



Painless and accurate medical image analysis using deep reinforcement learning with task-oriented homogenized automatic pre-processing

Di Yuan ^a, Yunxin Liu ^{a,*}, Zhenghua Xu ^{a,*}, Yuefu Zhan ^b, Junyang Chen ^c, Thomas Lukasiewicz ^d

^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 Department of Radiology, Hainan Women and Children's Medical Center, Haikou, China

^c College of Computer Science and Software Engineering and the Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ), Shenzhen University, Shenzhen, China

^d Department of Computer Science, University of Oxford, Oxford, United Kingdom

ARTICLE INFO

Keywords:

Homogenized automatic pre-processing
Deep reinforcement learning
Medical image analysis
Policy gradient
Pediatric pneumonia classification

ABSTRACT

Pre-processing is widely applied in medical image analysis to remove the interference information. However, the existing pre-processing solutions mainly encounter two problems: (i) it is heavily relied on the assistance of clinical experts, making it hard for intelligent CAD systems to deploy quickly; (ii) due to the personnel and information barriers, it is difficult for medical institutions to conduct the same pre-processing operations, making a deep model that performs well on a specific medical institution difficult to achieve similar performances on the same task in other medical institutions. To overcome these problems, we propose a deep-reinforcement-learning-based task-oriented homogenized automatic pre-processing (DRL-HAPre) framework to overcome these two problems. This framework utilizes deep reinforcement learning techniques to learn a policy network to automatically and adaptively select the optimal pre-processing operations for the input medical images according to different analysis tasks, thus helping the intelligent CAD system to achieve a rapid deployment (i.e., painless) and maintain a satisfactory performance (i.e., accurate) among different medical institutes. To verify the effectiveness and advantages of the proposed DRL-HAPre framework, we further develop a homogenized automatic pre-processing model based on the DRL-HAPre framework to realize the automatic pre-processing of key region selection (called HAPre-KRS) in the pneumonia image classification task. Extensive experimental studies are conducted on three pediatric pneumonia classification datasets with different image qualities, and the results show that: (i) There does exist a hard-to-reproduce problem in clinical practices and the fact that having different medical image qualities in different medical institutes is an important reason for the existing of hard-to-reproduce problem, so it is compelling to propose homogenized automatic pre-processing method. (ii) The proposed HAPre-KRS model and DRL-HAPre framework greatly outperform three kinds of state-of-the-art baselines (i.e., pre-processing, attention and pneumonia baseline), and the lower the medical image quality, the greater the improvements of using our HAPre-KRS model and DRL-HAPre framework. (iii) With the help of homogenized pre-processing, HAPre-KRS (and DRL-HAPre framework) can greatly avoid performance degradation in real-world cross-source applications (i.e., thus overcoming the hard-to-reproduce problem).

1. Introduction

In the past decades, medical imaging technology has been widely used in clinical diagnosis and treatment to help doctors understand the physical condition of patients [1]. But reading and analyzing such a large amount of medical images will cost clinicians and radiologists a lot of time and energy, and may also result in misdiagnosis due to the limitation of expertise and/or fatigue; so, in recent years, deep learning techniques have been increasingly applied in clinical computer-aided

diagnosis (CAD) for automatic medical image analysis [2]. Due to the differences in the conditions of imaging equipments, surroundings and patients, medical images usually contain a lot of interfering information (e.g., noise, artifacts, irrelevant objects, etc.); therefore, pre-processing is needed for almost all the deep-learning-based medical image analysis models to achieve satisfactory performances.

At present, the pre-processing of medical images mainly relies on the professional knowledge and experience of clinical experts, i.e., for

* Corresponding authors.

E-mail addresses: 202111402002@stu.hebut.edu.cn (Y. Liu), zhenghua.xu@hebut.edu.cn (Z. Xu).

different interference information in different types of medical images, we have to rely on the professional knowledge and experience of clinical experts to determine what kinds of pre-processing operations should be applied to make the features in medical images easier to learn. However, this will inevitably result in the following two problems in the application of deep-learning-based CAD systems in clinical practices. (i) **Hard to deploy quickly**: In order to help deep medical image analysis models achieve satisfactory performances, clinical experts are needed to assist in the pre-processing of medical images, which not only greatly consumes the very limited social medical resources, but also make the rapid deployment of deep-learning-based CAD systems in clinical practice become difficult. (ii) **Hard to reproduce**: We notice that a model that performs well on a specific medical image dataset of a medical institution often difficult to achieve similar performances in the same medical image analysis task of other medical institutions using the same types of medical images (whether the model is re-trained or used directly); we believe an important reason of this failure is the lack of homogenized pre-processing, i.e., due to the personnel and information barriers between different medical institutions, it is difficult for other medical institutions to conduct the same pre-processing operations on the new medical image datasets, so the difference in data quality after pre-processing inevitably leads to the different qualities of model training and inference.

In order to verify our argument on the hard-to-reproduce problem and the need of homogenized pre-processing in deep-learning-based medical image analysis tasks, preliminary experimental studies on pediatric pneumonia classification are conducted. Specifically, a widely used ResNet model is first applied on a well pre-processed public pediatric pneumonia X-ray image dataset from Guangzhou Women and Children's Medical Center [3] (image examples are shown in Fig. 1(a)–(c)), resulting in a classification F1-score at around 0.82; however, when the ResNet model (with the same setting) is then applied on another pediatric pneumonia X-ray images without pre-processing (see Fig. 1(d)–(f)) from another medical institution, Hainan Women and Children's Medical Center, the classification F1-score is dramatically dropped down to around 0.64; finally, by conducting the key region selection pre-processing on the latter dataset (see Fig. 1(g)–(i)), the F1-score rises back to around 0.68. This demonstrates that an important reason (certainly not the only reason) for the hard-to-reproduce problem (i.e., a good medical image analysis model is difficult to reproduce effectively in other medical institutions) is that the quality of medical images used for model training and inference in different medical institutions is different, and these quality differences are largely (but not completely) caused by the differences in pre-processing operations between them; so when the quality difference of their images is reduced, the performance difference between them can also be effectively reduced (e.g., the performance gap in the above example is narrowed from 0.18 to 0.14 because the quality gap is narrowed). Consequently, if homogenized medical image pre-processing can be achieved between different medical institutions, it will greatly reduce the image quality difference, thus alleviating the hard-to-reproduce problem of deep-learning-based CAD systems in clinical practices. However, as mentioned above, achieving high-quality medical image pre-processing heavily relies on the assistance of clinical experts, and due to the different subjective judgments, experiences and expertise of different experts, it is very difficult for different medical institutions to achieve high-quality and homogenized medical image pre-processing.

Motivated by these observations, in this paper, we propose a Deep-Reinforcement-Learning-based task-oriented Homogenized Automatic Pre-processing (Abbreviated as DRL-HAPre) framework to overcome the above two problems and achieve painless (resolving the hard-to-deploy-quickly problem) and accurate (alleviating the hard-to-reproduce problem) medical image analysis in clinical practices. Specifically, DRL-HAPre framework mainly consists of a task network and an auxiliary network, where the task network is a deep model used to achieve the corresponding medical image analysis tasks (e.g., we

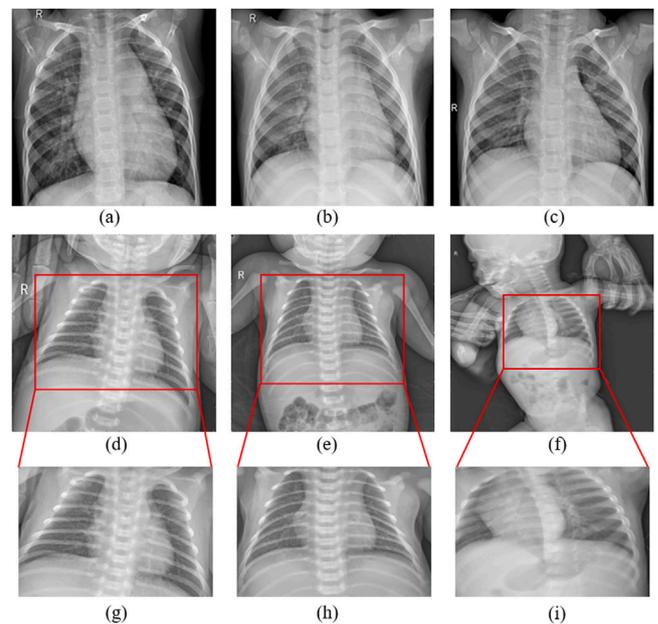


Fig. 1. (a)–(c) are three examples of pre-processed pediatric pneumonia X-ray images in a public dataset provided by Guangzhou Women and Children's Medical Center [3], (d)–(f) are examples of pediatric pneumonia X-ray images collected from Hainan Women and Children's Medical Center without pre-processing, (g)–(i) are results of conducting key region selection pre-processing on (d)–(f).

can use ResNet for classification tasks, U-Net for Segmentation tasks, etc.), and the auxiliary network is a policy network that aims to automatically and adaptively select the optimal pre-processing operations for the input medical images according to different analysis tasks. To ensure the pre-processing policies generated by the auxiliary network is optimal for the input image on the given task, the deep-reinforcement-learning-based policy gradient method is utilized to interactively train the auxiliary network using rewards that are calculated based on the results of the task network. Therefore, DRL-HAPre framework has the following advantages: (i) the pre-processing operations conducted in DRL-HAPre framework are fully automatic, i.e., do not need the assistance of clinical experts, so it can overcome the hard-to-deploy-quickly problem and make the deployment of CAD systems painless in pre-processing; (ii) when applying a well trained DRL-HAPre framework to other medical institutes for the same type of medical image analysis tasks (i.e., cross-source applications), since the same auxiliary network is used to generate pre-processing policies for new data, DRL-HAPre can alleviate the hard-to-reproduce problem using task-oriented homogenized pre-processing and will achieve much more stable cross-source performances than the conventional deep solutions.

To verify the effectiveness and advantages of the proposed DRL-HAPre framework, we continue to use the pediatric pneumonia classification task as a study case, and further develop a homogenized automatic pre-processing model based on the DRL-HAPre framework to realize the automatic pre-processing of key region selection (called HAPre-KRS) in the pneumonia image classification task. Specifically, given an input medical image, HAPre-KRS first estimates its original classification loss using a ResNet-based task network; then an unsupervised deep detection network is used as the auxiliary network to find its key region, and the detected key regions are send into the task network to obtain an augmented loss; finally, a reward is estimated (based on the difference between the original and augmented loss) and fed-back to learn the auxiliary detection network using policy gradient [4].

In summary, the main contributions of this work are as follows:

- We identify the challenging hard-to-deploy-quickly and hard-to-reproduce problems in the existing pre-processing solutions, and

propose a deep-reinforcement-learning-based task-oriented homogenized automatic pre-processing (DRL-HAPre) framework to overcome these two problems and help achieve painless (i.e., due to automation) and accurate (due to Homogenized quality enhancement) medical image analysis in clinical practices.

