ **do**
  Sample a training image (x, y) from the (X, Y) pool
  Load the segmentation network U_σ
  Initialize $t = 0$, $j = 0$, $r^t = 0$
  repeat
    Load the encoder of U_σ as the feature extraction module E_σ to get the state $s^t = E_\sigma(x)$
    With probability ϵ, select a random augmentation operation a^t;
    otherwise, select $a^t = \text{argmax}_a Q(s^t, a; \theta, \alpha, \beta)$
    Performing action a^t on (x, y) to get the augmentation image (x^t, y^t)
    Optimize the model parameter σ by fine-tuning with (x^t, y^t) to get σ^*
    Calculate a new dice $DSC_t = Dice\ (U_{\sigma^*}(Z), W)$
    Calculate the difference between the two segmentation dices DSC_t and DSC_{t-1}, DSC_t and DSC_{init}, respectively
    $D_r = DSC_t\text{-}DSC_{t-1}$
    $D_{abs} = DSC_t\text{-}DSC_{init}$
    if $D_{abs} > 0$
      Update reward $r^t = D_r$
    else if $D_{abs} \leq 0$ & $D_r > 0$
      Update reward $r^t = -e^{-D_r}$
    else
      Update reward $r^t = -m$
    end if
    if $r^t < 0$
      $j \leftarrow j + 1$
    end if
    Update $(x, y) = (x^t, y^t)$
    Update $t = t + 1$
    Perform a gradient descent step on $\mathcal{L}_p$ with respect to the network parameters θ, α, and β
  until $j = j_{max}$
  $\mathbb{G} = \mathbb{G} \cup (x, y)$
end for

Output: Augmentation training set $\mathbb{G}$

A larger DSC indicates better segmentation model performance, while a smaller DSC suggests poorer performance. Therefore, we use the DSC to calculate and constrain the value of the reward, aiming to encourage the agent to select the best augmentation transformation that improves the performance of the segmentation model. Specifically, we first obtain the initial segmentation accuracy DSC_{init} of the model $U(\sigma)$ on the validation set $\mathbb{V}$. Then, we use the image $a^t \odot x_i^{t-1}$ after the tth transformation and its label $a^t \odot y_i^{t-1}$ to fine-tune the model $U(\sigma)$ to obtain $U(\sigma)_t$ and its segmentation accuracy DSC_t. After that, the difference D_{abs} between the two segmentation accuracies is calculated:

$$D_{abs} = DSC_t - DSC_{init}. \tag{9}$$

We refer to the difference D_{abs} between the segmentation accuracy of the model $U(\sigma)$ after each fine-tuning and its initial segmentation accuracy as the absolute difference. D_{abs} represents the impact of this augmented image on the performance of the model $U(\sigma)$ after t times transformations. Therefore, the absolute difference D_{abs} contains information about the effectiveness of performing multi-step transformations (i.e., augmentation sequences) on the original image x_i. Introducing D_{abs} for R_{dc} facilitates the evaluation of each transformation in terms of the validity of the sequence.

Furthermore, in the iterative training of the agent, we calculate the difference D_r between DSC_t and the DSC_{t-1}. Here, DSC_{t-1} is obtained from the image fine-tuning model $U(\sigma)$ transformed in the previous step. D_r is referred to as the relative difference:

$$D_r = DSC_t - DSC_{t-1}. \tag{10}$$

This relative difference considers the change in model accuracy between two adjacent transformations and is used to indicate the effect of the tth transformation on the model U_{t-1}. Thus, D_r contains information about the effectiveness of performing a single-step transformation on image x_i^{t-1}.

Finally, by considering both the absolute difference between the model accuracy after each augmentation and the initial accuracy, as well as the relative difference between two adjacent augmentations, we design a comprehensive reward function. The reward function is doubly constrained by the absolute and relative differences, enabling it to provide the corresponding reward or penalty to the agent based on the three cases of segmentation model performance change. The first case is that the difference D_{abs} between DSC_t and DSC_{init} is greater than 0, indicating that the augmentation sequences obtained by the agent are valid. If the difference D_r between DSC_t and DSC_{t-1} is also greater than 0, i.e., the tth transform is also valid for image x_i^{t-1}, a reward D_r will be given to the agent as positive feedback. This encourages the agent to choose the most valid augmentation transformation for the image and ensures the validity of the augmentation sequence. If D_r is less than 0, then although the augmentation sequence is still valid, the tth transformation is not conducive to improving the segmentation model performance. In this case, a penalty D_r is given to the agent as negative feedback. The second case is $D_{abs} \leq 0$ and $D_r > 0$. In this case, although the t-th transformation is valid, the augmentation sequence is invalid. If D_r is fed back to the agent directly as in other methods, the agent will receive positive feedback when the augmentation sequence

is invalid, which is not good for guiding the agent to find the best augmentation sequence. In this case, we provide negative feedback to penalize the agent. However, since the t-th transformation is valid, we feed the agent a variable penalty value $-e^{-D_r}$ based on the effect of the t-th transformation on the performance improvement of the segmentation model. When the performance improvement is large, the penalty value is small; conversely, the penalty value is large. This allows for a balanced consideration of the augmentation sequence as well as the effectiveness of each transformation. The third case is $D_{abs} \leq 0$ and $D_r \leq 0$. In this case, both the augmentation sequence and the tth transformation are ineffective in improving the performance of the segmentation model. Therefore, a larger negative feedback $-m$ is given to the agent as a penalty. Thus, the reward function is defined as:

$$R_{dc} = \begin{cases} D_r, & if\ D_{abs} > 0 \\ -e^{-D_r}, & if\ D_{abs} \leq 0\ \&\ D_r > 0 \\ -m, & if\ D_{abs} \leq 0\ \&\ D_r \leq 0. \end{cases} \tag{11}$$

Inspired by [48], to prevent the agent from experiencing difficulty in convergence during training because the variance of the immediate rewards is too large, many works [49] have used the reward clipping technique in training. Like these works, we also use the reward clipping technique to ensure stable learning for the agent. Specifically, we restrict the immediate reward r^t to be between $[-1, 1]$. That is, we cut all positive rewards greater than 1 at 1 and all negative rewards less than -1 at -1 while keeping other rewards unchanged.

Since we use the reward clipping technique to limit the immediate reward r^t to between $[-1, 1]$, we set m = 1 to penalize the case where $D_{abs} \leq 0$ and $D_r \leq 0$. The augmentation in this case is the worst because neither the augmentation sequence nor the tth transformation is effective in improving the performance of the segmentation model. Therefore the penalty value is maximum at this point.

The advantages of our reward function are as follows: (i) More accurate assessment of the validity of each transformation: We use D_{abs} and D_r to dually constrain the reward function and break down the specifics of the segmentation performance variation. Thus, we more specifically propose appropriate reward functions for different cases to accurately reflect the impact of each transformation on the segmentation model. (ii) Guarantee the validity of the augmentation sequence: Due to the dual constraints of the reward function by D_{abs} and D_r, the resulting reward signal is correlated with the validity of both the augmentation sequence and each transformation. The goal of the RL system is to maximize the cumulative reward, so the reward signal with dual constraints can guide the agent in choosing the best augmentation transformation while ensuring the validity of the augmentation sequence. In this way, we can provide more accurate feedback to the agent based on the variation of the segmentation model performance and alleviate the problem that some of the augmented data are invalid.

3.3. Automatic stopping mechanism

In ASDRL-AutoAug, we propose an automatic stopping mechanism, ASM, to remedy the problem of insufficient partial image augmentation. Specifically, when the transformed image continuously degrades the performance of the segmentation model, i.e., if the image has a decreasing trend in its augmentation trajectory, then the subsequent transformations are unnecessary. At this point, the image has been augmented sufficiently, so the agent automatically stops the process of augmenting.

Because the performance change of the segmentation model can intuitively be represented by our proposed R_{dc}, we use the reward value of each step to judge the necessity of each transformation and the adequacy of image augmentation, i.e., whether it is necessary to continue or stop the augmentation at this time. This allows the agent to automatically terminate the sequential transformation according to the image augmentation trajectory. Specifically, if $r^t > 0$, then the transformed image effectively improves the performance of the segmentation model, indicating that the current transformation is necessary and the augmentation of the image is insufficient. On the other hand, if $r^t \leq 0$, the transformed image does not contribute to the performance improvement of the segmentation model, signifying that the current transformation is non-essential. Therefore, when $r^t \leq 0$ occurs n times consecutively, the augmentation transformation of image x_i continues to have a negative impact on the segmentation model. In this case, the augmentation of x_i is deemed unnecessary. Consequently, we conclude its augmentation process and reset the environment to $U(\sigma)$. We use r^t as a judgment condition to formalize a stop signal, denoted as $Done$:

$$Done^t = \begin{cases} 1, & if\ r^t \leq 0\ continuous\ n\ times \\ 0, & otherwise. \end{cases} \tag{12}$$

When the environment provides a reward to the agent, it concurrently feeds a stop signal, prompting the agent to automatically stop the transformation of the current image when the augmentation necessity becomes insufficient.

The automatic stopping mechanism we propose offers two benefits. First, this intelligent approach ensures that each transformation before stopping the augmentation results in a smooth improvement in model performance. This is because once the transformed image begins to degrade the performance of the segmentation model, the agent promptly ends the augmentation process for that image. Second, ASM enables the construction of an approximately infinite search space, thereby improving the adequacy of image augmentation. This is attributed to the search space size, which can be represented as 12^t, where 12 denotes the number of actions. Given that our augmentation process lacks a restriction on the maximum number of times, in theory, as t tends to $+\infty$, the search space becomes approximately infinite. This allows the agent not to consider the maximum number of transformations and to explore the augmentation sequences that make the segmentation model perform sufficiently well in an approximately infinite search space.

