

Article

A Reinforcement Learning Algorithm for Automated Detection of Skin Lesions

Usman Ahmad Usmani ^{1,*}, Junzo Watada ², Jafreezal Jaafar ¹, Izzatdin Abdul Aziz ¹  and Arunava Roy ³

¹ Department of Computer and Information Science, Faculty of Science and IT, Universiti Teknologi Petronas, UTP, Seri Iskandar 32610, Malaysia; jafreez@utp.edu.my (J.J.); izzatdin@utp.edu.my (I.A.A.)

² Production and Systems, Graduate School of Information, Waseda University, Wakamatsu, Kitakyushu 8080135, Japan; junzo.watada@gmail.com

³ Department of Computer Science, School of Information Technology, Monash University, Bandar Sunway, Subang Jaya 47500, Malaysia; royism.arunava@gmail.com

* Correspondence: usman_19001067@utp.edu.my

Abstract: Skin cancers are increasing at an alarming rate, and detection in the early stages is essential for advanced treatment. The current segmentation methods have limited labeling ability to the ground truth images due to the numerous noisy expert annotations present in the datasets. The precise boundary segmentation is essential to correctly locate and diagnose the various skin lesions. In this work, the lesion segmentation method is proposed as a Markov decision process. It is solved by training an agent to segment the region using a deep reinforcement-learning algorithm. Our method is similar to the delineation of a region of interest by the physicians. The agent follows a set of serial actions for the region delineation, and the action space is defined as a set of continuous action parameters. The segmentation model learns in continuous action space using the deep deterministic policy gradient algorithm. The proposed method enables continuous improvement in performance as we proceed from coarse segmentation results to finer results. Finally, our proposed model is evaluated on the International Skin Imaging Collaboration (ISIC) 2017 image dataset, Human against Machine (HAM10000), and PH² dataset. On the ISIC 2017 dataset, the algorithm achieves an accuracy of 96.33% for the naevus cases, 95.39% for the melanoma cases, and 94.27% for the seborrheic keratosis cases. The other metrics are evaluated on these datasets and rank higher when compared with the current state-of-the-art lesion segmentation algorithms.

Keywords: image segmentation; skin lesion; deep deterministic policy gradient; multistep manner; skin cancer; melanoma; semantic segmentation; instance segmentation

check for
updates

Citation: Usmani, U.A.; Watada, J.; Jaafar, J.; Aziz, I.A.; Roy, A. A Reinforcement Learning Algorithm for Automated Detection of Skin Lesions. *Appl. Sci.* **2021**, *11*, 9367. <https://doi.org/10.3390/app11209367>

Academic Editors: Cecilia Di Ruberto, Alessandro Stefano, Albert Comelli, Lorenzo Putzu and Andrea Loddo

Received: 3 July 2021

Accepted: 4 September 2021

Published: 9 October 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

The largest organ in the human body is the skin. The disorganized and uncontrolled growth of skin cells lead to skin cancer formation and cancer can rapidly grow to other body parts. Skin cancer is a common kind of cancer worldwide. The deadliest form of skin cancer is melanoma, and its prevalence has been rapidly rising in the last 30 years [1]. Early diagnosis can increase one's chances of survival. Identification of the melanoma or suspected skin lesions is conducted by dermoscopy imaging, by detecting the pigmented skin lesions. The technique is non-invasive and detects possible lesions in the early stage. Because of the higher resolution of dermoscopic images and better visualization capabilities, dermatologists can use their own eyes to examine skin lesions. The decision-making process is time-consuming, requires a high degree of expert knowledge, and is biased (i.e., depending on the dermatologist's viewpoint). Convolutional neural networks (CNN) can detect melanoma in the same manner that dermatologists can [2], suggesting the potential for automated skin lesion analysis.

Automated skin lesion analysis is an essential part of computer-assisted diagnosis [3,4]. Existing artificial intelligence (AI) algorithms do not do an excellent job of adequately considering this clinical frame of reference. The diagnostic accuracy (*Acc*) can increase if

the detection algorithms employ “contextual” images gathered within the same patient to analyze whether the images represent a melanoma [5]. If the process is successful, the classifiers will be more precise and assist dermatological clinic work. The ABCDE recommendations [6,7] provide physicians and patients with a clear and comprehensive framework for diagnosing melanoma, asymmetry, abnormalities along the border, skin discoloration, unbalanced color palette, lesions larger than 6 mm in diameter, evolved lesions (shape, color, or size), indicating the potential melanoma [8,9]. The uneven melanoma borders are usually notched or scalloped, and uneven margins are defined vaguely. Consequently, the lesion segmentation process is often given the first performance preference to provide the regions of interest (ROI) or boundary information that has proven to aid the detection and classification tasks [10,11]. Still, the issue of automatic lesion segmentation remains unsolved.

In some skin lesions with light pigment, the color and the visual patterns of the pigment patches and the surrounding skin regions are highly similar, resulting in fuzzy and unclear boundaries, making the segmentation extremely difficult. Figure 1 shows the skin sample lesion images taken from the ISIC 2017 dataset [12], which have distinct class differences due to the various variations such as severe, blurred boundary, inhomogeneity, hair appearance, etc. Furthermore, due to the high quality of the original dermoscopic images, direct processing takes a long time and requires many resources. Therefore, to reduce the image’s size, downsampling is utilized first. The subtlety and finer textures are lost during this process, making it much more difficult to distinguish the boundaries of the lesion. The lesions include color-makers, veins, hairs, glues, and rulers, all of which change the texture and color distribution of the lesions and impede efficient learning. Because the images are collected from various institutions and hospitals, they have a diverse set of characteristics. Hairs and color-makers are seen in specific images, complicating the process of segmenting lesions. The lesion segmentation is challenging due to the problems mentioned above. Several multi-scale information in CNN architectures [13–15] or multi-task learning frameworks [16,17] are mentioned in the literature to address these problems. The fundamental idea behind the methods is to generate reliable predictions by using as much data as possible. On the other hand, these methods either incorporate detailed new training parameters or need additional labeling information, both of which are inapplicable in practice.

However, in the methods using fully convolutional networks (FCNs) [18,19], the FCN smoothens out the complicated structures while ignoring the minor details. In the stand-alone post-processing, step CRF is employed and is disconnected from the FCN. A downsampling sequence followed by a greater sampling rate rapidly results in a detailed loss and poor segmentation results. The network makes errors while dealing with oversized objects. Despite their large representative capacity, the multi-scale approach uses skip connections from the low-level encoder features that tend to utilize redundant data. Furthermore, the encoder feature’s contextual information is insufficient at the start of the network when coupled with the proper high-level decoder feature map, resulting in poor pixel-wise recognition performance. A series of downsampling and a higher sampling rate leads to loss of details and rough lesion segmentation results. When dealing with oversized objects, it assigns wrong labels and ignores them. Figure 2 presents the segmentation masks generated from our reinforcement learning (RL) algorithm on the PH2 dataset. As can be seen in the figure, the masks are clear.

RL is prominent in artificial intelligence applications, such as anomaly detection, robot control, computer vision, autonomous driving, and computer gaming. DL and RL are used to solve increasingly complex problems resulting from a breakthrough in deep learning (DL). Deep reinforcement learning (DRL) trains an intelligent agent to deal with the Markov decision process (MDP) issue that combines DL with RL. GrabCut [20] is a conventional interactive lesion segmentation method that divides the foreground and background regions using a coarse bounding box. Guotai et al. [21] proposed a CNN-based model that improved the lesion segmentation results by scribbling on the initial

segmentation results. When physicians delineate the ROI on a lesion image, they first perform the coarse segmentation to determine the majority of the ROI's area.

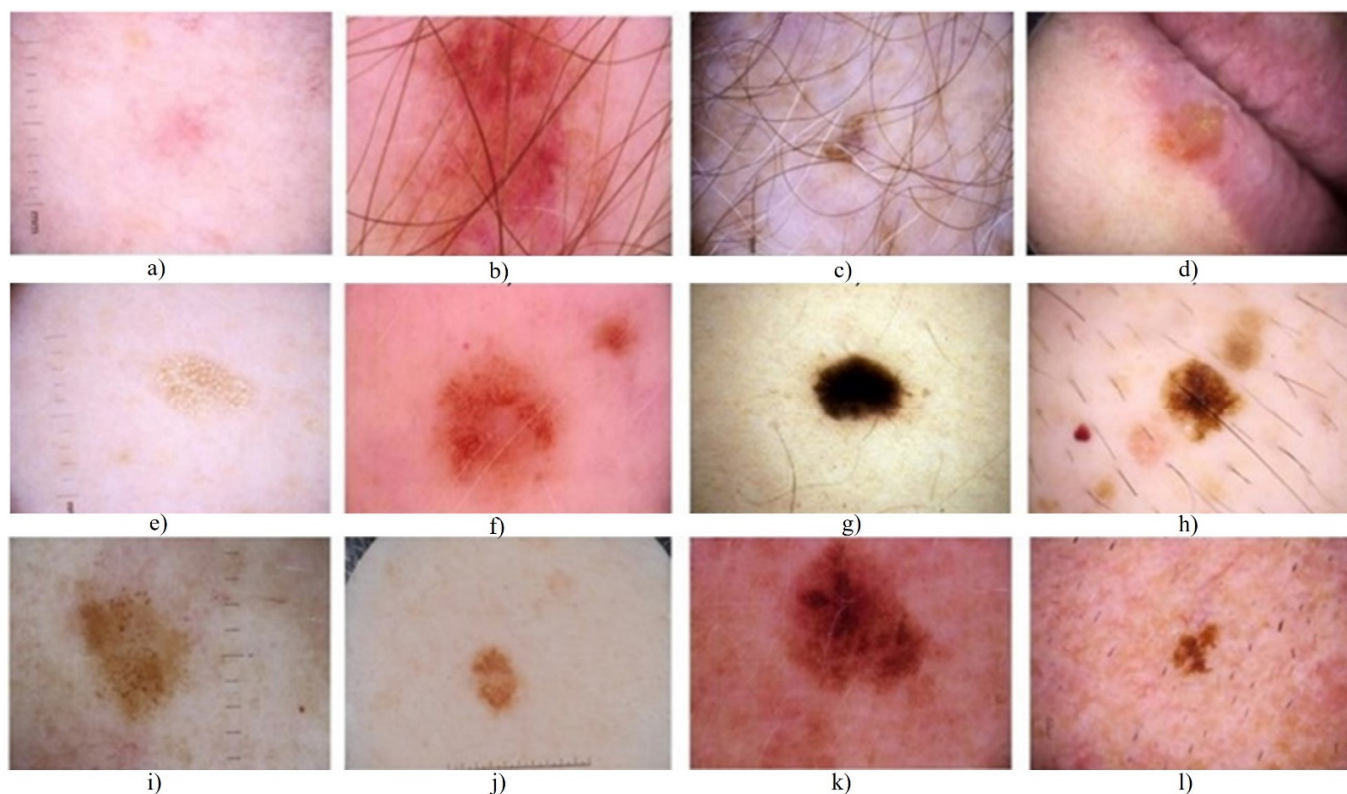


Figure 1. The skin sample lesion images taken from the ISIC 2017 dataset, which have different class differences within due to the various variations, such as severe (a,d,l), blurred boundary (b,d–f,h,l), inhomogeneity (b,i,f,h,k), hair appearance (g,c), poor contrast (h,j). The difficulties mentioned above are present in the skin lesion datasets. These datasets are used for the automated segmentation process.