- We further develop a homogenized automatic pre-processing model based on the DRL-HAPre framework to realize the automatic pre-processing of key region selection (called HAPre-KRS) in the pneumonia image classification task.
- Extensive experimental studies are conducted on three pediatric pneumonia classification datasets with different image qualities, and the results show that: (i) There does exist a hard-to-reproduce problem in clinical practices and the fact that having different medical image qualities in different medical institutes is an important reason for the existing of hard-to-reproduce problem, so it is compelling to propose homogenized automatic pre-processing method. (ii) The proposed HAPre-KRS model and DRL-HAPre framework greatly outperform three kinds of state-of-the-art baselines (i.e., pre-processing, attention and pneumonia baseline), and the lower the medical image quality, the greater the improvements of using our HAPre-KRS model and DRL-HAPre framework. (iii) With the help of homogenized pre-processing, HAPre-KRS (and DRL-HAPre framework) can greatly avoid performance degradation in real-world cross-source applications (i.e., thus overcoming the hard-to-reproduce problem).

The rest of this paper is organized as follows. Section 2 presents the related work and clarify the differences between our work and the references. The detailed methodology of the proposed DRL-HAPre framework and HAPre-KRS model is given in Section 3, which is followed by the experimental studies in Section 4. Section 5 discusses the social impact of the proposed works, while the conclusions and future works are given in Section 6.

2. Related work

In this section, we briefly discuss the closely related works on medical image analysis, deep reinforcement learning, and pre-processing.

Medical image analysis. Recently, deep learning technology has been increasingly applied in various fields of computer-aided diagnosis to help doctors analyze medical images. For example, in medical image classification, Yang et al. [5] propose a new training strategy called deep tree training, which is used to jointly train a series of networks constructed from the hidden layers of CNN in a hierarchical manner to help label medical images with some quality distortion. Devi et al. [6] propose an architecture having a three-stream network to better extract manual and automatic features from images to more accurately identify lung cancer. Zlocha et al. [7] use the feature pyramid network as the backbone of RetinaNet [8] and weak labels as auxiliary supervision to achieve the detection of lesions in CT images. Zhao et al. [9] use a feature pyramid network to extract local and global features of different receptive fields to achieve automatic detection for cancer metastasis from whole slide images. And Xu [2] et al. propose a new deep model ω -Net, which incorporates an additional expansive path into U-Net to import an extra supervision signal and obtain a more effective and robust image segmentation by dual supervision to achieve a more accurate medical image segmentation. Yu [10] et al. develop a new parallel dilated convolution module, which uses dilated convolutions to systematically aggregate multi-scale contextual information without losing resolution for dense segmentation to improve the segmentation accuracy. Most deep learning methods can only work on specific medical image analysis tasks, but our proposed framework is not limited by this. Although this paper only uses the DRL-HAPre framework for the pediatric pneumonia classification task to prove the effectiveness of our framework in homogenized automatic pre-processing, our framework

can be applied to other medical image analysis tasks by substituting different auxiliary and task networks.

Deep reinforcement learning. Inspired by the success of many deep reinforcement learning algorithms in the gaming domain [11], deep reinforcement learning has also been applied to medical image analysis. Akrouf et al. [12] propose to integrate a CNN classification model with a reinforcement-learning-based question answering agent for skin disease classification. To better identify underlying conditions, an agent learns how to ask patients if they have symptoms using the visual information provided by CNN and answers to the questions asked. Cheng et al. [13] study how to use semantic segmentation to generate a hard attention map that improves classification performance. Segmentation agents generate segmentation templates, which are trained through a reinforcement learning framework, and the reward is classification accuracy. To solve the problem of insufficient data in medical image classification, Ye et al. [14] study the synthetic sample selection problem for improved image classification, training a deep reinforcement learning agent to select synthetic images containing reliable informative features through proximal policy optimization, and classification accuracy as a bonus to ensure the quality of synthetic images for data augmentation purposes. Wang et al. [15] distinguish between benign and malignant breast nodules by proposing a multi-modal network combining four different types of ultrasonography. Similar to our work, Xu et al. [16] propose a deep selective attention method that aims to select valuable regions from raw images for classification. They develop a decision network to decide where to crop and whether the cropped patch is necessary for classification, then the classification network is trained on these selected patches and then provides feedback to the decision network to update its selection policy. Compared with the above methods, our method has two main differences: (i) The designed policy gradient solution in our work is different, its training procedure is much simpler and requires less supervisory information. (ii) We work on a different task: to our knowledge, we are the first deep reinforcement learning work that focuses on the importance of homogeneous pre-processing and use homogeneous automatic pre-processing to improve the medical image analysis models' performances.

Pre-processing. Pre-processing via image transformations has been used to improve generalization since the inception of convolutional networks. These works are similar to data augmentation. Following AlexNet [17], they have become a standard part of training pipelines. For object classification tasks, the transformations are selected to avoid changing the semantic category, i.e. translations, scales, color shifts, etc. In contrast to deep learning, data augmentation is rarely used in reinforcement learning. For example, Laskin et al. [18] propose a reinforcement learning model with augmented data by pre-processing operations and introduce two new data augmentations—random translation and random amplitude scale, and they prove that data augmentations alone can significantly improve the data efficiency and generalization of reinforcement learning methods operating from pixels, without any changes to the underlying reinforcement learning algorithm. Qin et al. [19] propose an automatic enhancement algorithm for medical image segmentation based on deep reinforcement learning so that the data enhancement strategy network is fed back and adjusted through the performance difference of the segmentation network before and after data enhancement to obtain the optimal data enhancement policy. Yarats et al. [20] use standard image transformations through pre-processing to perturb input observations as well as regularizing the Q-function learned by the critic so that different transformations of the same input image have similar Q-function values and further study image augmentation in sample-efficient reinforcement learning. However, these works mainly focus on how to improve the model's performance by using pre-processing, without paying attention to the necessity of automatic homogenization pre-processing. We are the first work aiming to discover the importance of homogenization pre-processing for the rapid application of CAD systems in different medical institutions, and use deep reinforcement learning to find the best homogenization automatic pre-processing for the corresponding task.

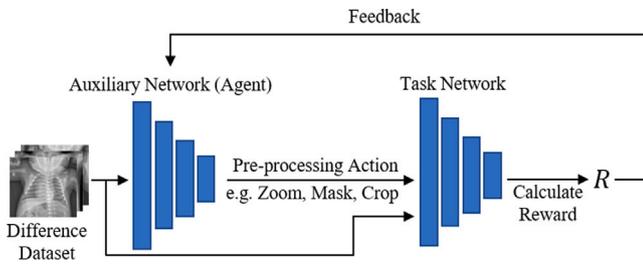


Fig. 2. Overview of our deep-reinforcement-learning-based task-oriented homogenized automatic pre-processing (DRL-HAPre) framework.

3. A deep-reinforcement-learning-based task-oriented homogenized automatic pre-processing framework

There are two major problems in existing medical image analysis tasks: first, medical images need to be pre-processed to improve their accuracy, which is time- and labor-consuming, and which requires different pre-processing according to different tasks, resulting in poor generalization. Besides, for some tasks with a small amount of data, it is necessary to train on similar medical images with a large amount of data before testing, but different data sources will reduce the model's accuracy and robustness. To solve the above problems, we propose a deep-reinforcement-learning-based task-oriented homogenized automatic pre-processing framework for medical image analysis, which is called *DRL-HAPre*. As shown in Fig. 2, *DRL-HAPre* framework is built based on policy gradients in deep reinforcement learning, which mainly consists of an auxiliary network for automatically selecting suitable pre-processing operations and a task network for achieving specific medical analysis tasks.

Specifically, this framework first utilizes the auxiliary network to automatically and adaptively select some pre-processing operations for the input images; the processed images are input into the task network for fine-tuning; then a reward is constructed by comparing the difference of task accuracies before and after pre-processing; finally, the reward is fed back to the auxiliary network to optimize the auxiliary network. Through such interactive training optimization, the auxiliary network is gradually able to automatically and adaptively select the optimal pre-processing operations (i.e., most conducive to improving the performance of the corresponding medical image analysis task) for each input image according to different medical image analysis tasks. Therefore, the *DRL-HAPre* framework has the following advantages: (i) the pre-processing operations conducted in the *DRL-HAPre* framework are fully automatic, i.e., do not need the assistance of clinical experts, so it can overcome the hard-to-deploy-quickly problem and make the deployment of CAD systems painless in pre-processing; (ii) when applying a well trained *DRL-HAPre* framework to other medical institutes for the same type of medical image analysis tasks (i.e., cross-source applications), since the same auxiliary network is used to generate pre-processing policies for new data, *DRL-HAPre* can alleviate the hard-to-reproduce problem using task-oriented homogenized pre-processing and achieve much more stable cross-source performances than the conventional deep solutions.

In this paper, to show the advantages of the *DRL-HAPre* framework, we focus on the task of pneumonia classification and further propose a homogenized automatic pre-processing model based on the *DRL-HAPre* framework to achieve the automatic pre-processing of key region selection in pneumonia image classification tasks, the resulting model is called *HAPre-KRS*. Specifically, as shown in Fig. 3, *HAPre-KRS* uses a pneumonia classification network (with parameters θ_c) as the task network and a detection network (with parameters θ_d) as the auxiliary network. Please note that *HAPre-KRS* is a specific model that is designed based on the *DRL-HAPre* framework to achieve homogenized automatic key region selection and to enhance the performances of

deep-learning-based pneumonia classification; to achieve other kinds of homogenized automatic pre-processing (e.g., zoom, mask, crop, etc.) in other medical image analysis tasks (e.g., segmentation or detection), the task network should be replaced by a network that matches the corresponding task properly (e.g., U-Net for segmentation tasks and RCNN for detection tasks), while the auxiliary network should be replaced by a new deep network which can help the agent to learn proper policies to achieve the corresponding pre-processing operations in a homogenized and automatic way.

As shown in Fig. 3, the detailed procedure of *HAPre-KRS* is as follows. First, the original chest X-ray images are input into the classification network to get the original classification loss \mathcal{L}_{ori} . Then, the obtained key areas from the detection network are input to the classification network to get the augmented classification loss \mathcal{L}_{aug} . Finally, the framework can be further divided into two different update policies and the rewards according to the difference D_1 between the original classification loss \mathcal{L}_{ori} and the augmented classification loss \mathcal{L}_{aug} after getting a key area from the detection network. On the one hand, if the difference D_1 is less than 0, the key area is valid, and then returns to the detection network to continue training. On the other hand, if the difference D_1 is at least 0, then the key area needs to be updated. We first enlarge the key area separately by α times and shrink it by α times to explore the appropriate key area, that is, we input them into the classification network to get the new classifications \mathcal{L}_{expand} and \mathcal{L}_{shrink} , respectively. The new key area chooses the one with the minimum loss. See Algorithm 1 for details. Note that although simple pre-processing actions (i.e., shrink and expand) are selected here, actually different pre-processing actions can be selected for different tasks. After that, we get the new difference between the minimum loss and the original enhanced classification loss, and then the weighted sum of the new difference and the difference D_1 is fed back to the detection network as a reward to continue training.