4. Experiments

Extensive experiments have been conducted to evaluate our proposed Adaptive Sequence-length based Deep Reinforcement Learning (ASDRL) model for Automatic data Augmentation (AutoAug) in intelligent medical image analysis (ASDRL-AutoAug). In this section, we first present information about the datasets, experimental setup, evaluation metrics, and baselines in Sections Section 4.1 to 4.4. Second, extensive experiments are conducted to verify that the segmentation model utilizes our proposed ASDRL-AutoAug to improve the segmentation performance more effectively than utilizing state-of-the-art data augmentation methods. Then, ablation studies are conducted to further investigate the effectiveness and necessity of the proposed R_{dc} and ASM in ASDRL-AutoAug and to further demonstrate the effectiveness of the different feedbacks in the proposed reward R_{dc}. Finally, additional experiments are conducted aiming to demonstrate the adaptiveness and universality of ASDRL-AutoAug and the effect of hyperparameter n in ASM on the training time cost and the effectiveness of the augmented data.

4.1. Datasets

Empirical studies of three public datasets confirm that our method outperforms other baselines. As shown in Table 2, we evaluate our method using three medical image datasets, which contain features with small datasets, small objects, and complex segmentation details (e.g., segmented edges) that are more representative of current medical images.

Cardiac is a publicly available MRI dataset [50] for automatic segmentation of the heart. It has a small number of data and small segmented objects, making it more challenging for segmentation tasks. The dataset

Table 2
Datasets.

Datasets	Images	Size	Modality	Challenge	Source
Cardiac [50]	1,350	320 × 320	MRI	Small dataset with small segmented objects	King's College London
BUSI [51]	647	256 × 256	USI	Tiny dataset with large variability	Baheya Hospital of Cairo, Egypt
LiTs [50]	1588	512 × 512	CT	Small dataset with large variability	Several clinical sites

is provided by King's College London and contains 20 cases with a total of 1350 slices, each with a size of 320 × 320 pixels.

BUSI is a publicly available dataset of breast ultrasound images [51] for the automatic segmentation of breast cancer tumors. The difficulty of its segmentation is the smaller number of data and the larger anatomical variability. The dataset is provided by Baheya Hospital for Early Detection & Treatment of Women's Cancer, Cairo, Egypt. The images are divided into three categories, including normal (no tumor), benign tumor, and malignant tumor, with a total of 600 cases and 780 slices, each with a size of 256 × 256 pixels. We selected only the 647 slices containing tumors in the segmentation task.

LiTs [50] is a publicly available CT dataset for the automatic segmentation of livers and tumors. It has large anatomical variability. The dataset contains 131 cases, each containing approximately 512 × 512 CT slices, with the number of slices ranging from 29 to 299. Because the number of slices in both the Cardiac and BUSI datasets is around 1000 and the segmentation task is singular in both cases, for a fair comparison and analysis, we randomly selected 1588 slices from LiTs and focused solely on liver segmentation.

To train and evaluate the network, the following preprocessing is performed on the three medical image datasets mentioned above. First, for Cardiac, we transform these 3D images to 2D images according to the transverse section in a slice-by-slice manner [52]. Then, for all three datasets, we normalize all input images to have zero mean and unit std. Finally, for all datasets, 70% of the datasets are used for training, 10% for validation, and 20% for testing.

4.2. Implementation details

Our experiments are implemented based on the PyTorch framework. To evaluate the performance of ASDRL-AutoAug, we choose a randomly initialized U-Net as our original baseline. In our experiments, several state-of-the-art data augmentation methods applied to the field of medical image segmentation are selected as data augmentation baselines, and then their impact on downstream performance is evaluated by the same U-Net. For a fair comparison, we repeat each experiment five times and report the optimal results. In addition, we fix a random seed to ensure that the same training result and evaluation result can be obtained under the same hyperparameters.

For U-Net, we use the Adam [53] optimizer with an initial learning rate set to 2.5×10^{-4}, weight decay to 5×10^{-4}, and batch size of 8. We train on two NVIDIA GeForce GTX 2080 GPUs. For Dueling DQN, the Adam optimizer is used, and the learning rate is set to 0.001. The discount factor γ is set to 0.1, and the feedback value m is set to 1. The hyper-parameter n is set to 2 during training and to 1 during validation. For the ϵ-*decreasing* algorithm, the initial value of ϵ is set to 1. The decreasing value is 3.3×10^{-5} in each iteration with a minimum value of 0.01. We perform the training of Dueling DQN on four NVIDIA GeForce GTX 2080 GPUs. The learning rate is set to 5×10^{-6} when fine-tuning U-Net. Our source code has been released on https://github.com/zhxhebut/ASDRL-AutoAug.

4.3. Evaluation metrics

To show the effectiveness of our method, we use six widely used segmentation evaluation metrics including mean Intersection over Union (mIoU), Dice Similarity Coefficient (DSC), Positive Predict Value (PPV), Sensitivity (SEN), 95% Hausdorff distance (HD95) [54] and Average Surface Distance (ASD). Specifically, mIoU is the ratio of the intersection and concatenation of the ground truths and the predicted values, which can measure the model's ability to identify training samples and boundary inference, and effectively compare the performance of each model. DSC, also known as F1-measure or F1-score, is the summed average of PPV and SEN, which can be used to evaluate the overlap between the outputs and the ground truths comprehensively. PPV, also known as precision, is the proportion of true positive cases out of all samples predicted to be positive. SEN, also known as recall, is the proportion of the positive samples that are correctly predicted to be positive. HD95 represents the 95% quantified value of the maximum surface distance between the prediction and the ground truths, which can reduce the influence of outliers and make the evaluation more stable and reliable. ASD is the average distance between the set of prediction points and the set of ground truth points. It is worth noting that higher values for these metrics, except HD95 and ASD, mean better performance. Formally,

$$mIoU = \frac{1}{K+1}\sum_{i=0}^{k}\frac{TP}{TP+FP+FN}, \tag{13}$$

$$DSC = \frac{2\times PPV\times SEN}{PPV+SEN} = \frac{2TP}{2TP+FP+FN}, \tag{14}$$

$$PPV = \frac{TP}{TP+FP}, \tag{15}$$

$$SEN = \frac{TP}{TP+FN}, \tag{16}$$

$$HD95 = max_{k95\%}[d(P,G), d(G,P)], \tag{17}$$

$$ASD = \frac{d(P,G)}{|P|}, \tag{18}$$

where TP, FP, and FN are the number of true positive points, false positive points, and false negative points, respectively. T is the number of ground-truth points of that class, P is the number of predicted positive points, G is the number of ground-truth positive points, and $d(*)$ is a function to calculate the surface distance. Consequently, DSC, mIoU, PPV, and SEN range from 0 to 1; HD95 and ASD range from 0 to $+\infty$; and none of these metrics have specific units.

4.4. Baselines

In the downstream segmentation task, we perform supervised training on the randomly initialized U-Net [44] with the mixed training set $\mathbb{M}$ of initial training set $\mathbb{N}$ and augmented training set $\mathbb{G}$ and then test the segmentation accuracy on the test set $\mathbb{T} = \{(p_1, q_1), \ldots, (p_k, q_k)\}$.

To evaluate the performance of the proposed ASDRL-AutoAug, thirteen state-of-the-art data augmentation methods 3-Augment [13], Random Aug [7], Cutout [27], Cutmix [34], Mixup [33], saliencyMix [17], Random Erasing [15] , GridMask [29], SLSRL [41], DADA [38], KeepAugment [16], ADA-DRL [25], and RandAugment [20] were selected as the baselines. The reasons for selecting these thirteen methods as baselines are as follows. (i) 3-Augment and Random Aug are random data augmentation methods capable of randomly selecting a transform among the given image transformations with equal probability. In

Table 3
The results of applying the proposed ASDRL-AutoAug and the state-of-the-art baselines on three public datasets, where the best and the second-best results are bold and underlined, respectively.