The clinicians then fine-tune the coarse segmentation procedure in the form of a multi-step process. This segmentation approach is similar to interactive segmentation [22]. Since it includes painting many strokes with a brush on the foreground and background or creating a box around the foreground, the interaction is considered prior knowledge for segmentation. This strategy gathers past information and improves the segmentation algorithm's efficacy by interacting with the user. Inspired by these concepts, the proposed method is proposed as a multi-step segmentation process. Our method has the advantage of improving the segmentation performance by automatically gathering the previous knowledge, thus eliminating the need for human involvement.

The segmentation process is proposed as an MDP. In the next step, the segmentation mask predicted previously is taken as prior knowledge in the following step during the segmentation process. The agent executes an action for the segmentation process in every step that depends on the current segmentation mask and the input image. Our method derives its inspiration from the stroke-based stylization approach [23]. We propose a segmentation executor that helps draw a brushstroke for the input segmentation mask to designate the ROI. A neural network is used in the segmentation executor, which converts the continuous action parameters to brushstrokes. It can be used in different ways, and after some specific number of phases, the final segmentation results are achieved. The location and shape to get the fine-grained segmentation is established by using the continuous action parameters set.

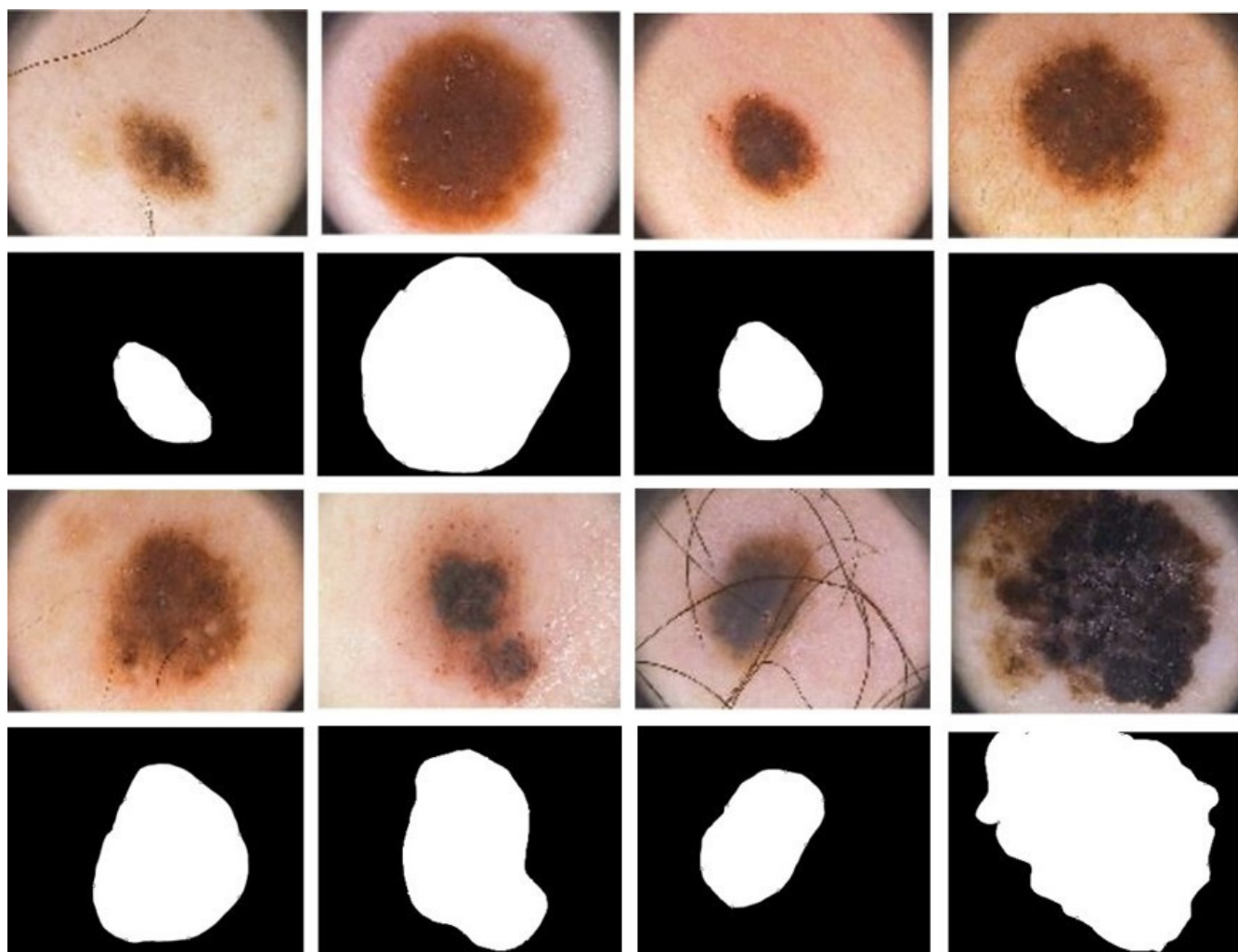


Figure 2. Segmentation masks generated from our RL algorithm on the PH2 dataset.

Deep Q-Networks (DQN) [24] is one of the most often utilized DRL algorithms. However, it is only restricted to solving the discrete action space problems. We use the Deep Deterministic Policy Gradients (DDPG) [25] algorithm for solving the continuous action space. DDPG is a learning method that simultaneously learns a policy and Q-function. Before employing the Q-function to learn the policy, it uses the off-policy data and Bellman's equation. To the best of our knowledge, this is the first attempt to represent and solve the skin lesion image segmentation problem as an MDP using DDPG. Since the DDPG is heavily dependent on searching the correct hyperparameters for the current task, we choose the action bundle, a suitable hyperparameter for the algorithm, thus increasing its stability.

The following are the main contributions of this work:

- The skin lesion image segmentation is proposed as an MDP. It is solved with the DDPG algorithm, similar to how the physicians delineate the lesion image ROIs.
- The proposed skin image segmentation executor is based on the quadratic Bezier curve (QBC) and uses the action bundle as a hyperparameter to further improve the *Acc* of the segmentation process.
- We use a modified experience replay memory (ERM) to train the segmentation agent efficiently. The ERM helps in efficiently utilizing the previous experiences by learning multiple times.

- We perform a quantitative statistical analysis of our skin lesion segmentation results to show the reliability of our segmentation method and compare our results to the current state-of-the-art approaches.

The structure of the article is as follows: Section 2 describes the current state-of-the-art methods, Section 3 presents our proposed RL method, overview, and the details of the experimental setup; Section 4 presents the results and discussion of our method. Finally, we conclude our article in Section 5.

2. Related Work

Many strategies used in the skin lesion segmentation are established in the literature, including region-merging-based approaches [26], active contour models [27–29], thresholding-based methods [17]. Many conventional methods [28,29] based on morphological processes and clustering algorithms are proposed in the literature. The skin lesion is split into the foreground and background regions using K-means clustering by Jafari et al. [30]. Similarly, Ali et al. [31] suggested that skin lesions be segmented using fuzzy C-means (FCM). The contour is produced regularly in another important class of techniques called the active contour models [27–29] as it approaches the pigmented regions boundaries. After generating candidate regions using threshold-based methods, the active contour models are directed by multi-direction gradient vector flow (GVF) snake [29], local histogram fitting energy [26], and can be used to enhance the course segmentation.

On the other hand, the traditional methods often use complex pre-and post-processing processes and a slew of data-dependent intermediate stages. Consequently, the performance of the conventional method primarily depends on these phases, requiring the design step to be done carefully when working with a variety of datasets. They will fail if the boundaries of the pigmented regions are unclear, and the skin conditions are complex. Deep CNN models have excelled in several computer vision applications [32–34], including advanced skin lesion segmentation. In general, convolution and pooling methods are used in basic CNN models. Deeper neural networks can extract more semantic and abstract characteristics using the learned kernels (e.g., components and shape).

The output feature maps of classification neural networks often shrink over time (by subsampling). Consequently, a probability vector with values ranging from 0 to 1, and a dimension equal to the number of categories, is generated. This is an encoding method in which the abstract and semantic properties encode the images as the neural network grows in depth. A segmentation neural network has a similar fundamental structure to a neural network classifier. Still, it also has a decoding route that attempts to improve the output resolution (through upsampling), such that the output segmentation mask size matches the input image size. Based on the above, Jafari et al. [35] proposed segmentation as a classification problem for skin lesion analysis. The image patches inputs of various sizes centered on a single pixel, and the output for that pixel is the projected label. In this scenario, the consideration of pixel context information present locally is adopted. Since this method relies on pixel-level prediction, dense prediction is needed, and then the research moved towards combining decoding pathway CNN to do the lesion segmentation. Due to its success, Ronneberger et al. [36] created the popular U-Net, which is extensively utilized in medical image segmentation applications.

Several U-Net-based melanoma segmentation and classification methods have been proposed [37–39]. Liu et al. [11] used the dilated convolution after the finish of every convolutional block present in the original U-Net to extend the proposed technique receptive field. Abhishek et al. [40] enhanced performance by integrating and choosing different color bands dependent on color changes. Yuan et al. [41] proposed a framework based on convolution–deconvolution. A Jaccard distance-based loss function was considered apart from the conventional cross-entropy loss. Al-Masni et al. [42] developed a full resolution convolutional neural network (FrCN) that learns the properties of full resolution for each input data individual pixel without subsampling. Bi et al. [43] proposed training the distinct CNN models for every class known using the category information. The hierarchical

development model-based stepwise integration (PSI) model was used to improve the output of lesion segmentation. Sarker et al. [44] proposed pyramid-pooling networks with dilated residual networks to segment skin lesions. The combination of endpoint error loss results in negative log-likelihood sharp boundaries. Xie et al. [16] suggested skin lesion segmentation as a mutual bootstrapping CNN method and classification, in which one job bootstraps the other.

Long et al. [45] first suggested a fully convolutional network (FCN) with a skip architecture based on a standard classification network to segment an entire image swiftly. Karthik et al. [46] used the Leaky ReLU and FCN in the final model layers frameworks to separate the ischemic lesions. Milletari et al. [47] proposed a V-Net-based architecture to segment the medical images in 3D and 2D formats. Many interactive segmentation algorithms are developed, with the physicians assisting the whole segmentation task. Olaf et al. [36] proposed a CNN and FCN-based biomedical image (cell) segmentation architecture. The classification network, led by a coarse segmentation network trained explicitly for this purpose, is guided by the expected coarse mask. Simultaneously, class-specific localization maps are produced by classification activation mapping (CAM). Then the concatenation is done into a U-Net-like network to improve the coarse mask prediction. DEXTR (deep extreme cut) [34] showed that using extreme points (contours corner points) as CNN input may improve the nature images instance segmentation results [34]. On the other hand, reference [34] uses extreme point inputs, the quality of which define the segmentation's efficiency. According to research [48,49], the auxiliary function of boundary/edge prediction helps in instance segmentation.