3.1. Policy gradient training

Following the policy gradient context [4], our detection network is seen as a policy, the image batch is treated as a state s_t , and the key-area-based direct feedback, expanded or shrank by α times, is framed as the action a_t . In detail, at each t th training step, the classification network receives the optimal key area as the input and outputs a reward signal R_t . To accomplish this, our objective is to maximize the expected reward $J(\theta_d)$ and find the optimal detection policy:

$$J(\theta_d) = E_d(a_t; \theta_d)[R_t], \quad (1)$$

where $E_d(a_t; \theta_d)$ is the expected reward concerning the key area after taking an action a_t when the policy model has been parameterized with θ_d . The policy is learned through back-propagation, which requires the definition of the gradient of the expected reward for the model parameters. Following the REINFORCE rule [4], the gradient can be defined as:

$$\nabla_{\theta_d} J(\theta_d) = E_d(a_t; \theta_d)[\nabla_{\theta_d} \log(P(a_t|s_t; \theta_d)) \cdot R_t]. \quad (2)$$

The expected reward cannot be estimated and requires an approximation. As is common practice in the traction of policy gradient, we can achieve such approximation using the negative log-likelihood loss, which is differentiable for the model parameters, and can be properly weighted by the reward signal to obtain the detection policy loss presented as follow:

$$\mathcal{L}_d = \mathcal{L}_{FL}(\hat{a}_t, a_t) \cdot R_t, \quad (3)$$

where $\hat{a}_t \sim \pi \sim P(a_t|s_t; \theta_p)$, and \mathcal{L}_{FL} is the focal loss:

$$\mathcal{L}_{FL} = - \sum_{i=1}^N (1 - p(x_i))^{\gamma} \log(p(x_i)), \quad (4)$$

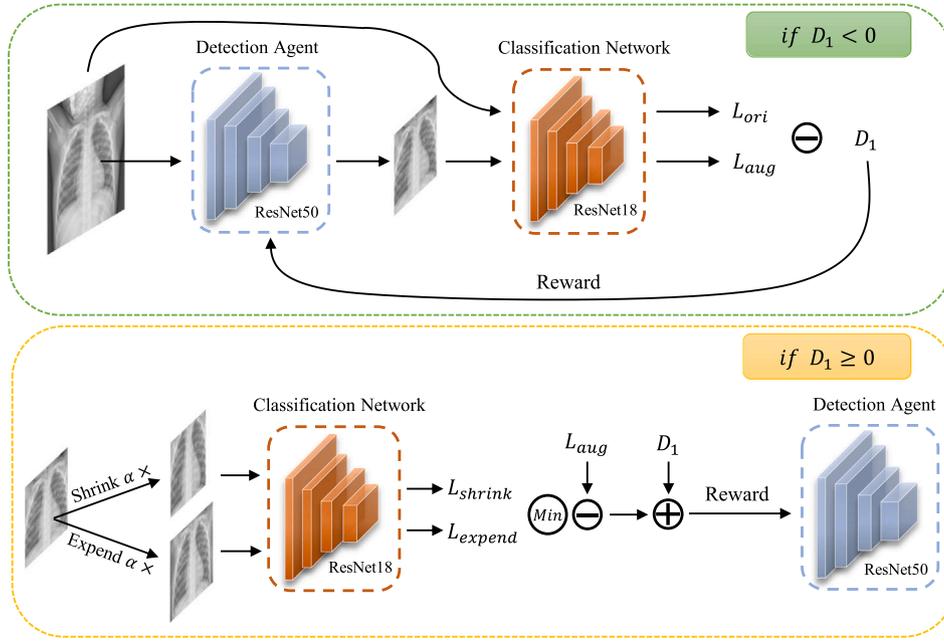


Fig. 3. Overview of a homogenized automatic pre-processing model based on DRL-HAPre framework to achieve the automatic pre-processing of key region selection in pneumonia image classification tasks (abbreviated as HAPre-KRS).

Algorithm 1 Deep Reinforcement Learning with Task-Oriented Homogenized Automatic Pre-processing Framework (DRL-HAPre)

Input: Training steps T ; training original images and labels (X, Y) ; classification loss function using cross-entropy loss \mathcal{L}_c and detection loss function using focal loss \mathcal{L}_d ; learning rates (α_c, α_d) ; the coordinate value of bounding box (i.e., key area) from the detection network $BB_{detection} = (x_{min}, y_{min}, x_{max}, y_{max})$; classification and detection networks (M_c, M_d) parameterized with (θ_c, θ_d) .

Output: M_c and M_d .

- 1: Initialize models parameters θ_d and θ_c by the ImageNet dataset
- 2: **for** $t = 1, \dots, T$ **do**
- 3: Sample a training batch (x, y) from the (X, Y) pool
- 4: Get the original classification loss $\mathcal{L}_{ori} = \mathcal{L}_c(M_c(x_t), y_t)$
- 5: Get a key area x_{key} of x_t from the detection network and its coordinate value $BB_{detection}$
- 6: Calculate a new classification loss $\mathcal{L}_{aug} = \mathcal{L}_c(M_c(x_{key}), y_t)$
- 7: Calculate the difference between the two classification losses $D_1 = \mathcal{L}_{aug} - \mathcal{L}_{ori}$
- 8: **if** $D_1 < 0$
- 9: Determine the optimal key area $BB_{optimal} = BB_{detection}$
- 10: Calculate policy gradient reward $R_t = e^{D_1}$
- 11: Calculate a detection loss $\mathcal{L}_{update} = \mathcal{L}_d(M_d(x_{key}), M_d(BB_{optimal})) \cdot R_t$
- 12: Update M_d with the perform gradient descent update $\theta_d \leftarrow \theta_d - \alpha_d \nabla_{\theta_d} \mathcal{L}_{update}$
- 13: Update M_c with the classification network $\theta_c \leftarrow \theta_c - \alpha_c \nabla_{\theta_c} \mathcal{L}_{update}$
- 14: **else**
- 15: Expand the key area by α times and shrink the key area by α times to get the classification \mathcal{L}_{expand} and \mathcal{L}_{shrink} , respectively
- 16: Calculate a new classification loss $\mathcal{L}_{new} = \text{minimum}(\mathcal{L}_{expand}, \mathcal{L}_{shrink})$
- 17: Choose the action and determine the optimal key area $BB_{optimal} = \{BB_{expand}, BB_{shrink}\}$
- 18: Calculate a new difference $D_2 = \lambda \cdot (\mathcal{L}_{new} - \mathcal{L}_{aug}) + (1 - \lambda) \cdot D_1$
- 19: **if** $D_2 < 0$
- 20: Calculate policy gradient reward $R_t = e^{D_2}$
- 21: **else**
- 22: Calculate policy gradient reward $R_t = D_2$
- 23: **end if**
- 24: Calculate a new detection loss $\mathcal{L}_{update} = \mathcal{L}_d(M_d(x_{key}), M_d(BB_{optimal})) \cdot R_t$
- 25: Update M_d with the perform gradient descent update $\theta_d \leftarrow \theta_d - \alpha_d \nabla_{\theta_d} \mathcal{L}_{update}$
- 26: Update M_c with the classification network by the optimal key area $\theta_c \leftarrow \theta_c - \alpha_c \nabla_{\theta_c} \mathcal{L}_{new}$
- 27: **end if**
- 28: **end for**

where $(1 - p(x_i))^\gamma$ is a modulating factor with tunable focusing parameter $\gamma \geq 0$. By using pixel-wise focal loss, we can preserve spatial information of the deviation of $P(a_i | s_i)$ from a_i . We then update θ_d by computing $\nabla_{\theta_d} L$. Thus, the classification network parameters θ_c are

updated with gradient descent by using the cross-entropy loss $\mathcal{L}_{CE} = -\sum_{i=1}^N p(x_i) \log(q(x_i))$ between the classification of the key area samples and the original target labels. In our experiments, we perform stochastic gradient update for both θ_d and θ_c at each batch step.

Table 1

Datasets information. BP: bacterial pneumonia, VP: viral pneumonia, NP: normal, and MP: manual process.

Datasets	Images	BP	VP	NP	MP	Ages	Size	Source
Ped-Pneumonia	5856	2780	1493	1583	✓	1–5	[384, 2916] × [127, 2713]	Guangzhou Women and Children's Medical Center
RSNA-Pneumonia	3811		1962	1849	✗	All-age	1024 × 1024	National Institutes of Health Clinical Center
Ours	3742	1088	295	2359	✗	0–14	[512, 3408] × [512, 3032]	Hainan Women and Children's Medical Center

3.2. Task-oriented homogenized automatic pre-processing

The implementation of the overall HAPre-KRS algorithm is summarized in Algorithm 1. The design of the reward is crucial to the convergence of the detection network. Using the change in training loss as a reward, as is done in neural architecture search [21], results in a weak reward signal hardly discernible from the expected changes in loss during training. Similarly, approximating rewards with a critic network introduces unnecessary overhead and slows down convergence. We propose a stable reward R_t referring to [13].

We first feed the original image to the classification network to get the original classification loss \mathcal{L}_{ori} and then feed it to the detection network to get the key area and the augmented classification loss \mathcal{L}_{aug} . After that, the difference D_1 between the two classification losses is calculated, which is defined as:

$$D_1 = \mathcal{L}_{aug} - \mathcal{L}_{ori}. \quad (5)$$

Finally, we design a comprehensive feedback reward mechanism that can be divided into two updated strategies in detail according to the difference. The first update strategy is that the difference D_1 between \mathcal{L}_{ori} and \mathcal{L}_{aug} is less than 0, which means that the key area obtained by the detection network is effective. We directly feed this key area and the detection loss back to the detection network and classification network to continue training our model. At this time, it is necessary to give the model a reward as positive feedback, so that the detection network does not need to make major changes; but D_1 is less than 0, and direct feedback to the model will make the optimization goal of detection network confusing. Therefore, we need to ensure that (i) the loss weight of the detection network cannot be a negative number when D_1 is less than 0, and (ii) the smaller the D_1 , the smaller the weight, that is, the smaller the penalty for the detection network. We find that using an exponential function is the most efficient; the policy gradient reward function is defined as:

$$R = e^{D_1}. \quad (6)$$

Otherwise, if the difference D_1 is at least 0, then the key area obtained by the detection network is not optimal, and another update strategy is required. First, we expand the key area by α times and shrink it by α times simultaneously to get a new optimal key area. Specifically, we input them, i.e., separately expanded key area and shrank key area, into the classification network again to obtain \mathcal{L}_{expand} and \mathcal{L}_{shrink} . The smaller between the two is the new key area and new enhanced classification loss \mathcal{L}_{new} . Then, the new difference D_2 between the new enhanced classification loss and the original enhanced classification loss, plus the weight of the original difference to get the new reward. Finally, the new classification loss \mathcal{L}_{new} and the new difference D_2 are respectively returned to the classification network and the detection network to continue training our model. Similarly, since the updated key area is effective, this difference D_2 is also less than 0. To ensure positive feedback to the model, if D_2 is less than 0, an exponential function is also used to define the reward function, otherwise, it is direct feedback to the model. Therefore, the final difference and the reward function are formulated as:

$$D_2 = \lambda(\min(\mathcal{L}_{expand}, \mathcal{L}_{shrink}) - \mathcal{L}_{aug}) + (1 - \lambda)D_1, \quad (7)$$

$$R = \begin{cases} e^{D_2} & \text{if } D_2 < 0, \\ D_2 & \text{else } D_2 \geq 0. \end{cases} \quad (8)$$