Datasets	Methods	DSC ↑	mIoU ↑	PPV ↑	SEN ↑	HD95 ↓	ASD ↓
Cardiac	U-Net [44]	0.7566	0.6605	0.7682	0.8011	8.0114	2.5489
	3-Augment [44]	0.7612	0.6545	0.7416	0.8264	10.0534	3.6543
	Random Aug [7]	0.7637	0.6570	0.7506	0.8398	7.6250	2.0446
	Cutout [27]	0.7663	0.6630	0.7426	0.8403	7.2624	2.2894
	Mixup [33]	0.7698	0.6749	0.7780	0.8123	11.9769	4.5322
	Cutmix [34]	0.7716	0.6592	0.7651	0.8286	8.4839	2.2331
	SaliencyMix [17]	0.7828	0.6895	0.7908	0.8137	6.5083	1.9798
	Random Erasing [15]	0.7881	0.6955	0.8119	0.7929	6.5787	2.6248
	GridMask [29]	0.7919	0.6993	0.8048	0.8151	6.9764	2.2509
	SLSRL [41]	0.7927	0.6951	0.7874	0.8368	5.7985	1.5747
	DADA [38]	0.7948	0.6947	0.7880	0.8446	9.1693	2.7868
	KeepAugment [16]	0.7953	0.6971	0.7995	0.8287	5.9727	1.9316
	ADA-DRL [25]	0.7963	0.7023	0.7889	0.8449	7.3819	2.2527
	RandAugment [20]	0.8002	0.7031	0.8053	0.8359	5.7941	1.8948
	Ours	**0.8122**	**0.7126**	**0.8140**	**0.8495**	**4.8361**	**1.2844**
BUSI	U-Net [44]	0.7121	0.6059	0.7487	0.7882	28.4540	10.0333
	3-Augment [44]	0.7205	0.6077	0.7204	0.7989	38.2277	14.8325
	Random Aug [7]	0.7313	0.6260	0.7333	0.8226	30.7283	11.1877
	Cutout [27]	0.7357	0.6315	0.7480	0.8101	27.1448	10.0033
	Mixup [33]	0.7413	0.6438	0.7824	0.7908	26.2881	9.7748
	Cutmix [34]	0.7469	0.6410	0.7343	0.8277	29.6330	11.4890
	SaliencyMix [17]	0.7521	0.6476	0.7590	0.8198	27.1902	9.2420
	Random Erasing [15]	0.7528	0.6513	0.7633	0.8263	27.7175	11.2242
	GridMask [29]	0.7638	0.6640	0.7822	0.8250	28.1392	9.8012
	KeepAugment [16]	0.7649	0.6641	0.7909	0.8111	24.4839	9.5988
	SLSRL [41]	0.7661	0.6691	0.7818	0.8247	24.4643	8.9452
	DADA [38]	0.7681	0.6653	0.7652	0.8329	32.2603	10.6599
	ADA-DRL [25]	0.7760	0.6749	0.7851	**0.8397**	23.6102	8.6173
	RandAugment [20]	0.7767	0.6767	0.7993	0.8251	22.6456	8.5182
	Ours	**0.7936**	**0.6943**	**0.8102**	0.8350	**22.6225**	**7.8293**
LiTs	U-Net [44]	0.7868	0.6859	0.8879	0.7493	23.5688	4.6414
	3-Augment [44]	0.8067	0.7323	0.8857	0.7741	17.2253	4.6966
	Random Aug [7]	0.8081	0.7204	0.8583	0.7867	20.6727	5.2243
	Cutout [27]	0.8093	0.7139	0.9014	0.7575	21.1350	3.7483
	Mixup [33]	0.8172	0.7251	0.8939	0.7750	16.1688	3.5270
	Cutmix [34]	0.8229	0.7386	0.8914	0.7826	15.4874	5.5498
	SaliencyMix [17]	0.8276	0.7352	0.9107	0.7958	19.9658	3.4425
	Random Erasing [15]	0.8327	0.7418	0.9103	0.7885	18.5983	3.9739
	GridMask [29]	0.8417	0.7583	0.9011	0.8031	14.9505	3.9117
	SLSRL [41]	0.8504	0.7715	0.8747	0.8487	16.7822	4.9585
	KeepAugment [16]	0.8585	0.7876	0.9087	0.8503	13.0604	2.8398
	DADA [38]	0.8655	0.7926	0.9167	0.8313	11.4246	3.4568
	ADA-DRL [25]	0.8748	0.8075	0.9031	0.8697	10.8029	2.6642
	RandAugment [20]	0.8777	0.8126	0.9229	0.8548	7.5151	2.2261
	Ours	**0.8951**	**0.8362**	**0.9248**	**0.8866**	**6.2526**	**2.1777**

particular, 3-Augment selects a transform from the given three transformations, while Random Aug selects a transform from the 12 transformations in Table 1; (ii) Cutout generates richer images by region erasing. Random Erasing and GridMask introduce random and multi-scale grid masking for region erasing, respectively, and the image erasing process is more complicated; (iii) KeepAugment is a single-image mixing technique based on saliency features, Mixup mixes two images and corresponding labels based on the mixing factor (alpha), Cutmix and SaliencyMix introduce randomness and saliency for multi-image mixing, respectively. These methods can force the model to better learn the representational information of the images; and (iv) RandAugment is an automatic data augmentation based on grid search; SLSRL and ADA-DRL are automatic data augmentation based on DRL for medical images; and DADA is a differentiable automatic data augmentation. All four of these methods can find suitable combinations among given augmentation transformations.

4.5. Main results

To demonstrate the effectiveness of our proposed ASDRL-AutoAug, we conducted experiments on three public datasets and compared ASDRL-AutoAug with thirteen state-of-the-art baselines for medical image data augmentation. The quantitative experimental results are shown in Table 3, and the visualization of the effects of different data augmentation methods on the performance of the segmentation model is shown in Fig. 3. Through quantitative and qualitative analysis, we draw the following conclusions.

As shown in Table 3, first, SaliencyMix, Random Erasing, GridMask, and KeepAugment outperform 3-Augment and Random Aug for three datasets with respect to DSC, mIoU, and PPV. This is because image erasing and image mixing, two advanced image augmentation transformations can force the model to better learn the representational information of the images by generating richer images. This demonstrates that increasing the diversity of training samples can improve the performance of deep models. Second, we observe that SaliencyMix, Random Erasing, GridMask, SLSRL, DADA, and KeepAugment generally outperform Cutout, mixup, and Cutmix on three datasets. This is because Random Erasing and GridMask generally mask more accurately by randomly determining the mask size and applying multi-scale grid masking; SaliencyMix and KeepAugment use saliency information to increase or retain salient regions more accurately; SLSRL and DADA automatically select more appropriate augmentation probabilities or operations for the images. This proves that more accurate transformations for different images are better than blindly selected

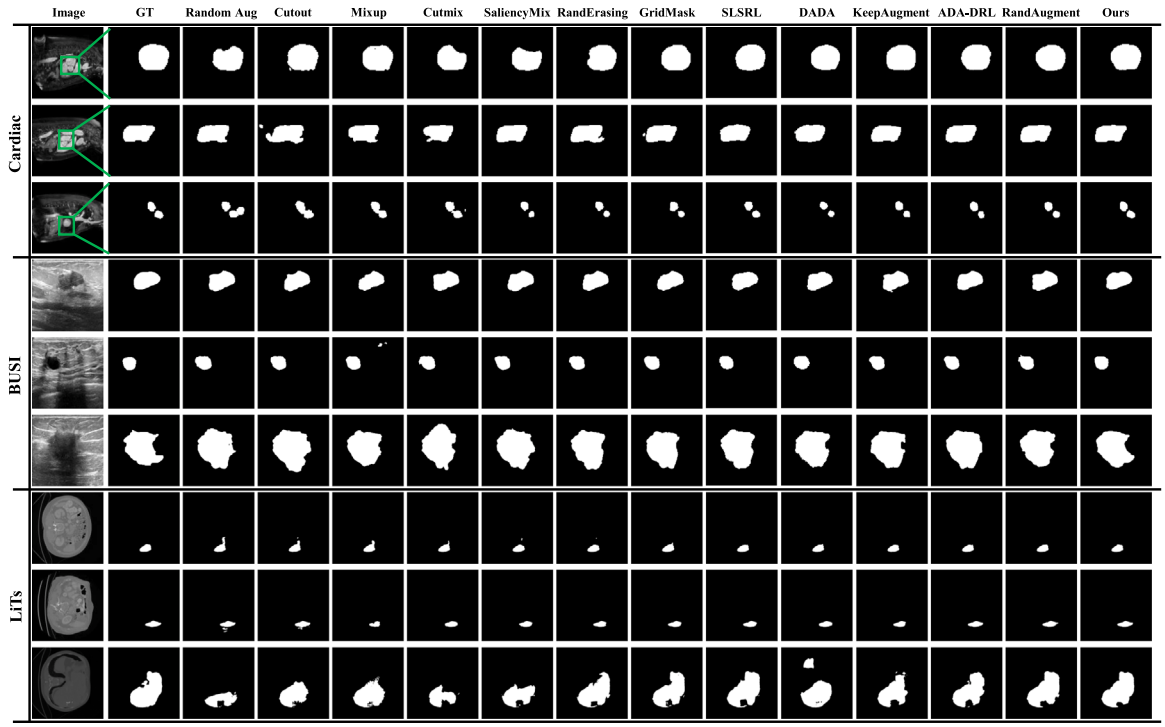

Fig. 3. Examples of visualized segmentation results of our proposed ASDRL-AutoAug and the baselines on three public datasets.

transformations. Then, we find that ADA-DRL outperforms SaliencyMix, Random Erasing, GridMask, SLSRL, DADA, and KeepAugment on DSC, mIoU, and SEN. This is because ADA-DRL automatically selects the sequence of basic transformations for each image by training with the task network interaction, such that the selection of augmentation transformations is more accurate than random or saliency-based methods. This further demonstrates that more accurate transformations for different images can better improve the performance of the deep models. In addition, we observe that RandAugment generally outperforms the previous data augmentation methods on three datasets. This is because RandAugment automatically finds the number and magnitude of data augmentation transformations by grid search. This demonstrates that the suitable number of transformations for different images is different and that setting the maximum length of the augmentation sequence based on human a priori leads to the insufficient augmentation of some images.

Finally, we find that our proposed ASDRL-AutoAug achieves more significant improvements over all baselines on the Cardiac, BUSI, and LiTs datasets, which proves that our proposed ASDRL-AutoAug is more effective than the current data augmentation methods. Specifically, we first find that ASDRL-AutoAug is 1.2%, 0.95%, 0.21%, and 0.46% higher than the second-best results on the Cardiac dataset for DSC, mIoU, PPV, and SEN, respectively, and 0.9580 and 0.6104 lower for HD95 and ASD, respectively; while on the BUSI dataset, DSC, mIoU, and PPV are 1.69%, 1.76%, and 1.09% higher, and HD95 and ASD are 0.0231 and 0.6889 lower, respectively than the second best result. SEN is the second-best result and very close to the best result; On the LiTs dataset, ASDRL-AutoAug outperforms the second-best result by 1.74%, 2.36%, 0.19%, and 1.69% for DSC, mIoU, PPV, and SEN, respectively, and 1.2625 and 0.0484 for HD95 and ASD, respectively. This demonstrates that ASDRL-AutoAug achieves a better performance than state-of-the-art data augmentation methods in medical image segmentation tasks. In particular, we observe that ASDRL-AutoAug improves more on the BUSI dataset than on Cardiac datasets (e.g., DSC, and mIoU). This observation suggests that ASDRL-AutoAug is still effective and even more significantly improved on datasets with smaller data volumes. The reasons for the superior performance of ASDRL-AutoAug are as follows. (i) ASDRL-AutoAug can automatically increase the diversity of training images by selecting the data augmentation sequence that is more conducive to improving model accuracy for each image through interactive training; (ii) ASDRL-AutoAug uses the reward R_{dc} with dual constraints to improve the accuracy of the evaluation for transformations, and can solve the problem that some of the augmented images are invalid by accurately selecting the augmentation transformations; and (iii) ASDRL-AutoAug uses the automatic stopping mechanism ASM to select different numbers of transformations for different images adaptively, which can solve the problem that some of the images are not augmented adequately.