A loss function based on the Dice index is proposed to enhance the segmentation network. In addition to FCN-based techniques, several DL-based image segmentation algorithms have been proposed, including polygon-RNN [50], DeepLab V3+ [51], and multi-task network cascades [52]. In recent years, novel approaches for various applications, such as area extraction [53], wound intensity correction [54], and automated lung nodule categorization, have been developed [55]. Although there are positive effects of the therapies discussed above, a few pieces of literature have examined how physicians compute the ROI in skin imaging. The RL helps in imitating the demarcation technique of a physician. RL is significantly progressing in many applications, by combining RL with DL, DQN [24], DDPG [25], proximal policy optimization (PPO) [56], and asynchronous advantage actor critic (AC) are examples of deep neural networks used in DRL techniques for agent training. DeepMind achieves human-level game-playing abilities using DRL [23]. Therefore, other researchers have begun to use DRL for a range of problems, such as recommendation systems [57], game simulators, the Internet of Things, and adaptive packet scheduling [58]. DRL methods show promise in image classification, landmark identification [59], object localization [60], visual navigation [61], large-scale 3D point clouds semantic parsing [62], and face recognition [63].

Sahba et al. [64] developed a system for the image segmentation of prostate images based on RL Q-learning [65], which helps in finding the best values suitable for the subcategory of images and enhances the extraction of the ROI from the image. In contrast, Q-learning is limited to a narrow set of states and actions. Several researchers have tried to use DQN for image segmentation in recent years, combining Q-learning with CNN. DeepOutline [66] is an end-to-end deep RL framework for semantic image segmentation that works similarly to a user sketching the outlines of objects in an image with a pen. This approach is also proposed as an MDP. SeedNet is a game-changing seed generation method for interactive segmentation [67]. In each of these methods, DQN is utilized for training an image segmentation agent. DQN, on the other hand, is unable to maintain continuous activity, demanding additional operations to address the issue. In this article, we use the DDPG algorithm directly to segment lesions to save time and effort.

3. Proposed Method

This section addresses the public available skin lesion datasets, preparation of the ground truth images, and our proposed RL method. The ISIC-2017 Skin Lesion Challenge dataset [12] and the PH² dataset [13] and Human against Machine (HAM 10,000) [68] are three public datasets used by our method. In addition, we scaled all of the images to a size of 361×256 pixels to increase *Acc* and reduce computational costs.

3.1. ISIC-2017 Segmentation Dataset

The ISIC is a leading organization in terms of the availability of skin lesion image datasets. In addition, it provides expert annotations for the lesion images that can be used by several automated computer-aided diagnosis (CADx) applications. These applications use these datasets to detect melanoma and other cancers. This organization holds annual skin lesion competitions to inspire more researchers to develop CAD applications to identify lesions and promote skin cancer awareness [27]. The ISIC 2017 skin lesion dataset includes 2750 images, with 2000 in the training set, 150 in the validation set, and 600 in the test set. The algorithms must attain high Sensitivity (*Sen*) and Specificity (*Spe*) values to ensure that the lesions are correctly segmented. Unfortunately, when the ISIC challenge 2018 [27] was held previously, they did not release the ground truth of their training dataset. As a result, we focus our evaluation on the ISIC-2017 dataset.

3.2. PH² Dataset

The PH² dataset contains 200 images, 160 of which are naevus (both atypical and normal naevus) and 40 melanomas [13]. The ground truth in this dataset offers the true and precise boundaries of skin lesions. This dataset acts as an alternative test dataset for DL models trained on the ISIC-2017 segmentation training set. The ISIC challenge dataset contains several dermoscopic skin lesion images, which are collected by various dermatoscopes and camera devices worldwide. Consequently, the color normalization and illumination pre-processing must be done using the color constancy method. To process the datasets, we utilized the shades of the gray algorithm [69]. Figure 3 shows our skin lesion segmentation results by our proposed neural network architecture.

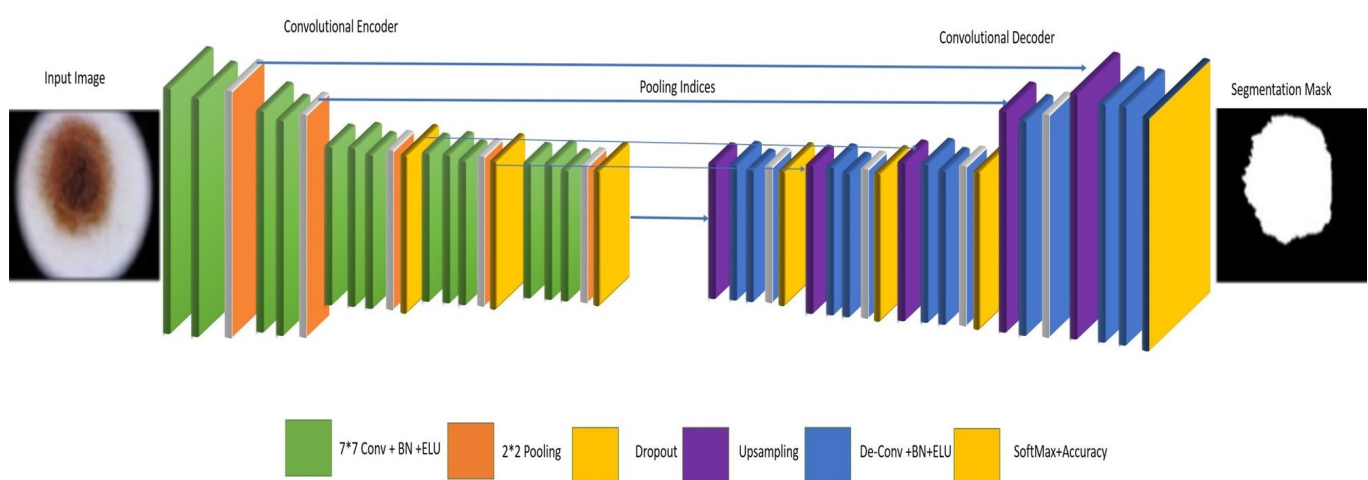


Figure 3. Segmentation of skin lesions by our proposed neural network architecture.

3.3. Overview of Our RL Method

In this article, we propose a groundbreaking multi-stage segmentation strategy based on DRL to detect skin lesions. In each step, a segmentation agent is trained to find the best segmentation technique based on the previous step's evaluation results. In this article, the DDPG algorithm trains the segmentation agent to solve the MDP problem. DL and deterministic policy gradient (DPG) are combined in the DDPG algorithms [70]. The actor

uses the low-dimensional state space to make choices. The advantage of the DDPG is that it classifies policies that outperform the actor.

Since DDPG is an off-policy algorithm, it provides a huge replay buffer, enabling it to learn from a wide variety of unrelated transformations. The predicted gradient of the action-value function is a more appealing version of the DPG. Because of its simplicity, the DPG can be calculated far more accurately than the traditional stochastic policy gradient. In high dimensional action spaces, DPG algorithms outperform their stochastic equivalents drastically. DDPG is a non-policy model-based algorithm for studying continuous action. It uses DQN's ERM and slow learning goal networks and is built on DPG, which operates through continuous action spaces. Compared to traditional methods, such as the level set, Chan-Vese, and snakes, the method proposed does not require any technological skills.

The neural network optimizes the method proposed based on the segmentation results from the prior step. Thus, the network evolves techniques for segmenting skin lesion images without the need for specialized knowledge. DDPG is used to address problems that require continuous action space based on the AC architecture. It is suggested to solve two challenges: overcoming the delayed RL issue for neural networks and creating a self-learning framework based on a neural network that needs no training or reinforcement from the context. The AC paradigm is a mix of value and policy-based methods. The first uses an implicit system for learning the value function and an action-based value function to achieve the policy. Conversely, policy optimization refers to explicitly defined model capacity, such as the policy gradient (PG) [71]. Our RL-based image segmentation algorithm is shown in Algorithm 1 below:

Algorithm 1 RL based image segmentation.

Randomly initializing actor network $\mu(s | \theta^\mu)$ and critic network $Q(s, a | \theta^Q)$ with weights θ^Q and θ^μ .
 Initializing of the target networks μ' and Q' and weights $\theta^{\mu'} \leftarrow \theta^\mu, \theta^{Q'} \leftarrow \theta^Q$
 Initializing of experience replay memory R
for episode $e = 1, N$ **do**
 Initializing a random process M for exploration of actions
 Received s_1 initial observation state
 for $x = 1, T$ **do**
 Select action parameter set $a_t = \mu(s_t | \theta^\mu) + N_t$ accordingly to the exploration noise and the current policy
 Feed the action parameters $(A_{s_0}, A_{s_t}, A_{s_{t+1}}, \dots, A_{s_T})$ in the segmentation executor.
 Feed the updated segmentation mask $S_{mt} + 1$ and the ground truth for computation of reward function $r(t)$.
 Execution of actions a_t and observing reward r_t and observation of new state s_{t+1}
 Storing transition (s_t, a_t, r_t, s_{t+1}) in R
 Sampling of a random mini-batch (s_i, a_i, r_i, s_{i+1}) of N transitions from R
 Set $y_i = r_i + \gamma Q'(s_{i+1}, \mu'(s_{i+1} | \theta^{\mu'}) | \theta^{Q'})$
 Feed the ground truth S_{mt} in the critic network
 Feed the reward $r(t)$ and long term expected return Q to the evaluation network.
 Evaluation of the segmentation policy focused on reward $r(t)$ and the long-term return Q .
 Updating critic by minimize of the loss: $L = \frac{1}{N} \sum_i (y_i - Q(s_i, a_i | \theta^Q))^2$
 Using the sampled policy gradient to update the actor policy:

$$\nabla_{\theta^\mu} J \approx \frac{1}{N} \sum_i \nabla_a (Q(s, a) | \theta^Q) \Big|_{s=s_i, a=\mu(s_i)} \nabla_{\theta^\mu} \mu(s | \theta^\mu) | s_i$$

 Updating the target networks:
 $\theta^{Q'} \leftarrow \tau \theta^Q + (1 - \tau) \theta^{Q'}$
 $\theta^{\mu'} \leftarrow \tau \theta^\mu + (1 - \tau) \theta^{\mu'}$
 end for
end for

The actor critics are made up of two networks: a value network and a policy network. The former is called the critic, and the latter is called the actor. In the AC network, the actor's responsibility is to learn policy, while the critic helps evaluate the decision taken by the actor. The actor aspires to get enhanced performance, while the critic aspires to be more precise and accurate. Iterative optimization is used in the training process, following

the theory of adversarial networks, due to the interdependence and interaction in both the actor and the critic [72]. Moreover, our solution lets the off-the-shelf segmentation executor perform a segmentation mechanism focused on the action parameters present as a continuous set. The segmentation executor performs a brushstroke centered as a series of action parameters and draws it onto a segmentation mask taken as an input to improve the *Acc* of the segmentation process. The architecture of our proposed method is depicted in Figures 3 and 4.