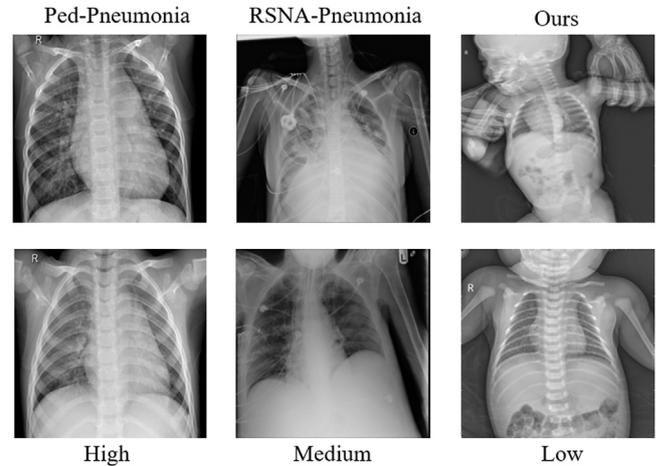


Fig. 4. The two images in the first column are examples of pre-processed pediatric pneumonia X-ray images in the Ped-Pneumonia dataset, whose image quality is high. Those in the second column are examples of the RSNA-Pneumonia dataset, whose quality is lower than Ped-Pneumonia but still better than ours (i.e., with medium quality). The images in the third column are examples of our dataset, which is not pre-processed and with low quality.

The resulting reward signal is related to key area quality, rather than reflecting the stochasticity in the training of the classification network.

The advantages of our model are as follows: (i) We do not need any labels outside the classification labels. (ii) We use the key area as input to the classification network, which can pay attention to important information while filtering irrelevant information. This is better than directly adding an attention module, because (i) the soft attention is not robust and may focus on other interfering information; (ii) pneumonia infects multiple regions of the terminal airway, alveoli, and interstitium, and our approach enables the classification network to enforce a focus on the key area; and (iii) our entire training process is online, which can improve the accuracy of both detection network and classification network to ensure homogeneous automatic pre-processing.

4. Experiments

Extensive experiments have been conducted to evaluate the proposed DRL-HAPre framework by comparing the HAPre-KRS model with the state-of-the-art baselines on two public datasets (one is well pre-processed and the other is without pre-processing) and a pediatric pneumonia dataset collected by ourselves.

4.1. Datasets

Three pediatric pneumonia X-ray image datasets are used in this work, where 70% of the contained images are used as training set, 10% as validation set, and 20% for testing. The detailed information of these three datasets is presented in the followings, the statistical results are summarized in Table 1, and some examples are shown in Fig. 4.

Ped-Pneumonia [3] is a public pediatric pneumonia dataset of children aged 0–5 from Guangzhou Women and Children's Medical Center.

This dataset contains 1583 normal images, 2780 bacteria pneumonia, and 1493 viruses pneumonia, with widths ranging from 384 to 2916 and heights ranging from 127 to 2713. It has been pre-processed by professionals; so, as shown in Fig. 4, its images are with less noise, higher contrast, and a more uniform distribution of types than another two datasets. Consequently, we use it to validate the performances of our method and the baselines on **high quality** data. However, the pre-processing needs assistance from clinical experts, so it is time-consuming and labor-intensive and will encounter the hard to deploy quickly and hard to reproduce problems.

RSNA Pneumonia Detection Challenge¹ (RSNA-Pneumonia) is a public pneumonia dataset from the National Institutes of Health Clinical Center. It contains 26,684 images of size 1024×1024 (i.e., 6012 pneumonia images and 20,672 normal ones). This dataset is a hybrid dataset, where some of the images are well-pre-processed while some others are without or with limited pre-processing; As Shown in Fig. 4, the latter images contain some interference so its image quality is less than Ped-Pneumonia dataset but still better than the dataset collected by ourselves. Consequently, we randomly select 3811 images (which are without or with limited pre-processing) as the **medium quality** dataset in our experiments.

Our dataset is a real pediatric pneumonia dataset of children aged 0 – 14 collected by ourselves from July 2016 to September 2021 at Women and Children’s Medical Center. It contains 2359 normal images, 1088 bacteria pneumonia, and 295 viruses pneumonia, with widths ranging from 512 to 3408, and heights from 512 to 3032. We do not conduct any pre-processing on this dataset, so the images contain a lot of interference information (e.g., the doctor’s hand to immobilize children), which are exactly the real images obtained in daily clinical practices. As shown in Fig. 4, this dataset has the lowest image quality, So we use it as the **low quality** dataset in our work.

4.2. Experimental settings

Our experiments are implemented using PyTorch and run on a NVIDIA GeForce GTX 2080Ti GPU and evaluated on the above three datasets using the same experimental settings. The implementation details of the proposed HAPre-KRS are as follows. The model is optimized using Adam [22] as the optimizer, where the learning rates of the classification network and detection agent network are set to $1e-5$ and $1e-6$, respectively, and the batch size is set to 1. All images are scaled to 640×640 , and the scaled images are randomly flipped vertically with a probability of 0.5.

To show the effectiveness of our models, we use the widely adopted accuracy (Acc), precision (Pre), recall (Rec), and F1 score (F1) as the evaluation metrics. Specifically, accuracy measures the percentage of correctly classified samples out of the total number of samples. Precision is the percentage of true positive samples of all predicted positive samples. Recall measures the probability that positive samples are correctly classified as positive. F1 score is the harmonic mean of PPV and Sensi, which thus can evaluate the model’s performances more comprehensively from the perspectives of both precision and recall. The higher the values of these metrics, the better the performance. Formally,

$$Acc = \frac{TP + TN}{TP + FP + TN + FN}, \quad (9)$$

$$Pre = \frac{TP}{TP + FP}, \quad (10)$$

$$Rec = \frac{TP}{TP + FN}, \quad (11)$$

$$F1 = \frac{2 * (Pre * Rec)}{(Pre + Rec)} = \frac{2 * TP}{2TP + FP + FN}, \quad (12)$$

where TP , TN , FP , and FN denote true positive (the number of positive samples that are correctly classified as positive), true negative (the number of negative samples that are correctly classified as negative), false positive (the number of negative samples that are incorrectly classified as positive), and false negative (the number of positive samples that are incorrectly classified as negative), respectively.

4.3. Baselines

In order to show the superior performances of the proposed DRL-HAPre framework achieved by homogenized automatic pre-processing, we compare the designed HAPre-KRS model with three kinds of state-of-the-art baselines, which are closely related to this work, i.e., pre-processing, attention, and pneumonia classification baselines. For a fair comparison, ResNet [23] is used as the backbone of our proposed DRL-HAPre and the pre-processing and attention-based baselines.

Specifically, since our work proposes to use deep reinforcement learning to overcome the existing two problems of pre-processing solutions and achieve homogenized automatic pre-processing, we first compare our work with five state-of-the-art deep learning or reinforcement learning based pre-processing solutions. (i) **Random Aug**: it is a pre-processing strategy that randomly applies various commonly used data augmentation methods to each image. (ii) **Cutout** [24]: this work conducts a pre-processing to randomly cut out parts of the sample, where the pixel values are set as 0. (iii) **Cutmix** [25]: it is an advanced version of Cutout, which also randomly cuts out parts of a given sample but randomly fills the cut area with pixel values of the same area in another data in the training set instead of assigning the pixel values as 0. (iv) **Mixup** [26]: it proposes to mix two random samples according to a pre-defined ratio (i.e., add proportionally) in the pre-processing. (v) **DRL-DA** [19]: it is a deep reinforcement learning based data augmentation solution, where DRL is used to explore the most effective sequence of pre-processing operations for each image.

Furthermore, since the functionality of key region selection pre-processing in our proposed HAPre-KRS is very similar to the widely used attention mechanism (i.e., make the learning of deep models focus on some key areas), we further compare our work with two state-of-the-art attention methods, CBAM [27] and DANet [28]. **CBAM** uses a convolutional block attention module to sequentially infer channel and spatial attention maps, which are then multiplied to the input feature map for adaptive feature refinement. **DANets** are dual attention networks that are proposed to combine parallel position and channel attention modules with the traditional dilated FCN to respectively model the semantic inter-dependencies in spatial and channel dimensions.

Finally, as the experiments are performed and the results are evaluated on the pediatric pneumonia classification tasks, three state-of-the-art deep learning based pneumonia classification solutions, DRE-Net [29], Two-Stage [30], and MAG-SD [31], are also selected as the baselines. (i) **DRE-Net** is a multi-scale based classification model, which introduces feature pyramid network [32] to extract features with different scales and combine the multi-scale features to improve the performances of pneumonia classification tasks. (ii) **Two-Stage** uses a detection network to focus on important areas first and then perform fine-grained classification tasks. (iii) **MAG-SD** is designed to resolve the problem of data imbalance, which uses multi-scale and attention techniques to achieve attention guided augmentations and obtain a new prediction loss based on a soft distance regularization.

4.4. Main results

Table 2 shows the results of the proposed HAPre-KRS model and the state-of-the-art pre-processing, attention, and pneumonia classification baselines on three pediatric pneumonia X-ray image datasets, Ped-Pneumonia, RSNA-Pneumonia, and our dataset, where the medical images are with respectively high, medium and low quality.

¹ Link: <https://www.kaggle.com/c/rsna-pneumonia-detection-challenge>.

Table 2

The results (in %) of applying the proposed HAPre-KRS and the state-of-the-art baselines on three pediatric pneumonia classification tasks on three different datasets, where the best result is bold, the second best one is underlined, and High, medium, and Low represents the data quality of the corresponding databases.