Fig. 3 shows the visualized segmentation results of ASDRL-AutoAug and the baselines on nine examples from three datasets. Specifically, the cardiac images (at the first three rows) show that: (i) the segmentation results of Random Aug, Cutout, Mixup, Cutmix, SaliencyMix, RandErasing, and GridMask are very incorrect and sometimes even under-segmented (e.g., Random Aug, Mixup, Cutmix, and SaliencyMix in the first row) or over-segmented (e.g., Cutout, RandErasing, and GridMask in the second row, and Random Aug, Cutout, Mixup, and Cutmix in the third row); (ii) KeepAugment, SLSRL, DADA, ADA-DRL, and RandAugment are better, but their performance is not satisfactory for the edge regions of the organs; (iii) the segmentation performance of our proposed ASDRL-AutoAug is much better than the baselines, and its segmentation results retain more details in the foreground regions and are very close to the ground truths. Similarly, we have the following observations for the breast images. (i) Random Aug, Cutout, Mixup, Cutmix, SaliencyMix, RandErasing, and GridMask cannot segment the tumor correctly, and under-segmentation (e.g., the fourth row) and over-segmentation (e.g., the sixth row) still exist; (ii) KeepAugment, SLSRL, DADA, ADA-DRL, and RandAugment do not segment the tumor edge region satisfactorily; (iii) among all data augmentation methods, the proposed ASDRL-AutoAug has the best segmentation results. Finally, the liver images (at the last three rows) show that: (i) the segmentation results of Random Aug, Cutout, Mixup, Cutmix, SaliencyMix, RandErasing, and GridMask suffer from under-segmentation and over-segmentation; (ii) the segmentation results of KeepAugment, SLSRL, DADA, ADA-DRL, and RandAugment have unsatisfactory performance in terms of the edge region of the organ; (iii) the segmentation

Table 4
Results of ablation experiments using different reward functions and using different stopping mechanisms, where the best and second best results are bold and underlined, respectively.

Datasets	Methods				DSC ↑	mIoU ↑	PPV ↑	SEN ↑	HD95 ↓	ASD ↓
	R_{abs}	R_r	R_{dc}	ASM						
Cardiac	✓	–	–	–	0.7881	0.6955	0.8119	0.7929	6.5787	2.6248
	–	✓	–	–	0.7963	0.7023	0.7889	0.8449	7.3819	2.2527
	–	–	✓	–	0.8023	0.7051	0.8011	0.8386	6.5036	1.5219
	✓	–	–	✓	0.7931	0.6926	0.7853	0.8429	6.4490	1.7082
	–	✓	–	✓	0.7996	0.7071	0.8060	0.8411	5.4744	1.5426
	–	–	✓	✓	**0.8122**	**0.7126**	**0.8140**	**0.8495**	**4.8361**	**1.2844**
BUSI	✓	–	–	–	0.7726	0.6684	0.7884	0.8289	32.3332	10.0995
	–	✓	–	–	0.7760	0.6749	0.7851	**0.8397**	23.6102	8.6173
	–	–	✓	–	0.7887	0.6908	0.8085	0.8322	23.6218	8.7044
	✓	–	–	✓	0.7755	0.6743	0.8072	0.8177	30.3277	10.0257
	–	✓	–	✓	0.7806	0.6827	0.7892	0.8202	28.2153	**7.5418**
	–	–	✓	✓	**0.7936**	**0.6943**	**0.8102**	0.8350	**22.6225**	7.8293
LiTs	✓	–	–	–	0.8670	0.7965	0.8920	0.8727	10.7099	3.2821
	–	✓	–	–	0.8748	0.8075	0.9031	0.8697	10.8029	2.6642
	–	–	✓	–	0.8869	0.8244	0.9085	0.8826	10.8770	3.6674
	✓	–	–	✓	0.8730	0.8009	0.9139	0.8502	10.7815	3.1534
	–	✓	–	✓	0.8840	0.8214	0.8990	**0.8914**	13.4559	4.0415
	–	–	✓	✓	0.8951	**0.8362**	**0.9248**	0.8866	**6.2526**	**2.1777**

performance of our proposed ASDRL-AutoAug is much better than the baselines. Thus, these visualization examples once again greatly demonstrate that ASDRL-AutoAug compensates for the shortcomings of existing data augmentation methods through the proposed reward R_{dc} and ASM, and achieves a better performance in medical image segmentation tasks.

4.6. Ablation studies

To demonstrate the effectiveness and necessity of each improvement proposed in ASDRL-AutoAug, we conduct ablation studies, and the experimental results are shown in Table 4. In addition, to further demonstrate the effectiveness of the different feedbacks in the proposed reward R_{dc}, we conduct ablation studies on three datasets, and the corresponding results are shown in Table 5.

4.6.1. Effectiveness and necessity of R_{dc} and ASM

To further investigate the effectiveness and necessity of our proposed reward R_{dc} and automatic stopping mechanism ASM, we conduct ablation studies on three datasets, and the results are shown in Table 4, where $R_{abs} = D_{abs}$ and $R_r = D_r$ represent the reward function that feeds back only D_{abs} or D_r in all cases. When the ASM is not used, the stopping mode is the same as ADA-DRL.

As shown in Table 4, we can find that, first, the automatic data augmentation method combining R_{dc} and ASM outperforms the other experimental results in the segmentation task on three datasets. Specifically, on the Cardiac dataset, the sixth row outperforms the second-best results by 0.99%, 0.55%, 0.8%, and 0.46% for DSC, mIoU, PPV, and SEN, respectively, and by 0.6383 and 0.2375 for HD95 and ASD, respectively; on the BUSI dataset, DSC, mIoU, and PPV are higher than the second-best results by 0.49%, 0.35%, and 0.17%, respectively, HD95 is 0.9877 lower than the second best result, and SEN, ASD are the second-best result and very close to the best result; on the LiTs dataset, DSC, mIoU, and PPV were 0.82%, 1.18%, and 1.09% higher than the second-best result, and HD95 and ASD were 4.4573 and 0.4865 lower, respectively, with SEN being the second-best result and very close to the best result. This demonstrates that the automatic data augmentation method combining R_{dc} and ASM can significantly improve the segmentation performance. This is because R_{dc} and ASM are targeted to solve different problems and can complement each other to better improve the effectiveness of data augmentation. Second, comparing the third row with the first and second rows (resp., the sixth row with the fourth and fifth rows), i.e., using R_{dc} and other rewards in the same stopping conditions, we find that R_{dc} performs better. For example, on three datasets, DSC and mIoU in the third row (resp., sixth row) are larger than in the first and second rows (resp., the fourth and fifth rows). This is because the accurate feedback by R_{dc} for the specific case of the augmentation trajectory can improve the accuracy of the assessment of the validity of each augmentation transformation. It is thus demonstrated that R_{dc} achieves an accurate assessment of each augmentation transformation by dual constraints, which can guarantee the validity of the augmentation sequence, i.e., it can better improve the performance of the segmentation model. Finally, comparing the first and fourth rows, the second and fifth rows, and the third and sixth rows, respectively, we find that the DSC of the latter is greater than that of the former on three datasets. This observation suggests that data augmentation using ASM under the same reward conditions is more effective than data augmentation with the maximum number of transformations. This is because ASM solves the problem of limited sequence length in other automatic data augmentation algorithms, allowing the agent to find the augmentation transformations for each image in turn without considering the maximum number of transformations, thus ensuring that the augmentation sequence leads to a sufficient improvement in the segmentation model performance. Thus, this proves the effectiveness of ASM in automatic data augmentation algorithms.

The above observations show that both the proposed R_{dc} and ASM are effective and essential for ASDRL-AutoAug to achieve a superior medical image data augmentation.

4.6.2. Effectiveness of different feedbacks in R_{dc}

To further demonstrate that the proposed reward R_{dc} is accurate for the evaluation of the augmentation transformation in different cases, we conduct ablation experiments on the feedback functions in different cases, and the corresponding results are shown in Table 5, where $R_a = D_{abs}$, $R_b = -e^{-D_{abs}}$, $R_c = -m$, $R_d = D_r$, and $R_e = -e^{-D_r}$. In Section 4.6.1, we have proved the validity of ASM, so in this section, we only perform experiments under the condition of using ASM and do not verify the validity of ASM anymore.

In Table 5, first comparing the first and fourth rows of each dataset separately, we find that feeding R_d in all three cases is more effective than feeding R_a in all three cases. This is because although R_a is beneficial to ensure the validity of the augmentation sequence, the augmentation sequence is composed of each transformation in sequence. So, R_d ensures the validity of each transformation first before the validity of the sequence can be guaranteed with the aid of R_a. Second, comparing the first and second rows (resp., fourth and fifth rows), we find that when changing only the feedback function in the case of D_{abs}

Table 5
Ablation studies of the feedback functions under different conditions when using ASM, where $R_a = D_{abs}$, $R_b = -e^{-D_{abs}}$, $R_c = -m$, $R_d = D_r$, and $R_e = -e^{-D_r}$. The best and the second-best results are bold and underlined, respectively.