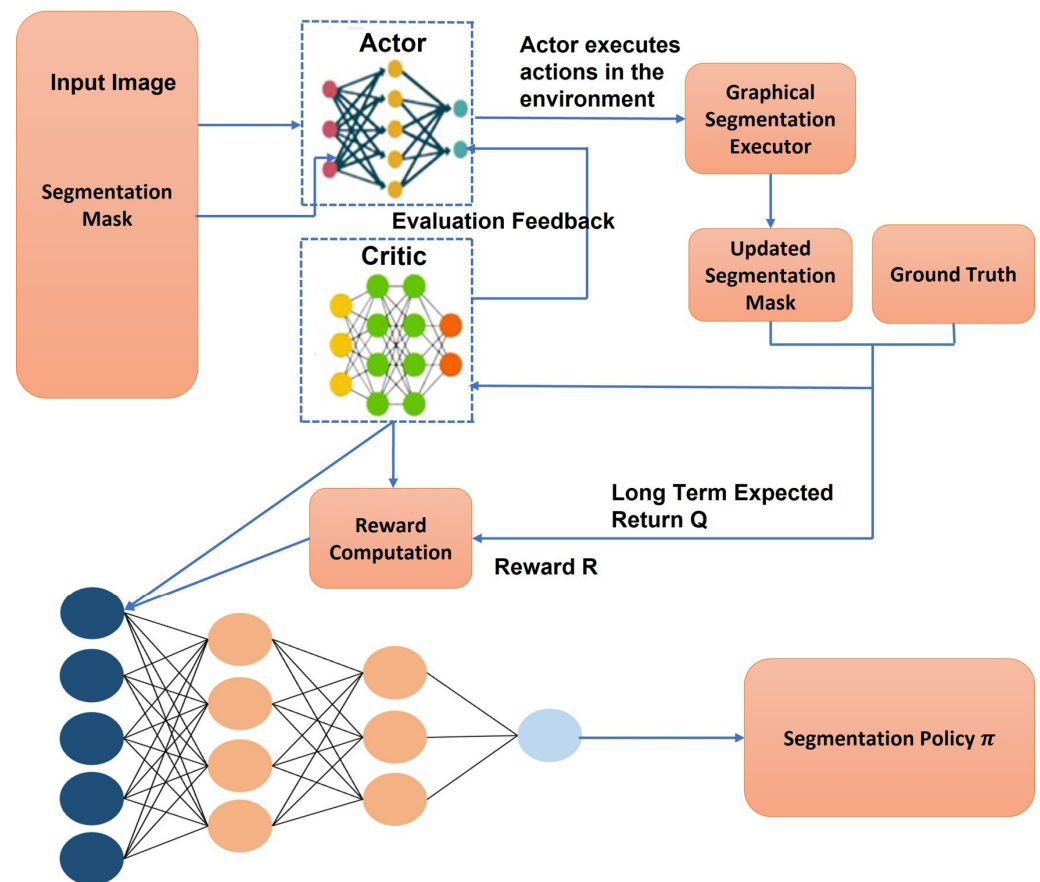


Figure 4. Our RL framework for the segmentation process. The set of actions selected by the actor depend on the new segmentation mask S_{mt} and the input image. The parameters are passed to the segmentation executor. $S_{mt} + 1$, the updated segmentation mask, is then produced by the segmentation executor. The new segmentation mask has three distinct features. The mask is used for the calculation of reward by comparison with the ground truth mask. Secondly, it is given as input to the critics and the ground truth to calculate the Q's long-term estimated return. Thirdly, it is used for updating the previously used S_{mt} segmentation mask. The critic actor and segmentation executor are both built using neural networks. The current segmentation methodology π is evaluated by the evaluator and is based on the long-term return Q and reward R.

A segmentation executor is used in the proposed method to perform the segmentation action, and it is based on neural networks. By using an initial segmentation mask and an input image, the agent attempts to characterize an action series $(A_{s0}, A_{st}, A_{st+1}, \dots, A_{sT})$ using the new segmentation strategy (S) for a task that requires segmentation (mapping of action A to state S). The role of synthesizing the texture of each stroke in this RL method is conceived as a sequential decision-making mechanism centered on the MDP mechanism, with a soft tuft brush acting as an RL agent. The probability of good action is very high at any stage, i.e., the decision should increase the compatibility between the future and the previous decisions. The *Acc* with which the segmentation process is carried out directly

affects the efficiency of the segmentation results. Segmentation methods are used to train the DDPG handler. The segmentation executor adopts the brushstroke chosen by the actor in process t and obtains a modified $S_{mt} + 1$ segmentation mask. In the segmentation procedure, these steps are repeated.

At the end of the segmentation phase, we have the final segmentation mask. The residual design of our model is built similar to ResNet-18 [32] and is used for the policy (actor) and the value (critic) network. Meanwhile, batch standardization is used by the policy network [73]. Our method's strength stems from integrating normalization into the model design, and the execution of normalization takes for every mini-batch of the training process. Batch normalization helps them train faster while reducing their time on the initialization process. It serves as a regularizer, obviating the need for dropout in some instances; by standardizing weights with translated ReLU (TReLU), the sensor network aids in the training of the model. Convolutional layers and fully connected layers are used in the segmentation network. The subpixel [74] technique is used in the segmentation executor to increase the brushstroke resolution. CoordConv layer is taken as the first layer by both the critic and the actor. The actor, critic, and the segmentation executor interaction framework is shown in Figure 4. The architecture details of these are described in Figure 5. In Section 3.4, the image segmentation method is depicted as an MDP process. The details of the segmentation executor and the steps used to improve our segmentation results are discussed in Section 3.3.

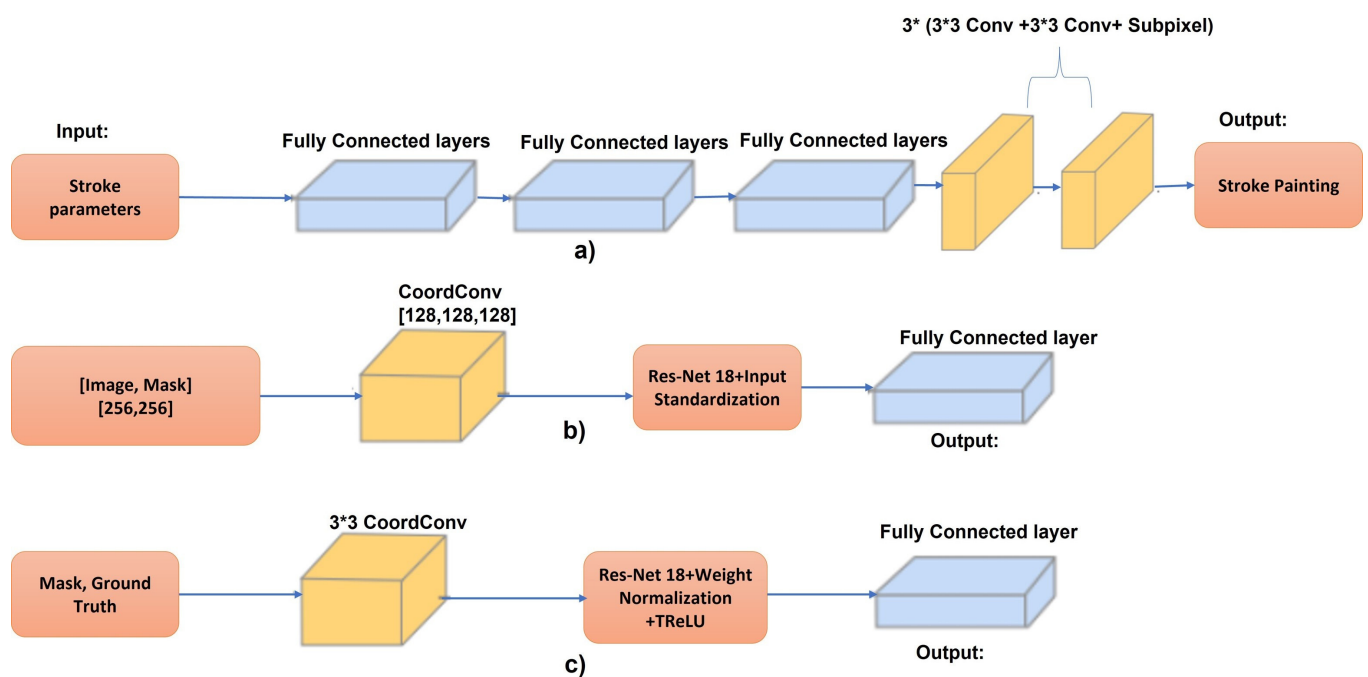


Figure 5. (a) The segmentation executor. (b) The actor and (c) The critic. The following are the network's architecture. The segmentation executor produces various segmentation results depending on the parameters taken for action. The action set is output by the actor as a series of five actions at a time, each with 8 parameters. The outputs of the critic are single valued for estimation. ResNet-18 is the architecture used.

The ERM for DDPG training is proposed in Section 3.4, which helps us obtain improved segmentation results

3.4. MDP for the Segmentation of Skin Lesion

The segmentation agent is used to find the ROI in this algorithm, with the skin lesion segmentation mechanism modeled as an MDP process. State, space S , operation A , and the reward features are the three main features of the MDP. This method's three-dimensional explanation is proposed as follows:

State: many of the agent's activities in the environment are included in the state space. The agent's decision will be based on this information. The state in this work consists of the image I , the current segmentation mask S_{mt} and the step-index t , which is defined as $S_t = (S_{mt}, I, t)$. S_{mt} is a mask for the segmentation process with a pixel scale of 0 or 255. The background pixels are 255 and 0. The original segmentation default value mask is 0. I is a representation of the lesion that requires segmentation. The index at step t is used to differentiate between the various phases of the segmentation. The state of the segmentation phase terminates in our multi-step segmentation process. The steps are maximized and performed until the training process is completed. Until the maximum number of stages has been achieved, the agent can enter the terminal state and continue to execute the final segmentation task.

Action: the action area includes any operation that the segmentation executor can conduct. In a state, the agent makes a policy declaration π in the space of action. The action is then used to adjust the brushstroke direction and shape, described as several parameters.

Reward function: The reward function defines the state to reward mapping in the RL task. Regularly, the agent's job is to maximize the amount of discounted future rewards R . This function signifies the immediate state reward when an action is changed, which helps evaluate the result's effectiveness for the decision taken by the agent. The M 's segmentation mask changes at every step during the process of training. Consequently, the mask's Acc is measured by the comparison of the truth ground mask for every step. The $L2$ mean square error is used as an arbitrary metric. R_{l2} is the default $L2$ reward mode. To better reflect each step's effect, we need an essential reward feature to take advantage of the $L2$ change pattern. The resemblance between the two images calculates the value of $L2$. If two images are the same, the loss of $L2$ is equivalent to 0. The reward function can be modeled using the difference in $Rl2$ between the two adjacent steps. R_{diff} 's reward function is shown in the Equation (1) below, where $S_{m}A_t$ denotes the last segmentation mask, and S_{mt} denotes the mask.

$$R_{diff} = L2(S_{mt} - 1, G) - L2(S_{mt}, G) \quad (1)$$

The reward function sends a favorable signal when the loss of $L2$ decreases and vice versa. Therefore, it is essential to predict Q 's long-term return value at each point to enhance the learning trends. The reward function underlines the effectiveness of each action selected. The consistency of the operation chosen over the entire duration of the segmentation is an essential function. $Q(S_t, A_t)$ is the A_t value defined in the S_t state.