Model	Ped-Pneumonia (High)				RSNA-Pneumonia (Medium)				Ours (Low)			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
ResNet-18 [23]	88.30	82.33	82.67	82.40	74.93	74.47	75.04	74.78	84.92	65.01	65.00	64.53
Random Aug	88.38	82.35	82.68	82.43	75.01	74.51	74.72	74.55	85.03	63.01	64.14	63.22
Cutout [24]	88.46	82.44	82.61	82.46	75.11	74.68	74.59	74.63	84.80	62.91	64.24	63.30
Cutmix [25]	88.41	82.38	82.70	82.47	75.07	74.53	74.80	74.59	84.91	62.88	64.52	63.27
Mixup [26]	88.44	82.37	82.65	82.41	74.98	74.51	74.77	74.58	84.74	62.21	64.30	63.13
DRL-DA [19]	88.48	82.40	82.71	82.50	75.27	75.57	74.82	74.68	84.08	62.29	63.26	62.53
CBAM [27]	88.50	81.63	<u>82.91</u>	82.53	75.24	75.51	75.20	74.94	85.25	62.51	<u>66.26</u>	63.68
DANet [28]	88.53	82.48	82.69	82.63	75.63	75.70	76.62	75.15	85.17	65.48	64.91	65.11
DRE-Net [29]	87.82	81.83	82.36	82.04	75.40	<u>76.72</u>	75.04	75.86	85.25	64.78	64.23	64.46
Two-Stage [30]	88.43	82.21	82.47	82.29	75.66	74.85	75.52	75.22	<u>85.30</u>	63.77	65.73	64.33
MAG-SD [31]	88.47	<u>82.53</u>	82.73	<u>82.64</u>	<u>77.05</u>	76.64	<u>78.28</u>	<u>77.32</u>	85.13	<u>65.53</u>	65.85	<u>65.37</u>
HAPre-KRS	88.84	82.57	83.11	82.79	78.22	79.36	79.85	79.50	86.48	68.05	69.26	68.34
Improvement	0.31	0.04	0.20	0.15	1.17	2.64	1.57	2.18	1.18	2.70	3.00	2.97

As shown in Table 2, we can first observe that the performances of HAPre-KRS and the baselines are all very different among the three datasets. For example, the F1-scores of the 11 baselines are all around 0.82 on the well-preprocessed high-quality dataset, Ped-Pneumonia; however, when these methods are applied on another two datasets with lower image quality (without pre-processing), their F1-scores dramatically decrease to around 0.75 and 0.64 on RSNA-Pneumonia and our dataset, respectively. This finding proves the existence of the hard-to-reproduce problem in clinical practices, i.e., models with good performances in a medical institution are not guaranteed to achieve satisfactory performances often in the same medical image analysis task of other medical institutions, because the lower medical image quality, the worse the deep model's performances.

Furthermore, we also find that conducting the key region selection pre-processing can greatly increase the ResNet model's performances on RSNA-Pneumonia (from 0.7478 to 0.7950 in F1) and our dataset (from 0.6453 to 0.6834 in F1), while the increase on Ped-Pneumonia is very marginal (only from 0.8240 to 0.8279 in F1). We believe this is because the key region selection pre-processing effectively increases the quality of medical images in RSNA-Pneumonia and our dataset, which thus makes it easier for the deep model to learn classification features from the images, however, as the images in Ped-Pneumonia have been well-preprocessed, introduce additional pre-processing will not significantly enhance the image quality and also the deep model's performances. Consequently, this sufficiently demonstrates that having different medical image qualities in different medical institutes is an important reason (but certainly not the only reason) for the existence of the hard-to-reproduce problem in clinical practices, so it is essential to propose a homogenized automatic pre-processing method.

Third, HAPre-KRS greatly outperforms all the state-of-the-art baselines in terms of all evaluation metrics on all three datasets. For example, the F1-scores of the best SOTA pre-processing baseline, DRL-DA, the best SOTA attention baseline, DANet, and the best SOTA pneumonia classification baseline, MAG-SD, are 0.7468 (resp., 0.8250 and 0.6253), 0.7515 (resp., 0.8263 and 0.6511), and 0.7732 (resp., 0.8264 and 0.6537), respectively, on RSNA-Pneumonia (resp., Ped-Pneumonia and our dataset), while that of HAPre-KRS is 0.7918 (resp., 0.8287 and 0.6768). Please also note that the superior performances are achieved by introducing only one pre-processing operation, i.e., key region selection pre-processing, into the auxiliary network, so the performances of the proposed DRL-HAPre framework actually can be easily further enhanced by adding more potential pre-processing operations (which is one of the interesting future works). These findings thus prove the superior performances of HAPre-KRS and the proposed DRL-HAPre framework, i.e., using task-oriented homogenized automatic pre-processing can effectively improve the feature learning ability of the corresponding deep models.

Finally, we also find that compared to the best state-of-the-art baselines (whose results are underlined), HAPre-KRS can achieve significant improvements on RSNA-Pneumonia and our dataset (e.g., the improvements w.r.t. the best baselines in F1-score are 0.0218 and 0.0297, respectively), but it only has marginal improvements on Ped-Pneumonia (the improvement of F1-score is only 0.0015). The reason is as follows: the performance improvement of HAPre-KRS is achieved by using automatic homogenized pre-processing to improve the quality of medical images; since the images in Ped-Pneumonia have already been pre-processed and with high-quality, using HAPre-KRS can only bring limited image quality improvement, and thus the performance improvement is also limited; differently, the images in RSNA-Pneumonia and our dataset are without pre-processing and with relatively low quality, so there is a lot of space for HAPre-KRS to improve the image quality and also the performances. This finding proves that the lower the qualities of the medical images, the greater the improvements of using our HAPre-KRS model and DRL-HAPre framework; so the proposed DRL-HAPre is a more superior choice for deep learning based medical image analysis with data that is not pre-processed.

4.5. Effects on cross-source applications

Since it is usually not easy to obtain sufficient medical images in a short period, in order to enhance the landing efficiency of the CAD systems in clinical practices, cross-source application of CAD models or systems is adopted, i.e., using a large dataset that is obtained in a medical institute to train a medical image analysis model or system, and then apply the well-trained model or system on some similar tasks based on the same kind of medical images in another medical institute. Due to the different image qualities in the different medical institutes, the cross-source application will usually result in the performance drop of the deep models; however, we believe that by using task-oriented homogenized automatic pre-processing, our DRL-HAPre framework can narrow the quality differences of medical images from different sources, which thus remedy the performance degradation in the cross-source application of CAD systems. Consequently, additional experiments are conducted to validate the superior performances of the proposed HAPre-KRS model and DRL-HAPre framework in the cross-source application, where the deep models ResNet and HAPre-KRS (using ResNet as the task network) are trained using the training sets of three datasets respectively and each of their resulting well-trained models is tested on all three testing sets. As shown in Table 3, this thus results in nine groups of results: three groups are trained and tested using the same datasets (in the diagonal of Table 3), while another six groups are trained and tested using different datasets. To show the stability and scalability of the proposed HAPre-KRS model and DRL-HAPre framework (i.e., their supremacy remains with different backbones),

Table 3

The results (in %) of cross-source applications (i.e., using the model trained on a data source to conduct predictions on other data sources), where “*” represents the corresponding ResNet with the specific depth, the improvements of same-source results (i.e., results obtained when the same datasets are used for both training and testing) are underlined, and the improvements of different-source results (i.e., results obtained by using medical images from different sources for training and testing) are in bold.

Train	Test Model	Ped-Pneumonia (High)				RSNA-Pneumonia (Medium)				Ours (Low)			
		Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
		Ped-Pneumonia											
	ResNet-18	88.30	82.33	82.67	82.40	70.11	69.51	68.37	68.90	72.43	59.58	59.38	59.72
	* + HAPre-KRS	88.84	82.57	83.11	82.79	73.24	72.55	71.59	71.78	77.74	61.60	62.48	61.89
	Improvement	<u>0.54</u>	<u>0.24</u>	<u>0.44</u>	<u>0.39</u>	3.13	3.04	3.22	2.88	5.31	2.02	3.10	2.17
	ResNet-34	87.49	82.64	81.51	81.63	70.62	69.35	67.82	68.25	72.55	59.87	59.04	59.13
	* + HAPre-KRS	88.32	82.80	82.08	81.93	73.73	71.66	71.02	71.23	78.37	62.43	60.80	61.77
	Improvement	<u>0.83</u>	<u>0.16</u>	<u>0.57</u>	<u>0.30</u>	3.11	2.31	3.20	2.98	5.82	2.56	1.76	2.64
	ResNet-50	88.92	81.95	82.92	82.29	70.77	70.42	68.73	69.29	73.48	60.56	61.08	60.49
	* + HAPre-KRS	89.69	83.16	83.40	82.87	73.33	72.33	71.92	72.19	80.24	61.55	62.11	61.66
	Improvement	<u>0.77</u>	<u>1.21</u>	<u>0.48</u>	<u>0.58</u>	2.56	1.91	3.19	2.9	6.76	0.99	1.03	1.17
RSNA-Pneumonia													
	ResNet-18	70.78	65.44	65.15	64.73	74.93	74.47	75.04	74.78	70.67	60.33	61.25	60.78
	* + HAPre-KRS	87.55	86.73	86.18	86.29	78.22	79.36	79.85	79.50	79.97	70.12	69.77	69.95
	Improvement	16.77	21.29	21.03	21.56	<u>3.29</u>	<u>4.89</u>	<u>4.81</u>	<u>4.72</u>	9.3	9.79	8.52	9.17
	ResNet-34	71.33	64.63	65.40	64.88	74.54	75.04	76.78	74.22	71.33	61.46	60.06	60.33
	* + HAPre-KRS	87.15	85.83	86.35	85.97	79.53	78.71	79.42	78.97	80.12	70.44	70.16	70.25
	Improvement	15.82	21.2	20.95	21.09	<u>4.99</u>	<u>3.67</u>	<u>2.64</u>	<u>4.75</u>	8.79	8.98	10.1	9.92
	ResNet-50	71.82	66.60	65.54	65.71	74.67	74.50	74.75	74.53	71.67	61.54	61.08	61.31
	* + HAPre-KRS	88.52	87.46	88.27	87.70	79.87	78.08	80.46	79.18	80.88	71.45	70.42	70.74
	Improvement	16.7	20.86	22.73	21.99	<u>5.2</u>	<u>3.58</u>	<u>5.71</u>	<u>4.65</u>	9.21	9.91	9.34	9.43
Ours													
	ResNet-18	65.12	59.25	61.67	60.17	61.25	59.55	60.40	59.67	84.92	65.01	65.00	64.53
	* + HAPre-KRS	80.81	74.44	76.50	75.78	66.13	63.25	66.47	64.70	86.48	68.05	69.26	68.34
	Improvement	15.69	15.19	14.83	15.61	4.88	3.7	6.07	5.03	<u>1.56</u>	<u>3.04</u>	<u>4.26</u>	<u>3.81</u>
	ResNet-34	65.32	62.40	61.23	61.55	61.56	59.73	58.52	58.80	84.56	64.31	64.41	63.66
	* + HAPre-KRS	81.06	73.42	75.44	74.56	66.37	63.14	63.88	63.48	86.88	68.14	68.77	68.04
	Improvement	15.74	11.02	14.21	13.01	4.81	3.41	5.36	4.48	<u>2.32</u>	<u>3.83</u>	<u>4.36</u>	<u>4.38</u>
	ResNet-50	58.89	52.65	50.34	50.95	61.68	59.73	60.44	60.04	84.83	62.72	64.38	63.41
	* + HAPre-KRS	80.33	73.45	72.27	72.77	66.84	64.41	64.70	64.58	86.62	67.57	68.36	67.68
	Improvement	21.44	20.8	21.93	21.82	5.16	4.48	4.26	4.54	<u>1.79</u>	<u>4.85</u>	<u>3.98</u>	<u>4.27</u>

Table 4

The results (in %) of applying the proposed framework for key region selection pre-processing and using the state-of-the-art attention baselines in ResNet-based medical image classification tasks using various pneumonia image datasets, where “*” represents the corresponding ResNet with the specific depth, the best results in a given metric are bold, the second best ones are underlined, and Improvement represents the gaps between the best and the second best results.