Datasets	Rewards			DSC ↑	mIoU ↑	PPV ↑	SEN ↑	HD95 ↓	ASD ↓
	$D_{abs} > 0$	$D_{abs} \le 0$ $D_r > 0$	$D_{abs} \le 0$ $D_r \le 0$						
Cardiac	R_a	R_a	R_a	0.7931	0.6926	0.7853	0.8429	6.4490	1.7082
	R_a	R_a	R_c	0.7984	0.7010	0.7817	**0.8588**	10.5937	2.7425
	R_a	R_b	R_c	0.7994	0.6977	0.8037	0.8326	7.9681	1.9105
	R_d	R_d	R_d	0.7996	0.7071	0.8060	0.8411	5.4744	1.5426
	R_d	R_d	R_c	0.8019	0.7061	**0.8197**	0.8290	6.2534	1.4012
	R_d	R_e	R_c	**0.8122**	**0.7126**	0.8140	0.8495	**4.8361**	**1.2844**
BUSI	R_a	R_a	R_a	0.7755	0.6743	0.8072	0.8177	30.3277	10.0257
	R_a	R_a	R_c	0.7770	0.6784	0.7684	0.8637	23.3161	9.2675
	R_a	R_b	R_c	0.7784	0.6767	0.7861	0.8431	23.8790	8.5102
	R_d	R_d	R_d	0.7806	0.6827	0.7892	0.8202	28.2153	**7.5418**
	R_d	R_d	R_c	0.7871	0.6875	0.7786	**0.8652**	30.8569	9.6859
	R_d	R_e	R_c	**0.7936**	**0.6943**	**0.8102**	0.8350	**22.6225**	7.8293
LiTs	R_a	R_a	R_a	0.8730	0.8009	0.9139	0.8502	10.7815	3.1534
	R_a	R_a	R_c	0.8747	0.8048	0.9057	0.8598	12.2362	3.7651
	R_a	R_b	R_c	0.8769	0.8068	0.8730	**0.9147**	14.0705	4.5612
	R_d	R_d	R_d	0.8840	0.8214	0.8990	0.8914	13.4559	4.0415
	R_d	R_d	R_c	0.8931	0.8311	0.9183	0.8913	7.2722	2.4703
	R_d	R_e	R_c	**0.8951**	**0.8362**	**0.9248**	0.8866	**6.2526**	**2.1777**

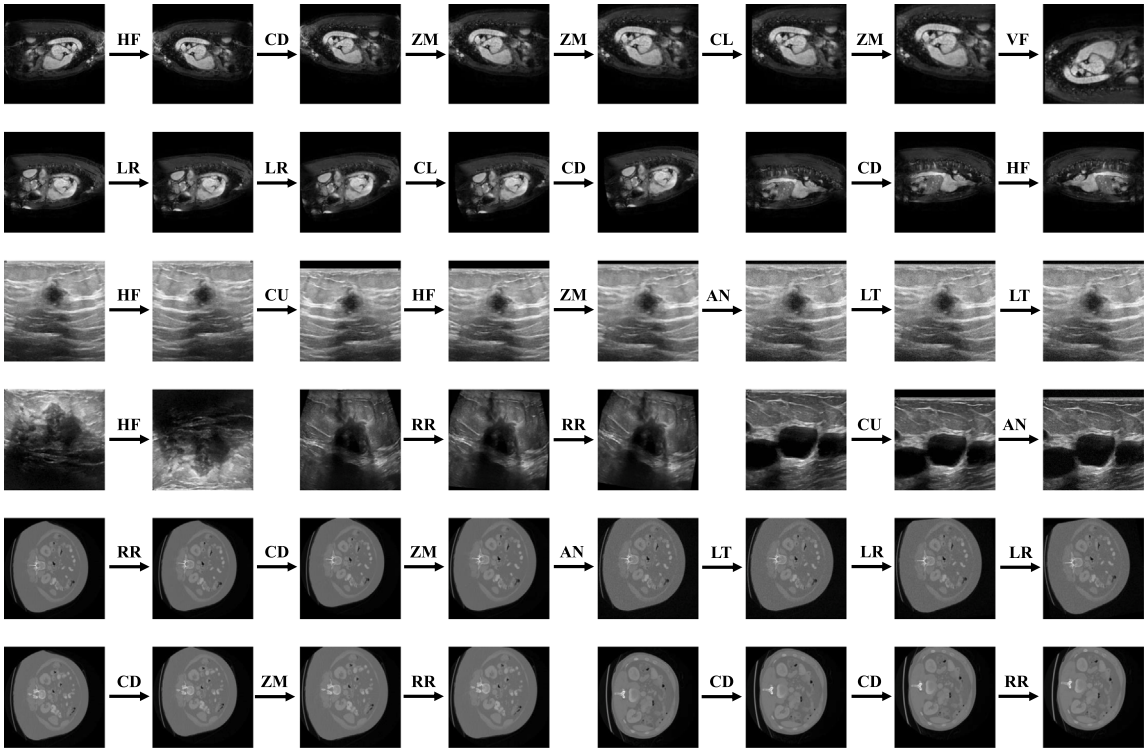

Fig. 4. Process of transforming images with the proposed ASDRL-AutoAug and the actions at each time step. The first two rows are three examples from the Cardiac dataset, the middle two rows are four examples from the BUSI dataset, and the last two rows are three examples from the LiTs dataset.

≤ 0 and $D_r \le 0$, the DSC of the latter improves on three datasets. This is because in this case, R_c has a higher penalty value than R_a (resp., R_c over R_d). This proves that the feedback R_c is more accurate in the case of $D_{abs} \le 0$ and $D_r \le 0$. Then, comparing the fifth and sixth rows, we find that when changing only the feedback function in the case of $D_{abs} \le 0$ and $D_r > 0$, the latter has a higher DSC and lower HD95 on Cardiac, and both DSC and PPV are higher on BUSI; on the LiTs dataset, the latter has higher DSC and mIoU. This is because R_e feeds the appropriate penalty value to the agent in this case of invalid augmentation sequences, but R_d is a positive reward value in this case. This proves that the feedback R_e is more accurate when $D_{abs} \le 0$ and $D_r > 0$. Finally, we find that the segmentation performance of the sixth row outperforms the other experimental results. Specifically, on Cardiac, the sixth row is 1.03% and 0.55% higher than the second-best result in terms of DSC and mIoU, respectively, and 0.6383 and 0.1168 lower in terms of HD95 and ASD, respectively, with PPV and SEN being the second-best results; on BUSI, DSC, mIoU, and PPV are higher than the second best results by 0.65%, 0.68%, and 0.3% respectively, HD95 is 0.6936 lower than the second-best result, and ASD is the second-best result and very close to the best result; on the LiTs dataset, DSC, mIoU, and PPV were 0.2%, 0.51%, and 0.65% higher, and HD95 and ASD were 1.0196 and 0.2926 lower, respectively, than the second-best results. This is because R_{dc} provides different feedback values in different cases of the augmentation trajectory, allowing a more accurate assessment of the effectiveness of the augmentation transformations. This proves

Table 6
Training time (in hours per epoch) and segmentation results of ASDRL-AutoAug under different settings of hyper-parameter n.

n	Cardiac			BUSI			LiTs		
	DSC ↑	HD95 ↓	time ↓	DSC ↑	HD95 ↓	time ↓	DSC ↑	HD95 ↓	time ↓
1	0.7973	6.2629	**0.8635**	0.7625	25.4043	**0.2386**	0.8725	11.6184	**1.7358**
2	0.8122	**4.8361**	2.1292	0.7936	**22.6225**	0.7206	**0.8951**	**6.2526**	4.0573
3	**0.8123**	6.5988	4.5653	**0.7940**	22.6475	1.0703	0.8814	12.2457	10.4841
4	0.8111	5.5189	9.2457	0.7925	23.2981	4.2367	／	／	／
5	／	／	／	／	／	／	／	／	／

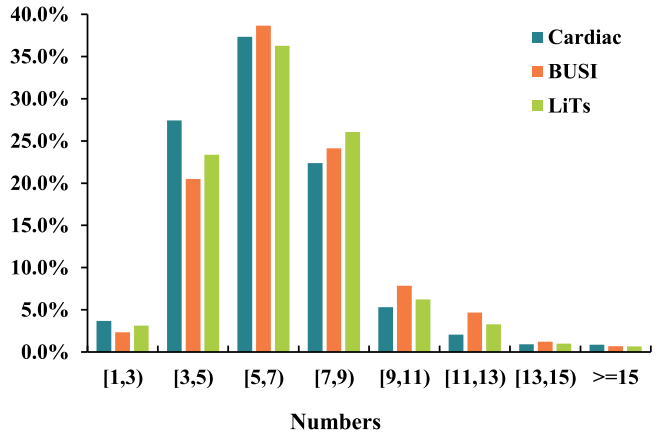

Fig. 5. The percentage of images in the dataset is calculated according to the number of transformations in the image augmentation sequence at the time of testing.

that the different feedback values provided by R_{dc} in different cases are valid.

4.7. Additional experiments

In this section, we aim to demonstrate the adaptiveness and universality of ASDRL-AutoAug and the effect of the hyperparameter n in ASM on the training time cost and the effectiveness of the augmented data.

4.7.1. Adaptiveness of ASDRL-AutoAug

To further demonstrate that the proposed ASDRL-AutoAug is adaptive, Fig. 4 visualizes the augmentation process of images in the three datasets of Cardiac, BUSI, and LiTs. We can observe that, first, the trained Dueling DQN tends to enlarge the segmented target regions or transform the images to different orientations during the augmentation process to generate well-diversified and effective augmentation images that are still recognizable after the transformation. This proves that the proposed ASDRL-AutoAug is effective. Second, the number of transformations used in the augmentation process is different for different images. In general, Dueling DQN automatically stops the augmentation of this image when the augmented image highlights the segmentation target. This proves that the proposed ASDRL-AutoAug is adaptive to the images.