The Bellman Theorem is used to measure $Q(S_t, A_t)$ and use the reward function. The reward function is denoted by $R(S_t, A_t)$. $Q(S_t, A_t)$ is shown in Equation (2) below as:

$$Q(S_t, A_t) = R(S_t, A_t) + \gamma \max_{a'} Q(S'_t, A'_t) \quad (2)$$

where $Q(S_t, A_t)$ is the value of Q for the selection of A_t from the state S_t . The reward function is represented by $R(S_t, A_t)$. γ signifies the discount factor that represents the advantage of future returns $Q(S_t + 1, (S_t + 1))$ in relation to the immediate reward $R(S_t, A_t)$. When γ value is zero, it is equal to the focus only on the immediate reward while ignoring all the long-term returns. When the γ value is zero, it is equal to the focus only on the immediate reward while ignoring all the long-term returns.

In the case of γ is 1, both the immediate reward and long-term returns are equally important. The segmentation policy is denoted by π . Using the Bellman equation, the critic helps estimate the decision based on the long-term return Q taken by the agent. The Bellman rule is used to establish the learning process of the long-term reward Q of the decision made by the agent. The value of Q depends on the activity and state. It is calculated using the Bellman equation. The critics' estimation increases the segmentation Acc . Instead of S_t and A_t , the critics are fed the ground truth of S_t . The new value function after modification of $V(S_t, G)$ undergoes training in the following Equation (3) below as:

$$V * (s) = \max \max_{a'} (R(S_t, A_t) + \gamma V * (s')) \quad (3)$$

Finally, the algorithm of DDPG is used for optimizing the MDP for lesion segmentation. Section 3.5 provides the detail of the hyperparameter action bundle and the segmentation executor.

3.5. Action Bundle and the Segmentation Executor

The mask on which the neural network draws the brushstroke as the ROI renderer implements the segmentation executor mentioned above. There are two advantages of the segmentation executors. First, it could be well distinguished and combined with DDPG. Secondly, a neural network can have a fine-grained quality. The segmentation executor is run by using learning algorithms that are supervised and executes on a vast number of samples taken for training, which are collected from various graphical rendering systems. Several segmentation executors, including triangles, circles, square Bézier curves, and B-spline curves, produce multiple brushstroke forms. A polynomial B-spline curve is more reliable to employ than a Bezier curve since its degree is independent of the number of control points. The B-spline curve offers local control over each section of the curve through control points. For a given parameter, the total of the basis functions equals one. Based on the experimental findings, QBC and B-spline significantly benefit in the segmentation of lesion images. Thus, we use the QBC and B-spline to segment the lesion images. The QBC action parameters are described in Equation (4) as follows:

$$A_t = (x_0, y_0, x_1, y_1, x_2, y_2, r_0, r_1), \quad (4)$$

where the three QBC control point coordinates are $(x_0, y_0, x_1, y_1, x_2, y_2)$, (P_0, P_1, P_2) , $A_t = (x_0, y_0, x_1, y_1, x_2, y_2, r_0, r_1)$.

The thickness of the two QBC endpoints (P_0, P_2) determines the parameters (r_0, r_1) .

Approximately eight motion parameters were obtained using neural networks, with the stroke proportions and forms being generally distinct. The path traced by a quadratic Bézier curve is given in the form of function $S(x)$ as seen in the Equation (5), given points P_0, P_1 , and P_2 . The following Equation (5) below describes the QBC theorem, and the interpretation can be made by taking the linear interpolation of the points corresponding to the linear Bézier curves from P_0 to P_1 and then from P_1 to P_2 respectively. Then we rearrange the Equation (5) for obtaining $S(x)$ in Equation (6) as follows:

$$S(x) = (1-x)[(1-x)P_0 + xP_1] + x[(1-x)P_1 + xP_2], \quad 0 \leq t \leq 1 \quad (5)$$

$$S(t) = (1-x)^2P_0 + 2(1-x)xP_1 + x^2P_2, \quad 0 \leq t \leq 1 \quad (6)$$

The tangents at QBC P_0 and P_2 converge at P_1 . The curve starts at P_0 in the direction of P_1 , where the curve ranges from 0 to 1 and curves from the direction of P_1 to the end of P_2 . A spline of the order n is a piecewise polynomial function having a degree $n-1$ present in a variable x . The knots are those values of x , where the intersection of the polynomial pieces occurs, and the listing is done in the ascending order as $\{t_0, t_1, t_2, \dots, t_n\}$. When the distinct knots are considered, the first $n-2$ derivatives of the polynomial components across each knot are continuous. Over a r knot, the spline is continuous only on the first $n-r-1$ derivatives of the spline. In Equation (6), there is a single spline $S_{i,n}(y)$ that obeys a given sequence of knots having a specific scaling factor satisfying Equation (7).

$$S_{i,n}(y) = \begin{cases} 0 & \text{if } y < x_i \text{ or } y \geq x_{i+n} \\ \text{non-zero} & \text{otherwise} \end{cases} \quad (7)$$

If the constraint is additionally added, such that $\sum_i S_{i,n}(y)=1$ for all y between the last and the first knot, then the factor of scaling $\sum_i S_{i,n}(y)$ becomes fixed. The resulting spline functions are called B-splines.

The defining of higher order B-splines in the form of a recursive Equation is shown in Equation (7) as follows:

$$S_{i,k+1}(y) = w_{i,k}(y)B_{i,k}(y) + [1 - w_{i+1,k}(y)]B_{i+1,k}(y) \tag{8}$$

$$\text{where } w_{i,k}(y) = \begin{cases} \frac{y-x_i}{y_{i+1}-x_i}, & x_{i+k} \neq t_i \\ 0 & \text{otherwise} \end{cases}$$

The action bundle strategy further improves *Acc* and is inspired by the frameskip [57], an effective hyperparameter for several RL tasks. The frameskip determines the granularity at which environmental agents are tracked, and the action to be used shall be selected. The skip frame parameter *K* allows the agent to take into repetition the actions at selected *K* frames. The connection between the associated states and the computational resources saved by this technique is explored. The connection between the different actions, referred to as the action bundle, is explored. To encourage the actor to delve further into the action space, the actor creates an action package by selecting *K* acts for the actions taken from the action space bundle. The segmentation executor then conducts *K* operations in a single action kit; thus, improving the segmentation result *Acc*. In Section 3.6, we discuss the ERM for DDPG.

3.6. Modified ERM for DDPG

The various DRL algorithm-training instances have been referred to as transformations. The five parameters for each transformation are the current state *S*, preferred actions *A* dependent on *S*, instant reward *R*, next step *S'*, and terminal, i.e., when the state undergoing execution comes to an end. The ERM stores transitions (*S, A, R, S', Terminal*), and random sampling prevents interaction between transformations. When the ERM retains many samples from the agent’s experience with the environment, a small batch of transformations for the training agent is randomly sampled from memory. The ERM is used to optimize the critic inputs for the best possible assessment, as seen in Figure 6.

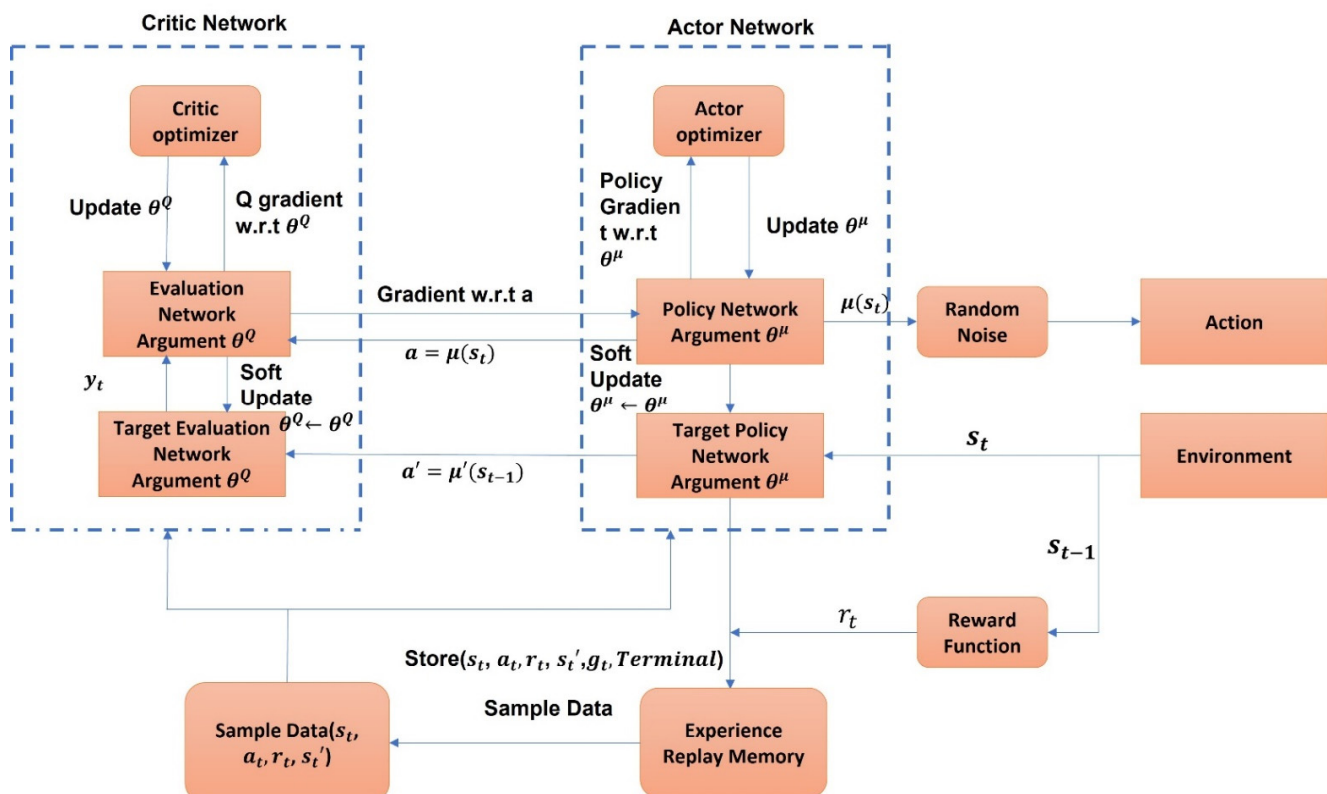


Figure 6. With Modified ERM. The ERM is used to adjust the critical feedback for correct assessment.

State S and Action A are given as an input to the critic network to maintain the Q 's long-term return. The performance of the critic depends on the actor's determination to determine the algorithm's efficiency and the correct policy π . The new parameter as a Ground Truth (GT) is added to optimize the necessary evaluation ability for segmentation assignment. The current transition is a new step ($S, A, R, S', GT, Terminal$). The ground truth and S' are sent to the critics for evaluation depending on the transition (new). On the other side, the presence of the ROI resembles the tissue that surrounds it, resulting in boundary ambiguity. In this situation, the whole scenario cannot be interpreted by the segmentation agent. The modified ERM is seen in Figure 6.

4. Results and Discussion

The performance of the RL algorithm is evaluated in this section. We conduct a qualitative and quantitative analysis of our proposed RL solution before comparing our results to those of various state-of-the-art segmentation algorithms and methods.