Model	Ped-Pneumonia (High)				RSNA-Pneumonia (Medium)				Ours (Low)			
	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1	Acc	Pre	Rec	F1
ResNet-18	88.30	82.33	82.67	82.40	74.93	74.47	75.04	74.78	84.92	65.01	65.00	64.53
* + CBAM	88.50	81.63	<u>82.91</u>	82.53	75.24	75.51	75.20	74.94	<u>85.25</u>	62.51	<u>66.26</u>	63.68
* + DANet	<u>88.53</u>	<u>82.48</u>	82.69	<u>82.63</u>	<u>75.63</u>	<u>75.70</u>	<u>76.62</u>	<u>75.15</u>	85.17	<u>65.48</u>	64.91	<u>65.11</u>
* + HAPre-KRS	88.84	82.57	83.11	82.79	78.22	79.36	79.85	79.50	86.48	68.05	69.26	68.34
Improvement	0.31	0.09	0.20	0.16	2.59	3.66	3.23	4.35	1.23	2.57	3.00	3.23
ResNet-34	87.49	82.64	81.51	81.63	74.54	75.04	76.78	74.22	84.56	64.31	64.41	63.66
* + CBAM	<u>88.12</u>	82.53	<u>81.73</u>	<u>81.71</u>	74.98	75.33	77.25	75.84	<u>85.28</u>	64.67	<u>65.47</u>	64.09
* + DANet	87.91	<u>82.77</u>	81.16	81.68	<u>75.57</u>	<u>76.61</u>	<u>77.76</u>	<u>76.96</u>	85.20	<u>65.11</u>	64.94	<u>64.73</u>
* + HAPre-KRS	88.32	82.80	82.08	81.93	79.53	78.71	79.42	78.97	86.88	68.14	68.77	68.04
Improvement	0.20	0.03	0.35	0.22	3.96	2.10	1.66	2.01	1.60	3.03	3.30	3.31
ResNet-50	88.92	81.95	82.92	82.29	74.67	74.50	74.75	74.53	84.83	62.72	64.38	63.41
* + CBAM	88.72	<u>82.11</u>	82.82	82.44	75.66	75.47	75.58	75.18	84.95	<u>64.67</u>	63.89	<u>63.77</u>
* + DANet	<u>89.20</u>	81.93	<u>82.98</u>	<u>82.45</u>	<u>76.10</u>	<u>75.84</u>	<u>76.25</u>	<u>75.76</u>	85.37	63.18	<u>64.79</u>	63.57
* + HAPre-KRS	89.69	83.16	83.40	82.87	79.87	78.08	80.46	79.18	86.62	67.57	68.36	67.68
Improvement	0.49	1.05	0.42	0.42	3.77	2.24	4.21	3.42	1.25	2.90	3.57	3.91

we use three ResNet models with different depths (i.e., ResNet-18, ResNet-34, ResNet-50) in our experiments.

As shown in Table 3, we first observe that when a ResNet model is trained in a dataset, the results of using different datasets for testing (called different-source results) are always lower than those of using the same datasets for testing (called same-source results), even if the testing medical images belong to a dataset with higher image quality. For example, when the ResNet-50 model is trained using RSNA-Pneumonia, the different-source results obtained by using medical images in Ped-Pneumonia and our dataset for testing (e.g., F1-scores are 0.6571 and 0.6131, respectively) are dramatically lower than the same-source results obtained using RSNA-Pneumonia for testing (e.g., F1-score is

0.7453), even if the medical images in Ped-Pneumonia have much better quality than those in RSNA-Pneumonia; similarly, when ResNet-50 is trained by our dataset, the different-source results (e.g., F1-scores are 0.5059 and 0.6004) are also always lower than the same-source results (e.g., F1-score is 0.6341), despite the image quality in another two datasets is much better than that in our dataset. This thus proves again the existence of hard-to-reproduce problems in the application of deep medical image analysis models in clinical practices.

Then, we further find that using the proposed HAPre-KRS model not only enhances the performances of ResNet models in all cases, but more importantly, it can also remedy the performance degradation in the cross-source application. This is because, by using the

homogenized pre-processing to narrow the quality gaps of images from different sources, the improvements of using HAPre-KRS for the different-source results (highlighted in bold in Table 3) are always much higher than those for the same-source results (with underlines in Table 3). For example, when ResNet-50 and the corresponding HAPre-KRS model are trained on RSNA-Pneumonia, the F1-score improvement obtained for the same-source case is 0.0465, while those obtained for different-source cases using Ped-Pneumonia and our dataset for testing are 0.2199 and 0.0943, respectively; i.e., the latter two are several times higher than the former. Similarly, when ResNet-18 and the corresponding HAPre-KRS model are trained using Ped-Pneumonia (resp., our dataset), the F1-score improvement of the same-source case is only 0.0039 (resp., 0.0381), while those of different source cases are 0.0288 and 0.0217 (resp., 0.1561 and 0.0503). Consequently, we can assert that by introducing higher improvements for the different-source results than the same-source results, our work can remedy the hard-to-reproduce problem in the practical cross-source application.

In addition, an interesting phenomenon is noted in Table 3 that with the help of the HAPre-KRS model, the cross-source application of deep CAD models can sometimes even obtain better performances than the corresponding same-source application. For example, when using ResNet-34 as the backbone, the F1-score of HAPre-KRS that is trained by RSNA-Pneumonia and tested by Ped-Pneumonia (i.e., 0.8597) is much higher than that uses Ped-Pneumonia for both training and testing (i.e., 0.8193); similarly, the ResNet-34 based F1-score of HAPre-KRS that is trained by RSNA-Pneumonia and tested by our dataset (0.7025) is much higher than that uses our dataset for both training and testing (i.e., 0.6804). This means that the HAPre-KRS model is trained using RSNA-Pneumonia and used in the same kind of medical image analysis tasks in other medical institutes, e.g., Ped-Pneumonia and our dataset, it not only will not have performance degradation but can even achieve better performances than training using the medical images that are from the same sources as the testing. Consequently, this finding demonstrates that the proposed HAPre-KRS model and DRL-HAPre framework not only reduce the performance degradation in the practical cross-source application, but it can also even help the cross-source application of CAD systems to sometimes achieve much better performances than training and testing the CAD systems using medical images from the same medical institute (i.e., same-source application of CAD), which significantly thus enhance the landing efficiency and effectiveness of CAD systems in real-world practices.

4.6. Compared to existing attention mechanisms

As the functionality of key region selection pre-processing in our proposed HAPre-KRS is very similar to the widely used attention mechanism (i.e., make the learning of deep models focus on some key areas), to show the effectiveness of the proposed DRL-HAPre framework more comprehensively, we further compare the performances of HAPre-KRS and the state-of-the-art attention methods (CBAM and DANets) using ResNet models with three different depths, i.e., ResNet-18, ResNet-34, and ResNet-50, on three pediatric pneumonia classification tasks.

Generally, as shown in Table 4, we observe that the performances of HAPre-KRS and the state-of-the-art attention baselines are always better than those of the corresponding ResNet models in all cases. This states that guiding the model to pay attention to the key areas is an effective way to improve the deep models' learning capabilities. Furthermore, we also find that the performances of HAPre-KRS and the attention baselines on Ped-Pneumonia is generally better than those in RSNA-Pneumonia and our dataset, this is because the higher the medical image quality, the better the deep model's performances. Finally, we note that, regardless of the depth of ResNet, HAPre-KRS always outperforms CBAM and DANet, and comparing to the second best results, the improvement gaps of HAPre-KRS on RSNA-Pneumonia and our datasets (e.g., ResNet-18 based improvements are respectively 0.0435 and 0.0323 in F1-score), i.e., the ones with medium or low image

Table 5

The results (in %) of the state-of-the-art supervised classification backbone models on our dataset, where Params represents the number of parameters in the corresponding model. The best results in the corresponding metrics are bold.

Model	Acc	Pre	Rec	F1	Params
ResNet-18	84.92	65.01	65.00	64.53	11.18M
ResNet-34	84.56	64.31	64.41	63.66	21.29M
ResNet-50	84.83	62.72	64.38	63.41	25.55M
DenseNet-121	84.53	60.18	61.42	60.06	6.87M
DenseNet-169	83.48	59.05	61.16	59.95	12.33M
VGG-16	84.58	60.87	66.22	62.53	134.27M
VGG-19	86.18	60.91	67.94	63.32	139.58M

quality, are always larger than those on Ped-Pneumonia (e.g., ResNet-18 based F1-score improvement is only 0.0016), i.e., the one with high image quality. This proves the effectiveness of HAPre-KRS more comprehensively: by conducting homogenized key region selection pre-processing, HAPre-KRS not only provides a more accurate key area for feature learning than the SOTA attention baselines, but it also greatly enhances the quality of medical images in RSNA-Pneumonia and our dataset and thus introduces significant performance improvements on these datasets, while the images in Ped-Pneumonia have been well-preprocessed so the enhancement is limited. In addition, since the HAPre-KRS model is a relatively simple implementation of the proposed DRL-HAPre framework in the pediatric pneumonia classification tasks, the effectiveness of HAPre-KRS also proves the effectiveness of our DRL-HAPre framework in achieving painless and accurate medical image analysis, where DRL-HAPre framework can be independently designed using appropriate auxiliary network and task network according to the corresponding medical image analysis tasks, and can also include various homogeneous pre-processing operations as needed.

We also show some visualized examples of HAPre-KRS and the attention baselines in Fig. 5 to exhibit the supremacy of HAPre-KRS that uses the task-oriented homogenized automatic pre-processing in localizing optimal key areas in pediatric pneumonia classification tasks, where four X-ray images from our dataset and their corresponding heatmaps generated by ResNet-18, HAPre-KRS and the attention baselines are shown. Specifically, the image shown at the first row of Fig. 5 is a medical image with a relatively obvious and clear lung area (i.e., images with relatively good quality); for this easy case, ResNet and ResNet with additional attention mechanisms can generally locate the lungs to a certain extent, but our model can locate the lung region more accurately. Furthermore, the examples shown at the second to forth rows of Fig. 5 are some medical images with low quality (i.e., the lung areas are either less obvious, with Small size, or containing some interference information); for these hard cases, ResNet and ResNet with additional attention mechanisms mistakenly pay attention to some interference areas, such as a doctor's hand, however, our model can still locate the interested lung region stably and accurately. Therefore, these visualized observations greatly demonstrate again that, by using the proposed task-oriented homogenized automatic pre-processing to dynamically adjust the detection network and classification network, HAPre-KRS is more effective than the state-of-the-art attention mechanisms in practical medical image classification tasks, especially for the images without pre-processing, because it can locate the key region, i.e., the lung area in this task, accurately and stably not only in easy cases but also in hard cases with low quality and interference information.