As shown in Fig. 5, we count the number of transformations of each image in the final augmentation sequence during the testing phase of Dueling DQN and calculate the percentage of images with different numbers of transformations in the training set. We can observe that, first, for the three datasets Cardiac, BUSI, and LiTs, although the images using 5 to 6 transformations are the most numerous, the images using 7 and more transformations still have a larger percentage. This suggests that the existing practice of setting a maximum value (e.g., 2 or 6) for the number of transformations for all images is not reasonable. In addition, there is a wide range of the number of transformations used for the images of the three datasets. Although there are more images with the number of transformations concentrated between 1 and 13, there are still images using more than 13 transformations, which indicates that the number of transformations of images is diverse when using ASDRL-AutoAug. This proves that the sequence lengths of different images are adaptive when using ASDRL-AutoAug to augment images.

4.7.2. Universality of ASDRL-AutoAug

To further demonstrate the universality of the proposed ASDRL-AutoAug, we apply ASDRL-AutoAug to five existing commonly used segmentation models: U-Net, DANet [55], ResUNet++ [56], U-Net++ [57], and U-Net3+ [58]. We compare the models trained without data augmentation, and models using four advanced data augmentation methods, SaliencyMix, GridMask, ADA-DRL, and RandAugment, using two evaluation metrics, DSC and HD95, on three datasets. The corresponding experimental results are shown in Fig. 6.

As shown in Fig. 6, the proposed ASDRL-AutoAug can be used for any segmentation model and can significantly improve the segmentation performance. Specifically, first, we observe that on the Cardiac, BUSI, and LiTs datasets, our proposed ASDRL-AutoAug improves the DSC of all segmentation models and significantly reduces the HD95 of all segmentation models. In particular, the worse the performance of the segmentation models, the more ASDRL-AutoAug improves the DSC. This demonstrates that the proposed ASDRL-AutoAug can be used for any segmentation model. In addition, ASDRL-AutoAug outperforms other data augmentation methods in both segmentation metrics on the three datasets, has the highest DSC improvement for all segmentation models, and reduces HD95 to a minimum. This demonstrates that the proposed ASDRL-AutoAug can significantly improve segmentation performance.

4.7.3. Effect of varying the hyper-parameter n

As shown in Eq. (12), the ASM in ASDRL-AutoAug uses the parameter n to determine the persistence of performance degradation of the segmentation model and then controls the feedback value of the stop signal. Therefore, the value of n can greatly affect the training time-cost and training quality of Dueling DQN, and thus the effectiveness of data augmentation. Therefore, we conduct experiments on three datasets, Cardiac, BUSI, and LiTs, to investigate the effect of varying the hyperparameter n during the training phase on the training time-cost of the Dueling DQN and the effectiveness of augmented data.

In Table 6, we use two commonly used segmentation metrics, DSC and HD95, to evaluate the effectiveness of the augmented data, and the average time per epoch to evaluate the training time of Dueling DQN. Specifically, first, for both DSC and HD95, we find that: (i) segmentation performance is the worst when $n = 1$, and when $n = 2$, DSC is the next highest on both the Cardiac and BUSI datasets, DSC is the highest on the LiTs dataset, and HD95 is the lowest on three datasets; (ii) DSC is the best or second-best when $n = 3$, respectively, but is close to DSC of $n = 2$; (iii) when $n = 4$ or 5, the training of Dueling DQN fails to converge. This is because, in the training phase, feeding back the penalty value only once ($n = 1$) is not enough to indicate that the performance of the segmentation model is continuously decreasing, and stopping augmenting the current image at this time is not conducive to improving the adequacy of the augmentation. When $n \leq 2$ (i.e., two or more consecutive feedbacks of penalty values), the persistence of the

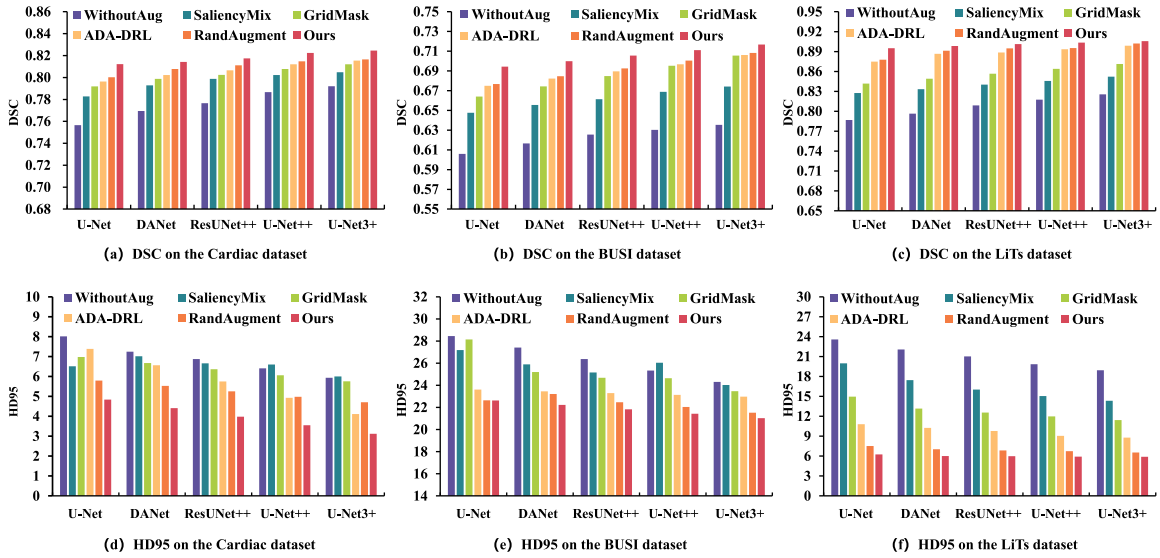

Fig. 6. Results of the proposed ASDRL-AutoAug and state-of-the-art baselines applied on different segmentation models.

segmentation performance degradation can be indicated. However, setting a larger value when n can indicate the persistence of performance degradation is unnecessary, because it will cause training difficulties to converge, such as in the case of $n = 5$. Second, in terms of training time, we find that: (i) on three datasets, the training time is the lowest when $n = 1$, and the training time is the next lowest when $n = 2$; (ii) when $n = 3$ or 4, the training time increases exponentially and is extremely high; and (iii) when $n = 5$, the training time is difficult to estimate. This is because using $n = 1$ in training is time-saving by sacrificing the adequacy of image augmentation. Since $n = 2$ is sufficient to account for the persistence of performance degradation, using more times of performance degradation to account for persistence when $n = 3$ or 4 is unnecessary and causes serious time loss. Therefore, we use 2 as the final chosen value in the training phase.

5. Discussion

5.1. Social impact of ASDRL-AutoAug

The algorithm can be widely used in realistic scenarios where the amount of data augmentation of medical images is required, effectively reducing the workload of doctors in data annotation and providing a more effective and reliable solution to the problem of insufficient finely annotated data in intelligent medical image analysis. We use cardiovascular radiotherapy as an example. Cardiovascular diseases, such as coronary heart disease, hypertension, and myocardial infarction, constitute a prevalent class of health issues with high mortality rates, impacting tens of millions of people worldwide. For various etiologies, imaging physicians are required to segment the heart chambers and identify lesion areas to formulate suitable treatment plans and monitor the treatment process. The adoption of automated segmentation techniques can substantially decrease physicians' diagnosis time. However, due to limitations in data annotation costs and concerns related to data privacy protection, acquiring large-scale, high-quality datasets becomes challenging. As a result, automated segmentation models encounter overfitting problems due to data scarcity, restricting the efficiency and performance of models when deployed clinically. While traditional data augmentation methods can augment the number and diversity of training sets, these methods usually can only process images for certain specific transformations. This limitation can introduce excessive noise and variation, leading to the loss of crucial information or resulting in data redundancy. However, our proposed ASDRL-AutoAug can automatically generate an appropriate data augmentation policy based on the characteristics of the data and the requirements of medical image segmentation models. This automated process significantly augments the size and diversity of the dataset without human intervention, consequently enhancing data utilization, model generalization ability, and robustness. Thus, our work significantly reduces the cost of medical image annotation and physician workload while safeguarding data privacy. Additionally, it enhances the performance of high-precision intelligent medical image segmentation systems in various clinical application scenarios, proving to be of great significance to the clinical practice of computer-aided diagnosis and treatment.

5.2. Limitations and future work

Our proposed automatic stopping mechanism, ASM, enables the agent to transform an image at a larger magnitude by selecting the same augmentation operation multiple times, regardless of the maximum number of augmentations. Therefore, interesting work in the future is to further improve the action space of ASDRL-AutoAug and attempt to constrain the operation amplitude to a certain range. This improvement would enable ASDRL-AutoAug to concurrently select the amplitude of the action based on visual information, thus allowing for a more refined transformation of the image. Second, we will also explore more intelligent and adaptive stopping criteria in our future work to further increase the performances and innovative value of the proposed work. In addition, in the future, we plan to conduct more experiments to investigate the extension of ASDRL-AutoAug to other vision tasks, such as object detection [59], image classification [60], and multimodal image segmentation [61,62]. This effort aims to address other challenging problems arising from limited medical images.