4.1. Experimental Setup

4.1.1. Implementation Details

The type of hardware used to assess network efficiency has a significant effect on its performance. The NVidia P100 form is used in the suggested RL method (16 G.B. RAM@1.32GHZ), where the performance is 9.3 TFLOPS. The train and test datasets are described in the device configuration discussed above. NVidia P100 is fitted with a 1.32 GHz processor and 16 G.B. of RAM for network training and benchmarking. The architecture uses Ubuntu 16.04, which is based on the programming language Python version 3.8. The Adam optimizer is used by the network, which has a learning rate of $[1e-4]$ and a micro-batch of 16. The decision-maker acts in the most simplistic form, receives a reward from the environment, and the environment shifts its state. The decision-maker then detects the state of the environment, takes the initiative, earns a reward, and so on. The state transformations are probabilistic and are solely determined by the current state and the behavior shown by the actor. The actor's reward is calculated by the behavior taken and the initial and current condition of the environment. The reward profit ratio for gamma is set to be at 0.850. The memory replay experience has been set to 600. The action bundle is set as $K = 5$, and the step number is set as $t = 3$.

4.1.2. Evaluation Metrics

To ensure the performance of the proposed models, the basic statistical parameters used in other literature works have been studied. (Sen) is calculated in Equation (9), as follows:

$$Sen = TP/TP + FN \quad (9)$$

It means that the number of lesion pixels in the image is distributed uniformly. Similarly, the parameter Spe determines if the pixels proportions have been correctly assigned to the image and is given in Equation (10) as follows:

$$Spe = TN/TN + FP \quad (10)$$

The rate of pixel classification referred to as Acc is determined in Equation (11), as follows:

$$Acc = TP/TN + TP + FN + FP \quad (11)$$

The spatial overlap that is present between the assigned binary mask and the segmented image is defined as the Dice coefficient ($Dice$), and is measured in Equation (12) as follows:

$$Dice = 2 TP/2 TP + FP + FN \quad (12)$$

The *Jaccard index* is the relationship between the binary labels and the pixel values analyzed for the input image. The Jaccard index is determined in Equation (13) as follows:

$$\text{Jaccard Index} = \frac{TP}{TP + FN + FP} \quad (13)$$

It is generally used to measure the change in the center of transformation present in the image axis. While true positive (*TP*) correctly depicts lesion pixels, false positive (*FP*) incorrectly depicts non-lesion pixels as lesions, true negative (*TN*) depicts all incorrectly labeled non-lesion pixels, and false negative (*FN*) represents the incorrectly identified lesion pixels.

A pixel's distance to a surface is defined in Equation (14) as follows:

$$HD(M, N) = \max_{m \in M} d(m, N), \max_{n \in N} d(n, M) \quad (14)$$

The Hausdorff distance between two surfaces *M* and *N* is computed using the difference between the predicted segmentation result and the ground truth. A lower HD number means the performance of the segmentation algorithm is good. The RVD value infers if the segmentation performed by the algorithm helps in the selection of the ROI area, more or less. The algorithm extracts a larger region in the ROI segmentation, if the value of RVD is positive or negative. The relative volume difference is calculated using the following RVD formula in Equation (15), as follows:

$$RVD = 100 \times \left[\frac{|M_{gt}|}{|M_m|} - 1 \right] \quad (15)$$

4.1.3. Evaluation and Comparison on the ISIC 2017 Dataset, HAM10000, and the PH2 Dataset

The trained neural networks for the CADx diagnosis of the pigmented skin lesions are challenging due to the lack and diversity of dermatoscopic image datasets. The HAM10000 dataset (Human against Machine with 10,000 training images) [68] is available to solve this problem. In this dataset, different modalities are used to collect and retain dermatoscopic images from various populations. The generated dataset contains 10015 dermatoscopic images that are utilized for training machine-learning algorithms. The cases in this dataset include all of the essential diagnostic categories as a representative collection in the pigmented lesions realms, such as intraepithelial carcinoma/actinic keratosis, Bowen's disease (akiec), benign keratosis-like lesions, basal cell carcinoma (bcc) (seborrheic keratosis, lichen planus-like keratosis, and solar lentigines), and dermatofibroma are some of the conditions (df).

The segmentation masks are evaluated on this dataset, and, as can be seen in Figure 7, the masks are very precise and clear. Figure 8a represents the initial input image with Figure 8b,c showing the segmentation of other methods that use a variety of loss functions. It shows the influence of the loss function TL and GN on the input images. In these, attention gates (AG) and group normalization (GN) analyze the input images and help in the prediction of the boundaries of skin lesions.

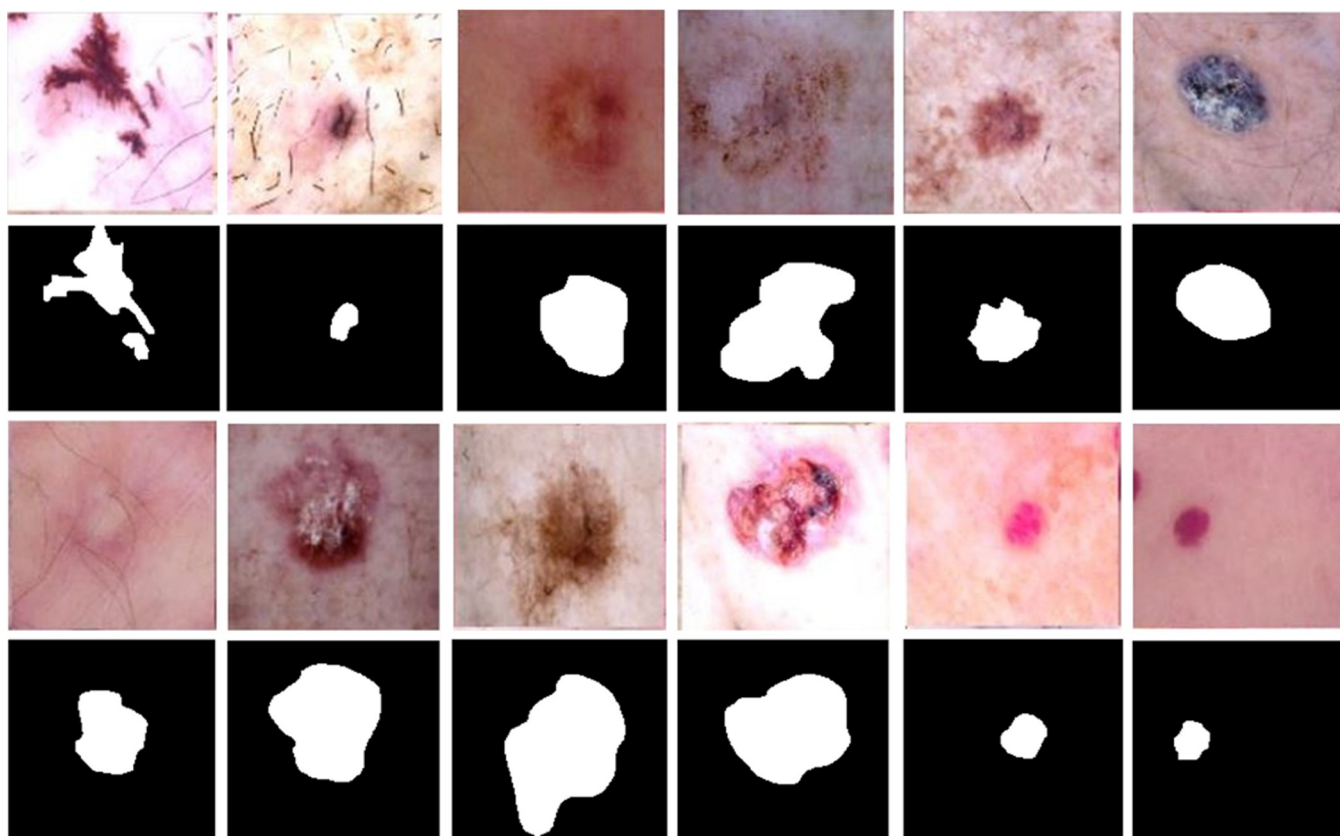


Figure 7. Evaluation results on the HAM10000 dataset. As can be seen in the second and the fourth rows, the results are precise and clear.

Figure 8d shows the lesion ROI computed by our RL algorithm. The results infer that the lesions are segmented in an accurate and precise manner. Our method reuses the CNN feature map and shortens the training and testing times, which helps us to train our networks from start to finish efficiently. Our algorithm enables the identification of high-dimensional hierarchical images. The detection approach is robust to changes in conditions, such as illumination and color balance. Our algorithm outperforms the alternative methods, as seen in Figure 9. Figure 9a shows the visual comparison evaluated on the ISIC 2017 image dataset (with a black background). Figure 9a shows the input images from the dataset. The provided binary mask (ground truth) is depicted in Figure 9b, whereas the method's prediction results (A.G. U-Net + GN) are depicted in Figure 9c.

Our RL algorithm results can be seen in Figure 9d. As can be seen, our algorithm is capable of segmenting lesions with great *Acc*. However, the segmentation mask results show the exact shapes of the lesions in some instances, which differ somewhat from the ground truth. Since the ground truth images are annotated manually, such minor errors can happen. Figure 10 shows the visual projections that have been derived from the proposed ISIC 2018 image dataset (present without the black background). Figure 10a shows the original image; Figure 10b shows the given binary mark (ground truth); Figure 10c the prediction effects of the method (Att U-Net + GN); Figure 10d the results of our RL algorithm. As can be seen, our segmentation mask corners are sharp and clear. The masks are much similar to the ground truth images.

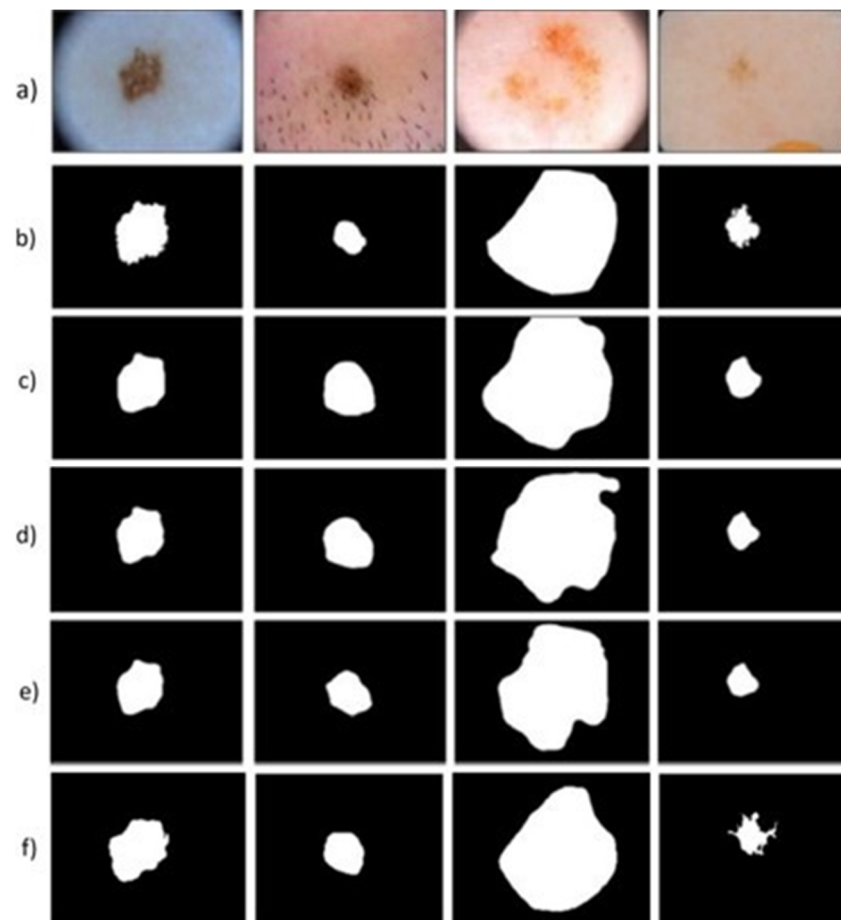


Figure 8. The ablation experiment segmentation diagrams for the input image. (a) The initial image taken from the dataset. (b) The image's ground truth; (c) Att U-Net + GN + TL; (d) Att U-Net + GN + FTL; (e) Att U-Net + GN + DL (TL: Tversky Loss; FTL: focal Tversky loss; DL = Dice loss). (f) Our RL algorithm results. The segmentation mask boundaries are sharper than the state-of-the-art methods, as shown.