4.7. The selection of task network

Furthermore, we conduct additional experiments to illustrate the rationality of selecting ResNet as the task network (i.e., backbone) in HAPre-KRS. Table 5 shows the results of all the state-of-the-art supervised classification backbone models. As shown in Table 5, the ResNet models generally achieve the best performances, while the

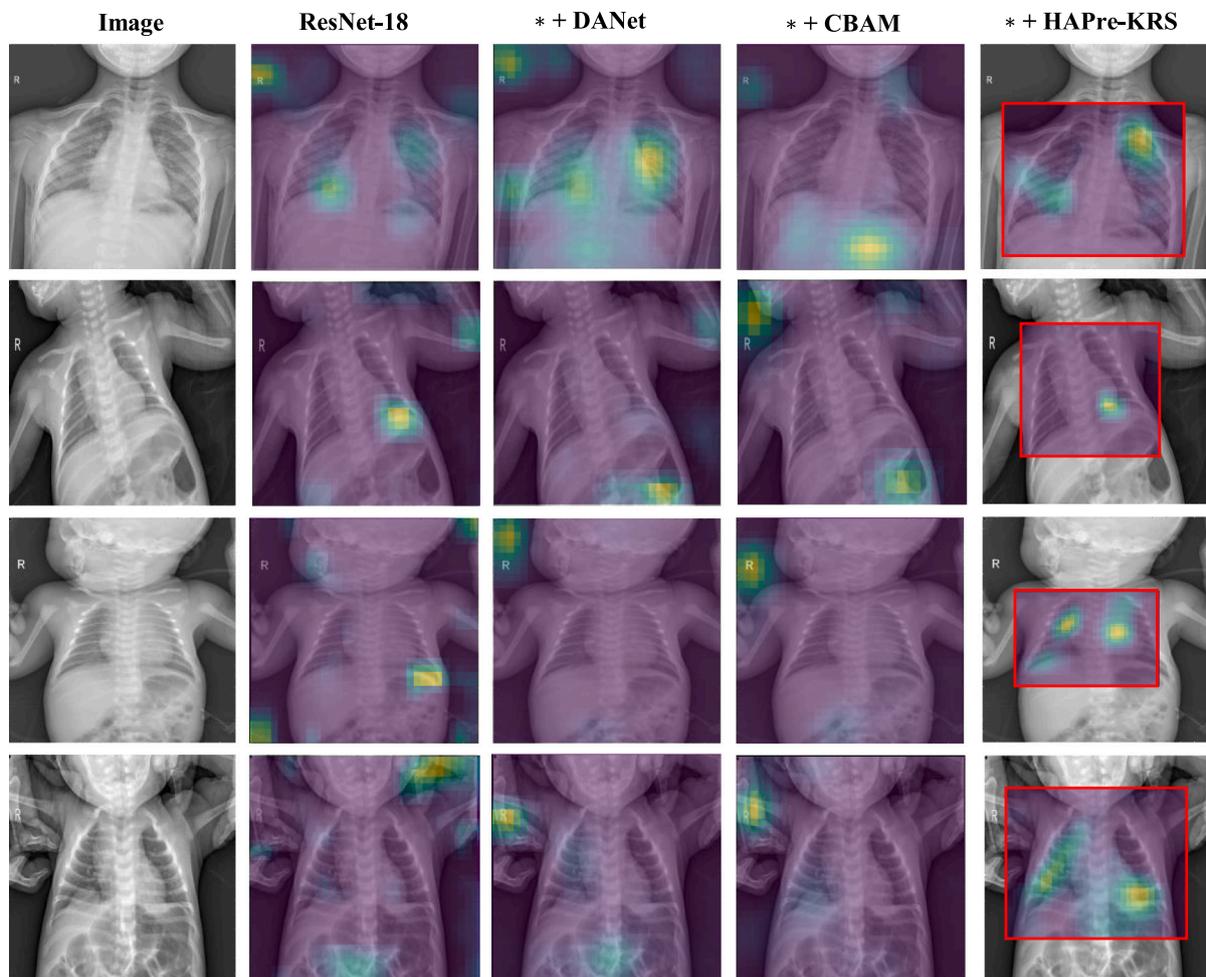


Fig. 5. Visualized examples of attention heatmaps of the proposed HAPre-KRS and the attention baselines. “*” represents the corresponding ResNet with the specific depth.

number of parameters used in the ResNet methods are relatively low; so we select the ResNet model as the task network (i.e., backbone) in the proposed HAPre-KRS.

5. Social impact for proposed framework

Medical image pre-processing is time-consuming and highly dependent on expert assistance, whose quality greatly affects the performances of subsequent deep-learning-based medical image analysis. However, there exist hard-to-deploy-quickly and hard-to-reproduce problems in the current medical image pre-processing solutions, which greatly limit the deployment efficiency and performances of intelligent medical image analysis models in clinical practices. Taking the pneumonia X-ray image classification as an example, the lesion areas in chest X-ray images of children are usually irregular, small, and contain interference information, such as motion artifacts and irrelevant objects (e.g., the doctor’s hands). Therefore, a commonly used pre-processing operation is to locate the key areas in the image and remove other irrelevant areas, i.e., key region selection. However, the pre-processing of key region selection heavily relies on experts’ experience in reading the chest X-ray images and accurately locating the key areas, which thus hinders the rapid deployment and cross-institutes application of the corresponding deep-learning-based medical image classification models. Therefore, by proposing the DRL-HAPre framework to automatically conduct the homogeneous pre-processing for medical image analysis models, our work can alleviate the hard-to-deploy-quickly and hard-to-reproduce problems, and help achieve painless and accurate medical image analysis in medical practices.

Consequently, besides the technical contributions, this work also brings great social benefits in the related research and clinical areas, e.g., accelerating the application process of the intelligent CAD system to significantly reduce the workload of doctors, and saves both time and money for patients.

In addition, by utilizing an auxiliary network to generate additional auxiliary information or signals and help the task network achieve better performances in a deep reinforced way, the proposed DRL-HAPre framework can not only be used to conduct task-oriented homogenized automatic pre-processing in pediatric pneumonia classification tasks, but also be used to boost the deep models’ performances in many closely-related application domains. First, DRL-HAPre can be used to improve the classification performance of other kinds of medical image classification tasks (e.g., COVID-19 classification [33], skin disease classification [34], etc.) in a similar way with similar or even more pre-processing operations. Second, medical image segmentation [35] is an important clinical task but it is a very time-consuming and labor-intensive task to obtain pixel-wise segmentation masks in practices, so DRL-HAPre can be further applied (with some simple modifications) to resolve this problem by using detection and classification models as the auxiliary network to provide bounding-box based weak supervision signals for the segmentation model based task network, which thus reduces the number of pixel-wise segmentation annotations needed for the deep segmentation model [36] to achieve satisfactory performances. Last but not least, DRL-HAPre can also be used to enhance the performances of medical image detection [37] using deep reinforcement learning; a potential way may be using a classification model as the auxiliary network to generate class activation maps to guide

the bounding-box selection in the deep detection model based task network. Consequently, the proposed DRL-HAPre can bring significant and beneficial impacts on these closely-related application domains.

6. Conclusion and future work

In this work, we first identified the existence of the hard-to-deploy-quickly and hard-to-reproduce problems in the pre-processing of the deep learning based medical image analysis. Then, a deep-reinforcement-learning task-oriented homogenized automatic pre-processing framework, DRL-HAPre was proposed to overcome these two problems and achieve painless (due to automation) and accurate (due to Homogenized quality enhancement) medical image analysis in clinical practices. Finally, based on the DRL-HAPre framework, a pre-processing model, HAPre-KRS, was further developed to achieve automatic and homogenized pre-processing of key region selection in the pneumonia image classification task.

Extensive experimental studies were conducted on three pediatric pneumonia classification datasets with different image qualities. The experimental results first proved the existence of the hard-to-reproduce problem and the fact that having different medical image qualities in different medical institutes is an important reason for the existence of a hard-to-reproduce problem in clinical practices, so it is compelling to propose homogenized automatic pre-processing method. Then, the results further exhibited that the proposed HAPre-KRS model and DRL-HAPre framework greatly outperform three kinds of state-of-the-art baselines (i.e., pre-processing, attention and pneumonia baseline), and the lower the medical image quality, the greater the improvements of using our HAPre-KRS model and DRL-HAPre framework. Finally, with the help of homogenized pre-processing, HAPre-KRS (and DRL-HAPre framework) can greatly avoid performance degradation in real-world cross-source applications and thus overcome the hard-to-reproduce problem.

As for the future scope of this work, although our experiments mainly focus on the automatic pre-processing of key region selection, we believe any pre-processing operations (e.g., the ones mentioned in [18]) that can reduce image interference information is suitable to be incorporated into the DRL-HAPre framework to achieve an automatic and homogenized medical image pre-processing. Furthermore, our framework not only can be applied in a single pre-processing operation but also be used to bring automation and homogenization to hybrid pre-processing tasks [19], where a combination of multiple pre-processing techniques are automatically and adaptively selected using DRL-HAPre framework to achieve optimal and homogenized pre-processing and then enhance the performances of subsequences medical image analysis tasks. Consequently, an interesting future research direction is to further explore and investigate the applications of the proposed framework in various homogenized automatic medical image (hybrid) pre-processing tasks in different clinical applications.

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 61906063, 62102265, and 62276089, by the Natural Science Foundation of Hebei Province, China, under the grant F2021202064, by the “100 Talents Plan” of Hebei Province, China, under the grant E2019050017, by the Open Research Fund from Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) under the grant GML-KF-22-29, by the Natural Science Foundation of Guangdong Province of China under the grant 2022A1515011474, by the Key Research and Development Project of Hainan Province, China, under the grant ZDYF2022SHFZ015, and by the Natural Science Foundation of Hainan Province, China, under the grant 821RC1131.