6. Conclusion

In this work, we identified two drawbacks of existing deep reinforcement learning-based data augmentation approaches: the problem of some augmented images being invalid and the problem of insufficient image augmentation. Subsequently, we proposed an Adaptive Sequence-Length based Deep Reinforcement Learning (ASDRL) model for Automatic Data Augmentation (AutoAug) in intelligent medical image analysis, denoted ASDRL-AutoAug. ASDRL-AutoAug addresses

these issues by utilizing a reward with dual constraints (R_{dc}) and an automatic stopping mechanism (ASM) to achieve more effective medical image augmentation. Specifically, the technical contribution of ASDRL-AutoAug is twofold: first, R_{dc} is proposed in ASDRL-AutoAug to address the problem of ineffectiveness in partially augmented images by providing a more accurate assessment of the validity of the augmentation transformations. Second, to tackle the issue of insufficient image augmentation, ASM is proposed. ASM automatically stops the augmentation process of each image based on the change in the augmentation trajectory of that image, ensuring a smooth improvement of the model performance before the augmentation is stopped. Extensive experimental studies on three real medical image segmentation datasets demonstrate that the proposed ASDRL-AutoAug is significantly more effective than current image augmentation solutions for medical image segmentation tasks. Both R_{dc} and ASM are shown to be effective and essential for the superior performance of ASDRL-AutoAug. Furthermore, we illustrate that R_{dc} evaluates the transformations more accurately than existing reward functions; and ASDRL-AutoAug is adaptive to different images in terms of sequence length and proves to be generalizable across different segmentation models.

CRediT authorship contribution statement

Zhenghua Xu: Conceptualization, Funding acquisition, Methodology, Project administration, Resources, Supervision, Writing – original draft, Writing – review & editing. **Shengxin Wang:** Conceptualization, Data curation, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft, Writing – review & editing. **Gang Xu:** Conceptualization, Formal analysis, Investigation, Methodology, Validation, Visualization, Writing – original draft. **Yunxin Liu:** Conceptualization, Methodology, Writing – review & editing. **Miao Yu:** Conceptualization, Writing – review & editing, Investigation, Resources. **Hongwei Zhang:** Conceptualization, Investigation, Writing – review & editing. **Thomas Lukasiewicz:** Conceptualization, Funding acquisition, Investigation, Resources, Supervision, Writing – review & editing. **Junhua Gu:** Conceptualization, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Acknowledgments

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, and by the Key Research and Development Project of Hainan Province, China, under the grant ZDYF2022SHFZ015. This work was also supported by the AXA Research Fund, France.

References

[1] D. Yuan, Y. Liu, Z. Xu, Y. Zhan, J. Chen, T. Lukasiewicz, Painless and accurate medical image analysis using deep reinforcement learning with task-oriented homogenized automatic pre-processing, Comput. Biol. Med. 153 (2023) 106487.
[2] Z. Xu, X. Zhang, H. Zhang, Y. Liu, Y. Zhan, T. Lukasiewicz, EFPN: Effective medical image detection using feature pyramid fusion enhancement, Comput. Biol. Med. (2023) 107149.
[3] Z. Xu, S. Liu, D. Yuan, L. Wang, J. Chen, T. Lukasiewicz, Z. Fu, R. Zhang, ω-Net: Dual supervised medical image segmentation with multi-dimensional self-attention and diversely-connected multi-scale convolution, Neurocomputing 500 (2022) 177–190.
[4] Z. Xu, B. Tian, S. Liu, X. Wang, D. Yuan, J. Gu, J. Chen, T. Lukasiewicz, V.C.M. Leung, Collaborative attention guided multi-scale feature fusion network for medical image segmentation, IEEE Trans. Netw. Sci. Eng. Early Access (2023) 1–15.
[5] D. Yuan, Z. Xu, B. Tian, H. Wang, Y. Zhan, T. Lukasiewicz, μ-Net: Medical image segmentation using efficient and effective deep supervision, Comput. Biol. Med. 160 (2023) 106963.
[6] M. Yu, M. Guo, S. Zhang, Y. Zhan, M. Zhao, T. Lukasiewicz, Z. Xu, RIRGAN: An end-to-end lightweight multi-task learning method for brain MRI super-resolution and denoising, Comput. Biol. Med. 167 (2023) 107632.
[7] M. Xu, S. Yoon, A. Fuentes, D.S. Park, A comprehensive survey of image augmentation techniques for deep learning, Pattern Recognit. 137 (2023) 109347.
[8] C. Shorten, T.M. Khoshgoftaar, A survey on image data augmentation for deep learning, J. Big Data 6 (1) (2019) 1–48.
[9] Z. Xu, J. Tang, C. Qi, D. Yao, C. Liu, Y. Zhan, T. Lukasiewicz, Cross-domain attention-guided generative data augmentation for medical image analysis with limited data, Comput. Biol. Med. 168 (2024) 107744.
[10] S. Yang, W. Xiao, M. Zhang, S. Guo, J. Zhao, F. Shen, Image data augmentation for deep learning: A survey, 2022, arXiv preprint arXiv:2204.08610.
[11] D. Han, J. Kim, J. Kim, Deep pyramidal residual networks, in: Proceedings of Computer Vision and Pattern Recognition, 2017, pp. 5927–5935.
[12] W. Liu, D. Anguelov, D. Erhan, C. Szegedy, S. Reed, C.-Y. Fu, A.C. Berg, SSD: Single shot multibox detector, in: Proceedings of European Conference on Computer Vision, 2016, pp. 21–37.
[13] H. Touvron, M. Cord, H. Jégou, DeiT III: Revenge of the ViT, in: Proceedings of European Conference on Computer Vision, 2022, pp. 516–533.
[14] L. Yang, L. Qi, L. Feng, W. Zhang, Y. Shi, Revisiting weak-to-strong consistency in semi-supervised semantic segmentation, 2022, arXiv preprint arXiv:2208.09910.
[15] Z. Zhong, L. Zheng, G. Kang, S. Li, Y. Yang, Random erasing data augmentation, in: Proceedings of the Association for the Advance of Artificial Intelligence, 34, (07) 2020, pp. 13001–13008.
[16] C. Gong, D. Wang, M. Li, V. Chandra, Q. Liu, KeepAugment: A simple information-preserving data augmentation approach, in: Proceedings of Computer Vision and Pattern Recognition, 2021, pp. 1055–1064.
[17] A. Uddin, M. Monira, W. Shin, T. Chung, S.-H. Bae, et al., SaliencyMix: A saliency guided data augmentation strategy for better regularization, 2020, arXiv preprint arXiv:2006.01791.
[18] A. Ferreira, J. Li, K.L. Pomykala, J. Kleesiek, V. Alves, J. Egger, GAN-based generation of realistic 3D data: A systematic review and taxonomy, 2022, arXiv preprint arXiv:2207.01390.
[19] E.D. Cubuk, B. Zoph, D. Mane, V. Vasudevan, Q.V. Le, Autoaugment: Learning augmentation strategies from data, in: Proceedings of Computer Vision and Pattern Recognition, 2019, pp. 113–123.
[20] E.D. Cubuk, B. Zoph, J. Shlens, Q.V. Le, RandAugment: Practical automated data augmentation with a reduced search space, in: Proceedings of Computer Vision and Pattern Recognition, 2020, pp. 702–703.
[21] K. Tian, C. Lin, M. Sun, L. Zhou, J. Yan, W. Ouyang, Improving auto-augment via augmentation-wise weight sharing, Adv. Neural Inf. Process. Syst. 33 (2020) 19088–19098.
[22] J. Xu, M. Li, Z. Zhu, Automatic data augmentation for 3D medical image segmentation, in: Proceedings of Medical Image Computing and Computer Assisted Intervention, 2020, pp. 378–387.
[23] P. Li, X. Liu, X. Xie, Learning sample-specific policies for sequential image augmentation, in: Proceedings of ACM International Conference on Multimedia, 2021, pp. 4491–4500.
[24] T.N. Minh, M. Sinn, H.T. Lam, M. Wistuba, Automated image data preprocessing with deep reinforcement learning, 2018, arXiv preprint arXiv:1806.05886.
[25] T. Qin, Z. Wang, K. He, Y. Shi, Y. Gao, D. Shen, Automatic data augmentation via deep reinforcement learning for effective kidney tumor segmentation, in: Proceedings of International Conference on Acoustics, Speech and Signal Processing, 2020, pp. 1419–1423.
[26] M.D. Bloice, P.M. Roth, A. Holzinger, Biomedical image augmentation using augmentor, Bioinformatics 35 (21) (2019) 4522–4524.
[27] T. DeVries, G.W. Taylor, Improved regularization of convolutional neural networks with cutout, 2017, arXiv preprint arXiv:1708.04552.
[28] K. Kumar Singh, Y. Jae Lee, Hide-and-seek: Forcing a network to be meticulous for weakly-supervised object and action localization, in: Proceedings of International Conference on Computer Vision, 2017, pp. 3524–3533.
[29] P. Chen, S. Liu, H. Zhao, J. Jia, GridMask data augmentation, 2020, arXiv preprint arXiv:2001.04086.
[30] Y. Kim, A.S. Uddin, S.-H. Bae, Local augment: Utilizing local bias property of convolutional neural networks for data augmentation, Access 9 (2021) 15191–15199.
[31] J.-W. Seo, H.-G. Jung, S.-W. Lee, Self-augmentation: Generalizing deep networks to unseen classes for few-shot learning, Neural Netw. 138 (2021) 140–149.
[32] J. Choi, C. Lee, D. Lee, H. Jung, SalfMix: A novel single image-based data augmentation technique using a saliency map, Sensors 21 (24) (2021) 8444.
[33] H. Zhang, M. Cisse, Y.N. Dauphin, D. Lopez-Paz, Mixup: Beyond empirical risk minimization, 2017, arXiv preprint arXiv:1710.09412.
[34] S. Yun, D. Han, S.J. Oh, S. Chun, J. Choe, Y. Yoo, CutMix: Regularization strategy to train strong classifiers with localizable features, in: Proceedings of International Conference on Computer Vision, 2019, pp. 6023–6032.