Figure 11a shows the original input image; Figure 11b the binary mask (ground truth); Figure 11c the segmentation results of simple U-Net; Figure 11d the segmentation results of the SE block on the specific U-Net; Figure 11e the segmentation results of the BCDU network (with 1 dense unit); Figure 11f segmentation results of the U-Net network (with all the 64 filters); Figure 11g Att-U-Net + GN + TL segmentation results. Finally, Figure 11h shows the results of our RL algorithm. As can be seen from the image, the results of the masks are precise. The background is reflected in the darker (black) area, while the foreground is the lighter (white) area.

This method is quick, and it helps in calculating the running time easily. This method is well balanced because there is a high degree of separation between the foreground and the background. The proposed number of parameters in a model, storage measurements, and inference speed is compared to other state-of-the-art models. The GN used in previous work tests the mean and variation of the channel groups since they have been standardized, e.g., the AG U-Net model scans images in epochs that can easily connect to other algorithms. In BCDU, the image upload epoch is 359 s, while the U-Net baseline is 165 s. It takes 133 s to encrypt a terabyte of info. For the NVIDIA Quadro K1200 GPU, the proposed solution's predicted performance (AG U-Net + GN + TL) is faster than the U-Net baseline with a 256×256 input scale [75].

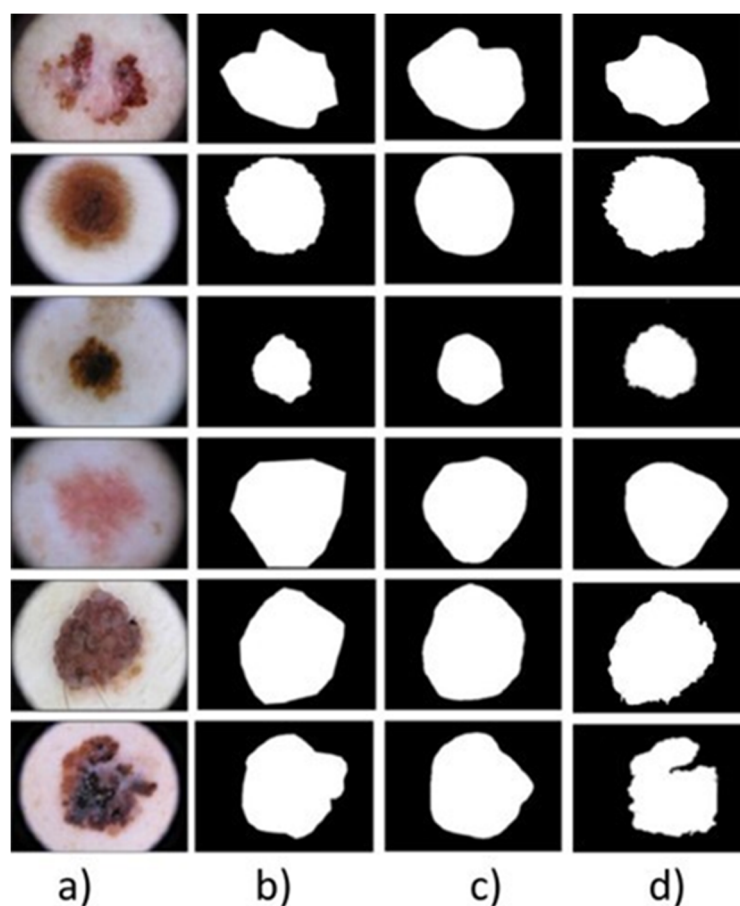


Figure 9. The visual projections derived from the ISIC 2018 image dataset. (a) The original image from the dataset. (b) The given binary mark (ground truth), (c) The prediction effects of the proposed method (Att U-Net + GN). (d) The results of our RL algorithm. Our algorithm is able to segment the lesions clearly. However, in some cases, it shows exactly the lesion shape in the segmentation mask, deviating a bit from the ground truth. However, as can be seen, the segmentation masks are clear.

The visual segmentation results in Figures 9–12 demonstrate the proposed RL method's better segmentation performance than the other state-of-the-art methods. The performance metrics visualization for the PH2 and the ISIC 2017 challenge dataset is shown in Figure 12. The statistical measures used by our methods, such as *Acc*, *Dice*, *Jaccard index*, *Sen*, and *specificity* are plotted for the PH2 and the ISIC 2017 skin segmentation dataset. The blue line denotes the performance metrics plotted for the PH2 dataset, and the red line shows the metrics for ISIC 2017 skin segmentation dataset. In Figure 12b, the statistical measures of our method denoted as *Dice*, *JSI*, *MCC*, and the overall of these statistical measures are plotted for the three categories of skin lesions: *naevus*, *melanoma*, and *seborrheic keratosis*. While the other blue line and the red line denote the PH2 and the ISIC 2017 dataset metrics, the highlighted points indicate the metric values such as *Sen*, *Spe*, and *Acc* for each of the categories on both datasets. Figure 12c shows the action bundle effect.

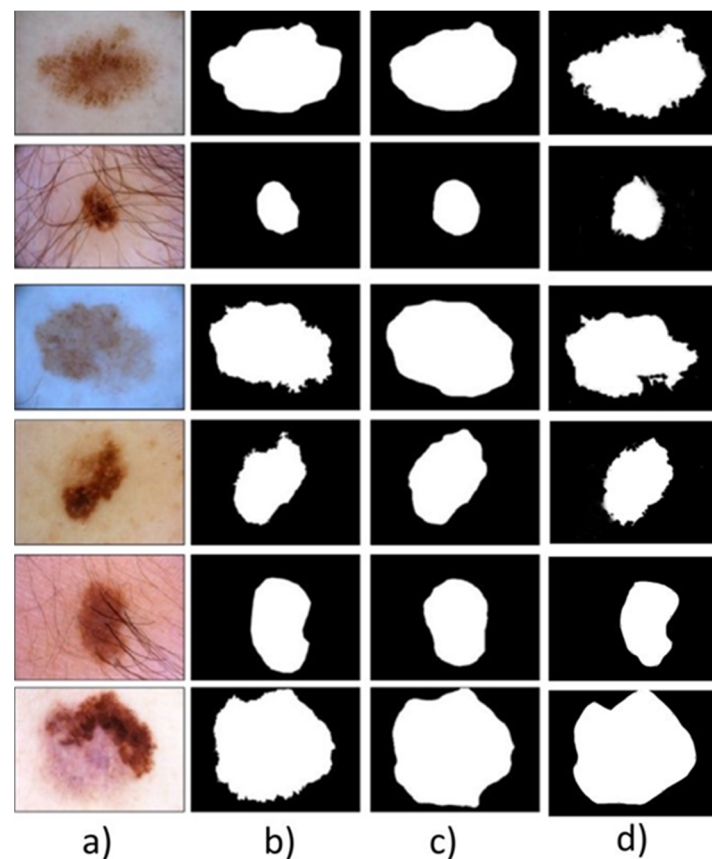


Figure 10. The visual images taken from ISIC 2018 image database model (present without the black frame). (a) The original image, (b) the given binary mark (ground truth), (c) the prediction effects of the method (Att U-Net + GN). (d) The results of our RL algorithm. As can be seen, our segmentation masks corners are sharp and clear.

The Dice index values taken for the ablation experiments are shown in Table 1. The various values of K are the different settings for the action bundle. We evaluate the effect of the value K on the Dice Values. We infer from the results that our segmentation results improve when we use both the modified ERM and action bundle. In Table 2, our approach works better for *Sen* and *Spe* than all of the algorithms, with values 98.59% and 97%. We aim to design these algorithms for high *Sen* and *Spe*. The *Sen* value is larger than other methods in terms of various statistical parameters taken for comparison. Table 2 compares the proposed model's segmentation efficiency to that of state-of-the-art methods on the PH2 dataset in terms of *Dice index*, *Jaccard index*, *Acc*, *Sen*, and *Spe*. The *Acc*, *Jaccard index*, and *Spe* values are 0.96, 0.92, and 0.97. The higher *Acc* shows that the ratio of the correctly segmented area over the ground truth is higher with respect to other methods. The percent overlap between the target mask and the predicted mask is higher. Table 3 shows the results of our suggested approaches and state-of-the-art algorithms on the ISIC 2017 dataset. The participants in the ISIC-2017 segmentation challenge segmented the borders of the lesion regardless of the lesion form. The value of the *Jaccard index*, *Sen*, *Spe*, *Dice index*, and *Acc* are 0.84, 0.95, 0.985, 0.957, and 0.9539. The percentage of positive individual pixel values that are correctly identified and the negative values that are correctly identified is higher than the other methods. The majority of the lesion regions have a darker background and a lighter boundary. As a result, the segmentation frameworks find the darker center but lacks the bright border regions. Our method outperforms all other segmentation algorithms for each skin lesion and can pick up the bright border regions.