References

- [1] Jiayi Shen, Casper J.P. Zhang, Bangsheng Jiang, Jiebin Chen, Jian Song, Zherui Liu, Zonglin He, Sum Yi Wong, Po-Han Fang, Wai-Kit Ming, et al., Artificial intelligence versus clinicians in disease diagnosis: Systematic review, *JMIR Med. Inform.* 7 (3) (2019) e10010, <http://dx.doi.org/10.2196/10010>.
- [2] Zhenghua Xu, Shijie Liu, Di Yuan, Lei Wang, Junyang Chen, Thomas Lukasiewicz, Zhigang Fu, Rui Zhang, ω -Net: Dual supervised medical image segmentation with multi-dimensional self-attention and diversely-connected multi-scale convolution, *Neurocomputing* 500 (2022) 177–190, <http://dx.doi.org/10.1016/j.neucom.2022.05.053>.
- [3] Daniel Kermayn, Kang Zhang, Michael Goldbaum, et al., Labeled optical coherence tomography (OCT) and chest X-ray images for classification, *Mendeley Data* 2 (2) (2018) <http://dx.doi.org/10.17632/rscbjbr9sj.2>.
- [4] Ronald J. Williams, Simple statistical gradient-following algorithms for connectionist reinforcement learning, *Mach. Learn.* 8 (3) (1992) 229–256, <http://dx.doi.org/10.1007/BF00992696>.
- [5] Yun Yang, Yuanyuan Hu, Xingyi Zhang, Song Wang, Two-stage selective ensemble of CNN via deep tree training for medical image classification, *IEEE Trans. Cybern.* 52 (9) (2021) 9194–9207, <http://dx.doi.org/10.1109/TCYB.2021.3061147>.
- [6] T. Arumuga Maria Devi, V.I. Mebin Jose, Three stream network model for lung cancer classification in the CT images, *Open Comput. Sci.* 11 (1) (2021) 251–261, <http://dx.doi.org/10.1515/comp-2020-0145>.
- [7] Martin Zlocha, Qi Dou, Ben Glocker, Improving RetinaNet for CT lesion detection with dense masks from weak RECISt labels, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 402–410, http://dx.doi.org/10.1007/978-3-030-32226-7_45.
- [8] Tsung-Yi Lin, Priya Goyal, Ross Girshick, Kaiming He, Piotr Dollár, Focal loss for dense object detection, in: Proceedings of the IEEE International Conference on Computer Vision, 2017, pp. 2980–2988, <http://dx.doi.org/10.1109/ICCV.2017.324>.
- [9] Zixu Zhao, Huangjing Lin, Hao Chen, Pheng-Ann Heng, PFA-Scannet: Pyramidal feature aggregation with synergistic learning for breast cancer metastasis analysis, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 586–594, http://dx.doi.org/10.1007/978-3-030-32239-7_65.
- [10] Fisher Yu, Vladlen Koltun, Multi-scale context aggregation by dilated convolutions, 2015, arXiv preprint, [arXiv:1511.07122](https://arxiv.org/abs/1511.07122).
- [11] Yuhang Song, Jianyi Wang, Thomas Lukasiewicz, Zhenghua Xu, Shangdong Zhang, Andrzej Wojcicki, Mai Xu, Mega-reward: Achieving human-level play without extrinsic rewards, in: Proceedings of the AAAI Conference on Artificial Intelligence, Vol. 34, 2020, pp. 5826–5833, <http://dx.doi.org/10.1609/aaai.v34i04.6040>.
- [12] Mohamed Akrouf, Amir-massoud Farahmand, Tory Jarmain, Latif Abid, Improving skin condition classification with a visual symptom checker trained using reinforcement learning, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 549–557, http://dx.doi.org/10.1007/978-3-030-32251-9_60.
- [13] Kaiyang Cheng, Claudia Iriondo, Francesco Calivá, Justin Kroegue, Sharmila Majumdar, Valentina Pedoia, Adversarial policy gradient for deep learning image augmentation, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2019, pp. 450–458, http://dx.doi.org/10.1007/978-3-030-32226-7_50.
- [14] Jiarong Ye, Yuan Xue, L. Rodney Long, Sameer Antani, Zhiyun Xue, Keith C. Cheng, Xiaolei Huang, Synthetic sample selection via reinforcement learning, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 53–63, http://dx.doi.org/10.1007/978-3-030-59710-8_6.
- [15] Jian Wang, Juzheng Miao, Xin Yang, Rui Li, Guangquan Zhou, Yuhao Huang, Zehui Lin, Wufeng Xue, Xiaohong Jia, Jianqiao Zhou, et al., Auto-weighting for breast cancer classification in multimodal ultrasound, in: Proceedings of International Conference on Medical Image Computing and Computer-Assisted Intervention, 2020, pp. 190–199, http://dx.doi.org/10.1007/978-3-030-59725-2_19.
- [16] Bolei Xu, Jingxin Liu, Xianxu Hou, Bozhi Liu, Jon Garibaldi, Ian O. Ellis, Andy Green, Linlin Shen, Guoping Qiu, Attention by selection: A deep selective attention approach to breast cancer classification, *IEEE Trans. Med. Imaging* 39 (6) (2019) 1930–1941, <http://dx.doi.org/10.1109/TMI.2019.2962013>.
- [17] Alex Krizhevsky, Ilya Sutskever, Geoffrey E. Hinton, Imagenet classification with deep convolutional neural networks, *Commun. ACM* 60 (6) (2017) 84–90, <http://dx.doi.org/10.1145/3065386>.
- [18] Misha Laskin, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, Aravind Srinivas, Reinforcement learning with augmented data, 2020, arXiv preprint [arXiv:2004.14990](https://arxiv.org/abs/2004.14990).
- [19] Tiexin Qin, Ziyuan Wang, Kelei He, Yinghuan Shi, Yang Gao, Dinggang Shen, Automatic data augmentation via deep reinforcement learning for effective kidney tumor segmentation, in: Proceedings of IEEE International Conference on Acoustics, Speech and Signal Processing, 2020, pp. 1419–1423, <http://dx.doi.org/10.1109/ICASSP40776.2020.9053403>.

- [20] Ilya Kostrikov, Denis Yarats, Rob Fergus, Image augmentation is all you need: Regularizing deep reinforcement learning from pixels, 2020, arXiv preprint [arXiv:2004.13649](https://arxiv.org/abs/2004.13649).
- [21] Barret Zoph, Quoc V. Le, Neural architecture search with reinforcement learning, 2016, arXiv preprint, [arXiv:1611.01578](https://arxiv.org/abs/1611.01578).
- [22] Diederik P. Kingma, Jimmy Ba, Adam: A method for stochastic optimization, 2014, arXiv preprint [arXiv:1412.6980](https://arxiv.org/abs/1412.6980).
- [23] Kaiming He, Xiangyu Zhang, Shaoqing Ren, Jian Sun, Deep residual learning for image recognition, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016, pp. 770–778, [http://dx.doi.org/10.1109/CVPR.2016.90](https://dx.doi.org/10.1109/CVPR.2016.90).
- [24] Terrance DeVries, Graham W. Taylor, Improved regularization of convolutional neural networks with cutout, 2017, arXiv preprint [arXiv:1708.04552](https://arxiv.org/abs/1708.04552).
- [25] Sangdoon Yun, Dongyoon Han, Seong Joon Oh, Sanghyuk Chun, Junsuk Choe, Youngjoon Yoo, Cutmix: Regularization strategy to train strong classifiers with localizable features, in: Proceedings of the IEEE/CVF International Conference on Computer Vision, 2019, pp. 6023–6032, [http://dx.doi.org/10.1109/ICCV.2019.00612](https://dx.doi.org/10.1109/ICCV.2019.00612).
- [26] Hongyi Zhang, Moustapha Cisse, Yann N. Dauphin, David Lopez-Paz, mixup: Beyond empirical risk minimization, 2017, arXiv preprint [arXiv:1710.09412](https://arxiv.org/abs/1710.09412).
- [27] Sanghyun Woo, Jongchan Park, Joon-Young Lee, In So Kweon, CBAM: convolutional block attention module, in: Proceedings of the European Conference on Computer Vision, 2018, pp. 3–19, [http://dx.doi.org/10.1007/978-3-030-01234-2_1](https://dx.doi.org/10.1007/978-3-030-01234-2_1).
- [28] Jun Fu, Jing Liu, Haijie Tian, Yong Li, Yongjun Bao, Zhiwei Fang, Hanqing Lu, Dual attention network for scene segmentation, in: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019, pp. 3146–3154, [http://dx.doi.org/10.1109/CVPR.2019.00326](https://dx.doi.org/10.1109/CVPR.2019.00326).
- [29] Ying Song, Shuangjia Zheng, Liang Li, Xiang Zhang, Xiaodong Zhang, Ziwan Huang, Jianwen Chen, Ruixuan Wang, Huiying Zhao, Yutian Chong, et al., Deep learning enables accurate diagnosis of novel coronavirus (COVID-19) with CT images, *IEEE/ACM Trans. Comput. Biol. Bioinform.* 18 (6) (2021) 2775–2780, [http://dx.doi.org/10.1109/TCBB.2021.3065361](https://dx.doi.org/10.1109/TCBB.2021.3065361).
- [30] Shubham Chaudhary, Sadbhawna Sadbhawna, Vinit Jakhethiya, Badri N. Subudhi, Ujjwal Baid, Sharath Chandra Guntuku, Detecting covid-19 and community acquired pneumonia using chest ct scan images with deep learning, in: Proceedings of 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing, 2021, pp. 8583–8587, [http://dx.doi.org/10.1109/ICASSP39728.2021.9414007](https://dx.doi.org/10.1109/ICASSP39728.2021.9414007).
- [31] Jingxiong Li, Yaqi Wang, Shuai Wang, Jun Wang, Jun Liu, Qun Jin, Lingling Sun, Multiscale attention guided network for COVID-19 diagnosis using chest X-ray images, *IEEE J. Biomed. Health Inf.* 25 (5) (2021) 1336–1346, [http://dx.doi.org/10.1109/JBHI.2021.3058293](https://dx.doi.org/10.1109/JBHI.2021.3058293).
- [32] Tsung-Yi Lin, Piotr Dollár, Ross Girshick, Kaiming He, Bharath Hariharan, Serge Belongie, Feature pyramid networks for object detection, in: Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2017, pp. 2117–2125, [http://dx.doi.org/10.1109/CVPR.2017.106](https://dx.doi.org/10.1109/CVPR.2017.106).
- [33] Yuanzhe Li, FeiFei Guo, Yang Cao, LiFeng Li, YanJun Guo, Insight into COVID-19 for pediatricians, *Pediatr. Pulmonol.* 55 (5) (2020) E1–E4, [http://dx.doi.org/10.1002/ppul.24734](https://dx.doi.org/10.1002/ppul.24734).
- [34] Viswanatha Reddy Allugunti, A machine learning model for skin disease classification using convolution neural network, *Int. J. Comput. Program. Database Manage.* 3 (1) (2022) 141–147, [http://dx.doi.org/10.1002/ppul.24734](https://dx.doi.org/10.1002/ppul.24734).
- [35] Ailiang Qi, Dong Zhao, Fanhua Yu, Ali Asghar Heidari, Zongda Wu, Zhennao Cai, Fayadh Alenezi, Romany F Mansour, Huiling Chen, Mayun Chen, Directional mutation and crossover boosted ant colony optimization with application to COVID-19 X-ray image segmentation, *Comput. Biol. Med.* 148 (2022) 105810, [http://dx.doi.org/10.1016/j.combiomed.2022.105810](https://dx.doi.org/10.1016/j.combiomed.2022.105810).
- [36] Hang Su, Dong Zhao, Hela Elmannai, Ali Asghar Heidari, Sami Bourouis, Zongda Wu, Zhennao Cai, Wenyong Gui, Mayun Chen, Multilevel threshold image segmentation for COVID-19 chest radiography: A framework using horizontal and vertical multiverse optimization, *Comput. Biol. Med.* 146 (2022) 105618, [http://dx.doi.org/10.1016/j.combiomed.2022.105618](https://dx.doi.org/10.1016/j.combiomed.2022.105618).
- [37] Keli Hu, Liping Zhao, Sheng Feng, Shengdong Zhang, Qianwei Zhou, Xiaozhi Gao, Yanhui Guo, Colorectal polyp region extraction using saliency detection network with neutrosophic enhancement, *Comput. Biol. Med.* 147 (2022) 105760, [http://dx.doi.org/10.1016/j.combiomed.2022.105760](https://dx.doi.org/10.1016/j.combiomed.2022.105760).