[35] B. Zoph, E.D. Cubuk, G. Ghiasi, T.-Y. Lin, J. Shlens, Q.V. Le, Learning data augmentation strategies for object detection, in: Proceedings of European Conference on Computer Vision, 2020, pp. 566–583.
[36] S. Lim, I. Kim, T. Kim, C. Kim, S. Kim, Fast AutoAugment, Adv. Neural Inf. Process. Syst. 32 (2019).
[37] R. Hataya, J. Zdenek, K. Yoshizoe, H. Nakayama, Faster AutoAugment: Learning augmentation strategies using backpropagation, in: Proceedings of European Conference on Computer Vision, 2020, pp. 1–16.
[38] Y. Li, G. Hu, Y. Wang, T. Hospedales, N.M. Robertson, Y. Yang, Differentiable automatic data augmentation, in: Proceedings of the European Computer Vision Conference, 2020, pp. 580–595.
[39] Y. Tang, B. Li, M. Liu, B. Chen, Y. Wang, W. Ouyang, Autopedestrian: An automatic data augmentation and loss function search scheme for pedestrian detection, IEEE Trans. Image Process 30 (2021) 8483–8496.
[40] S. Zhou, Y. Tang, M. Liu, Y. Wang, H. Wen, Impartial differentiable automatic data augmentation based on finite difference approximation for pedestrian detection, IEEE Trans. Instrum. Meas. 71 (2022) 1–11.
[41] D. Yang, H. Roth, Z. Xu, F. Milletari, L. Zhang, D. Xu, Searching learning strategy with reinforcement learning for 3D medical image segmentation, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 3–11.
[42] W. He, M. Liu, Y. Tang, Q. Liu, Y. Wang, Differentiable automatic data augmentation by proximal update for medical image segmentation, IEEE/CAA J. Autom. Sin. 9 (7) (2022) 1315–1318.
[43] Z. Liu, Q. Lv, Y. Li, Z. Yang, L. Shen, Medaugment: universal automatic data augmentation plug-in for medical image analysis, 2023, arXiv preprint arXiv: 2306.17466.
[44] O. Ronneberger, P. Fischer, T. Brox, U-Net: Convolutional networks for biomedical image segmentation, in: Proceedings of Medical Image Computing and Computer-Assisted Intervention, 2015, pp. 234–241.
[45] Z. Wang, T. Schaul, M. Hessel, H. Hasselt, M. Lanctot, N. Freitas, Dueling network architectures for deep reinforcement learning, in: Proceedings of International Conference on Machine Learning, 2016, pp. 1995–2003.
[46] G. Xu, S. Wang, Z. Xu, T. Lukasiewicz, Adaptive-masking policy with deep reinforcement learning for self- supervised medical image segmentation, in: Proceedings of IEEE International Conference on Multimedia & Expo, 2023, http://dx.doi.org/10.1109/ICME55011.2023.00390.
[47] Y. Bai, X. Ding, D. Hu, Y. Jiang, Research on dynamic path planning of multi-AGVs based on reinforcement learning, Appl. Sci. 12 (16) (2022) 8166.
[48] V. Mnih, K. Kavukcuoglu, D. Silver, A.A. Rusu, J. Veness, M.G. Bellemare, A. Graves, M. Riedmiller, A.K. Fidjeland, G. Ostrovski, et al., Human-level control through deep reinforcement learning, Nature 518 (7540) (2015) 529–533.
[49] M. Hessel, H. Soyer, L. Espeholt, W. Czarnecki, S. Schmitt, H. Van Hasselt, Multi-task deep reinforcement learning with popart, in: Proceedings of the AAAI Conference on Artificial Intelligence, 33, (01) 2019, pp. 3796–3803.
[50] A.L. Simpson, M. Antonelli, S. Bakas, M. Bilello, K. Farahani, B. Van Ginneken, A. Kopp-Schneider, B.A. Landman, G. Litjens, B. Menze, et al., A large annotated medical image dataset for the development and evaluation of segmentation algorithms, 2019, arXiv preprint arXiv:1902.09063.
[51] W. Al-Dhabyani, M. Gomaa, H. Khaled, A. Fahmy, Dataset of breast ultrasound images, Data in brief 28 (2020) 104863.
[52] Q. Yu, Y. Xia, L. Xie, E.K. Fishman, A.L. Yuille, Thickened 2D networks for efficient 3D medical image segmentation, 2019, arXiv preprint arXiv:1904.01150.
[53] D.P. Kingma, J. Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint arXiv:1412.6980.
[54] D. Karimi, S.E. Salcudean, Reducing the hausdorff distance in medical image segmentation with convolutional neural networks, Trans. Med. Imag 39 (2) (2019) 499–513.
[55] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, H. Lu, Dual attention network for scene segmentation, in: Proceedings of Computer Vision and Pattern Recognition, 2019, pp. 3146–3154.
[56] D. Jha, P.H. Smedsrud, M.A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, H.D. Johansen, ResUNet++: An advanced architecture for medical image segmentation, in: Proceedings of International Symposium on Multimedia, 2019, pp. 225–2255.
[57] Z. Zhou, M.M. Rahman Siddiquee, N. Tajbakhsh, J. Liang, UNet++: A nested U-Net architecture for medical image segmentation, in: Proceedings of Medical Image Computing and Computer Assisted Intervention, 2018, pp. 3–11.
[58] H. Huang, L. Lin, R. Tong, H. Hu, Q. Zhang, Y. Iwamoto, X. Han, Y.-W. Chen, J. Wu, UNet 3+: A full-scale connected UNet for medical image segmentation, in: Proceedings of International Conference on Acoustics, Speech and Signal Processing, 2020, pp. 1055–1059.
[59] D. Yao, Z. Xu, Y. Lin, Y. Zhan, Accurate and intelligent diagnosis of pediatric pneumonia using X-ray images and blood testing data, Front. Bioeng. Biotechnol 11 (2023) 1058888.
[60] Z. Xu, T. Li, Y. Liu, Y. Zhan, J. Chen, T. Lukasiewicz, PAC-Net: Multi-pathway FPN with position attention guided connections and vertex distance IoU for 3D medical image detection, Front. Bioeng. Biotechnol 11 (2023) 1049555.
[61] J. Zhang, S. Zhang, X. Shen, T. Lukasiewicz, Z. Xu, Multi-ConDoS: Multimodal contrastive domain sharing generative adversarial networks for self-supervised medical image segmentation, IEEE Trans. Med. Imaging Early Access (2023) 1–20, http://dx.doi.org/10.1109/TMI.2023.3290356.
[62] S. Zhang, J. Zhang, B. Tian, T. Lukasiewicz, Z. Xu, Multi-modal contrastive mutual learning and pseudo-label re-learning for semi-supervised medical image segmentation, Med. Image Anal. 83 (2023) 102656.

Zhenghua Xu received a M.Phil. in Computer Science from The University of Melbourne, Australia, in 2012, and a D.Phil in computer Science from University of Oxford, United Kingdom, in 2018. From 2017 to 2018, he worked as a research associate at the Department of Computer Science, University of Oxford. He is now a professor at the Hebei University of Technology, China, and a awardee of "100 Talents Plan" of Hebei Province. He has published about 50 papers in top journals and conferences, e.g., Nature Neuroscience, IEEE TMI, Medical Image Analysis, NeurIPS, CVPR, AAAI, IJCAI, ICDE, etc. His current research focuses on intelligent medical image analysis using deep learning, reinforcement learning, federated learning techniques.

Shengxin Wang is currently a master student at Hebei University of Technology, China. She received B.Eng. degree in Biomedical Engineering from Hebei University of Technology, China, in 2020. Her research interests lie in medical image analysis using deep reinforcement learning.

Gang Xu is currently a master student at Hebei University of Technology, China. He received B.Eng. degree in Automation Specialty from North China University of Science and Technology, China, in 2020. His research interests lie in medical image analysis using deep reinforcement learning.

Yunxin Liu is currently a Ph.D. student at Hebei University of Technology, China. He received the master degree in computer science from Jiangxi Normal University, Nanchang, China, in 2021. His research interests lie in medical image analysis using deep reinforcement learning.

Miao Yu is currently a PHD student in the State Key Laboratory of Reliability and Intelligence of Electrical Equipment, School of Health Sciences and Biomedical Engineering, Hebei University of Technology, Tianjin, China. She received Master degree in Hebei University of Technology, China, in 2018. Her research interests lie in medical image processing using deep learning methods.

Hongwei Zhang is with Tianjin University of Technology, he received the Ph.D. degrees from the Department of Computer Science and Engineering, Tianjin University of Technology, Tianjin, China. He is currently a lecturer with the School of Computer Science and Engineering, Tianjin University of Technology. His main research interests include blockchain, privacy protection and data security sharing.

Thomas Lukasiewicz is a Professor of Computer Science at the Department of Computer Science, University of Oxford, UK, heading the Intelligent Systems Lab within the Artificial Intelligence and Machine Learning Theme. He currently holds an AXA Chair grant on "Explainable Artificial Intelligence in Healthcare" and a Turing Fellowship at the Alan Turing Institute, London, UK, which is the UK's National Institute for Data Science and Artificial Intelligence. He received the IJCAI-01 Distinguished Paper Award, the AIJ Prominent Paper Award 2013, the RuleML 2015 Best Paper Award, and the ACM PODS Alberto O. Mendelzon Test-of-Time Award 2019. He is a Fellow of the European Association for Artificial Intelligence (EurAI) since 2020. His research interests are especially in artificial intelligence and machine learning.

Junhua Gu received the B.S degree in mathematics from Shanghai Jiaotong University, Shanghai, China, in 1988, the M.S. degree in computer science and the Ph.D. degree in electrical engineering from the Hebei University of Technology, Tianjin, China, in 1993 and 1997, respectively. He is currently a Professor at the Hebei University of Technology, China. He has authored more than 70 papers. His current research interests include big data, intelligent control, and intelligent transportation systems. Prof. Gu was awarded the Hebei New Century "333 Talent Project" Second Level Suitable Person.