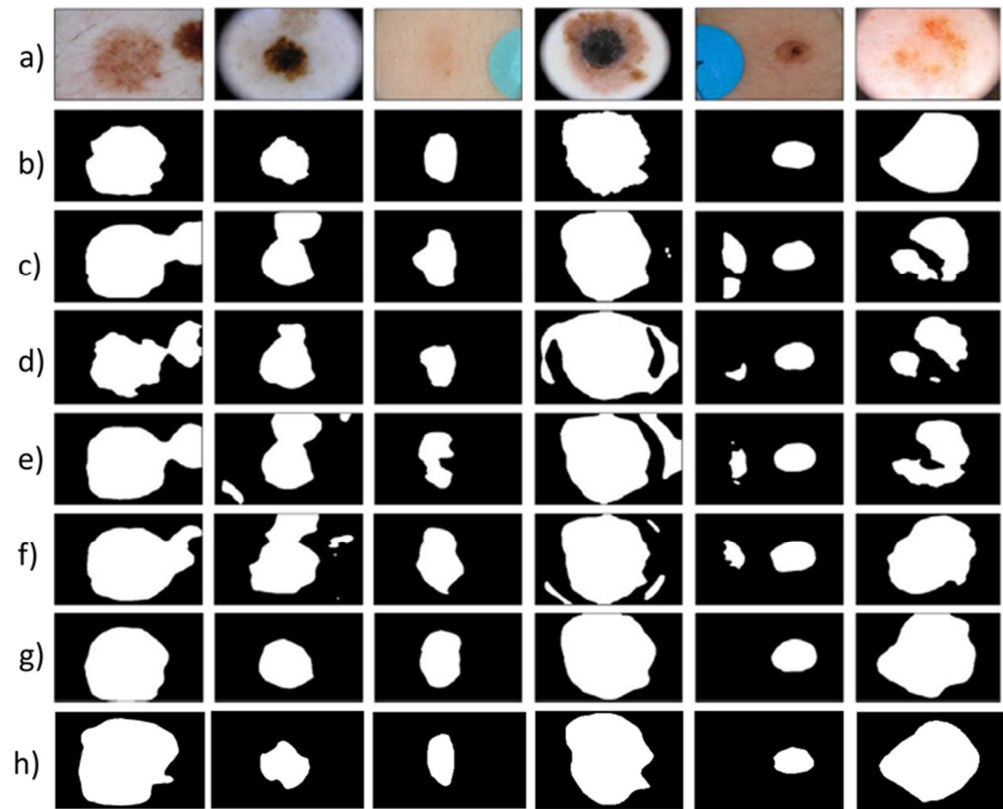


Figure 11. (a) Original input image, (b) ground truth images, (c) segmentation results of simple U-Net, (d) segmentation results of SE block on basic U-Net, (e) segmentation results of the BCDU network (with 1 dense unit), (f) segmentation results of the U-Net network (with all 64 filters), and (g) segmentation results of Att U-Net + GN + TL. (h) Results of our RL algorithm. As can be seen from the image, the results of the masks are clear. Our algorithm is able to segment the lesions with a high Acc. The other statistical measures are also higher than the other methods.

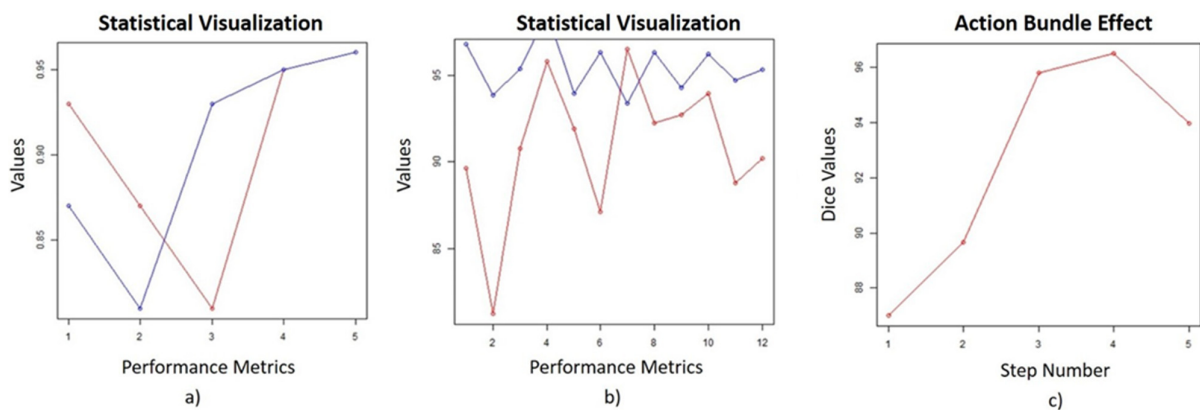


Figure 12. (a) The statistical measures used by our method such as Acc, Dice, Jaccard index, Sen, and Spe are plotted for the PH2 and the ISIC 2017 skin segmentation dataset. The blue line denotes the performance metrics plotted for the PH2 dataset and the red line shows the metrics for ISIC 2017 skin segmentation dataset. In (b), the statistical measures of our method denoted as Dice, JSI, MCC, and the overall of these statistical measures are plotted for the three categories of skin lesions: naevus, melanoma, and seborrheic keratosis. While the other blue line and the red line denotes the metrics for the PH2 and the ISIC 2017 dataset, the highlighted points indicate the metric values such as Sen, Spe, and Acc for each of the categories on both the datasets. (c) Shows the action bundle effect. The various values of K are the different settings for the action bundle. We evaluate the effect of the value K on the dice values. We infer from the results that our segmentation results improve when we use both the modified ERM and action bundle.

Table 1. The Dice index values taken for the ablation experiments. The various values of K are the different settings for the action bundle.

Statistical Measure	Basic Model	Modified ERM Included	With Action Bundle	With Both	K = 1	K = 3	K = 5	K = 7
Dice Index	93.00	93.98	94.0	95.7	93.0	93.98	95.79	94.0

Table 2. The proposed RL model's segmentation efficiency is compared to state-of-the-art methods. The statistical measures such as Dice score, Jaccard index, *Acc*, *Sen*, and *Spe* are evaluated on the PH2 dataset.

Method	Dice Score	Jaccard Index	Acc	Sen	Spe
U-Net [36]	0.89	0.81	0.94	0.93	0.94
U-Net (all 64 filters) [37]	0.90	0.81	0.94	0.93	0.95
SE_U-Net [51]	0.91	0.83	0.95	0.89	0.96
BCDU [52]	0.90	0.82	0.94	0.94	0.95
Attn_U-Net+GN [75]	0.91	0.83	0.95	0.94	0.95
FCN-16s [15]	0.88	0.80	0.91	0.93	0.88
DeepLab V3+ [51]	0.89	0.81	0.92	0.94	0.89
Mask R-CNN [48]	0.90	0.83	0.93	0.96	0.89
Ensemble-S [75]	0.93	0.90	0.83	0.96	0.92
Xie et al. [16]	0.88	0.80	0.92	0.98	0.86
Sarker et al. [44]	0.88	0.80	0.91	0.98	0.85
SLSNet [76]	0.90	0.81	0.94	0.87	0.95
Lina et al. [77]	0.87	0.79	0.94	0.88	0.95
Wang et al. [78]	0.89	0.82	0.87	0.62	0.94
Wibowo et al. [79]	0.88	0.80	0.93	0.86	0.96
Our RL algorithm (proposed)	0.94	0.92	0.96	0.9859	0.985

Table 3. Results comparison of our proposed RL method and the state-of-the-art methods on the ISIC skin lesion segmentation challenge 2017 dataset.

Method	Acc	Dice Score	Jaccard Index	Sen	Spe
First: Yading Yuan (CDNN model) [35]	0.934	0.849	0.765	0.825	0.975
Second: Matt Berseth (U-Net) [37]	0.932	0.847	0.762	0.820	0.978
U-Net [36]	0.901	0.763	0.616	0.672	0.972
SegNet [38]	0.918	0.821	0.696	0.801	0.954
FrCN [47]	0.940	0.870	0.771	0.854	0.967
Ensemble-S [75]	0.933	0.844	0.760	0.806	0.979
Xie et al. [16]	0.939	0.866	0.788	0.877	0.955
Sarker et al. [44]	0.941	0.871	0.793	0.899	0.950
SLSNet [76]	0.944	0.875	0.777	0.841	0.953
Lina et al. [77]	0.941	0.867	0.790	0.892	0.939
Wang et al. [78]	0.873	0.898	0.829	0.590	0.941
Wibowo et al. [79]	0.938	0.877	0.802	0.862	0.963
Our RL algorithm (proposed)	0.9539	0.957	0.840	0.950	0.985

The qualitative results show our model's segmentation mask performing better than the other models. The statistical measures of our RL algorithm are compared to the results of the standard algorithms, such as U-Net [36], SegNet [38], FrCN [42], etc. The *Acc*, dice values are 95.39% and 95.7%, ensuring that our algorithm results are superior to the other

state-of-the-art methods. From this, we infer that we can qualitatively and quantitatively segment skin lesions better than the other approaches.

Our findings are also compared to U-Net [36], SE U-Net [37], BCDU [52], DeepLab V3+ [51], and other methods. The SE U-Net [35] approach works in a manner analogous to the proposed technique as it simplifies the inter-class dependencies by adding several convolution layer parameters. However, it is unable to perform the segmentation of minor and complicated lesions. The value of RVD is negative and has a value of -0.0451 , which signifies our better performance compared to the other methods. In Tables 4 and 5, we compare our RL algorithm on three different lesion types: seborrheic keratosis (SK), naevus, and melanoma. It is compared with methods such as FCN-AlexNet [10], FCN16s [15], FCN 8s [41], etc. In Table 4, the algorithm outperforms other methods with an overall value of the Dice index as 93.98%, JSI as 88.79%, and MCC of 90.21%. The higher Dice and Jaccard index values indicate the better similarity of our qualitative results in comparison with the ground truth images.

Table 4. For the ISIC 2017 test set, the segmentation results of our proposed RL algorithm and the current state-of-the-art segmentation methods. The statistical measures used are dice score, Jaccard similarity index (JSI), Mathews correlation coefficient (MCC), and the melanoma category seborrheic keratosis (SK).

Method	Naevus			Melanoma			Seborrheic Keratosis			Overall		
	Dice	JSI	MCC	Dice	JSI	MCC	DICE	JSI	MCC	DICE	JSI	MCC
FCN-AlexNet [10]	85.61	77.01	82.91	75.94	64.32	70.35	75.09	63.76	71.51	82.15	72.55	78.75
FCN-32s [11]	85.08	76.39	82.29	78.39	67.23	72.70	76.18	64.78	72.10	82.44	72.86	78.89
FCN-16s [15]	85.60	77.39	82.92	79.22	68.41	73.26	75.23	64.11	71.42	82.80	73.65	79.31
FCN-8s [41]	85.33	76.07	81.73	80.08	69.58	74.39	68.01	56.54	65.14	81.06	71.87	77.81
DeepLabV3+ [51]	88.29	81.09	85.90	80.86	71.30	76.01	77.05	67.55	74.62	85.16	77.15	82.28
Mask R-CNN [48]	88.83	80.91	85.38	80.28	70.69	74.95	80.48	70.74	76.31	85.58	77.39	81.99
Ensemble-S [75]	87.93	80.46	85.58	78.45	68.42	73.61	76.88	66.62	74.05	84.42	76.03	81.51
Xie et al. [16]	88.87	81.69	85.93	83.05	74.01	77.98	81.71	72.50	77.68	86.66	78.82	83.14
Sarker et al. [42]	89.28	82.11	86.33	83.54	74.53	78.08	82.53	73.45	78.61	87.14	79.34	83.57
SLSNet [76]	86.59	78.76	79.80	92.12	79.25	79.53	86.12	74.52	77.12	88.27	77.54	78.81
Lina et al. [77]	87.12	80.35	85.14	86.25	78.69	80.25	84.35	81.32	83.25	85.90	80.12	82.88
Wang et al. [78]	88.12	79.14	80.12	89.12	77.24	80.37	86.37	83.40	81.42	87.87	79.90	80.63
Wibowo et al. [79]	86.32	79.45	81.22	85.67	76.27	80.27	85.39	79.58	79.38	85.79	78.40	80.29
Our RL algorithm	93.00	89.57	90.78	95.79	91.93	87.11	95.00	93.23	92.74	94.59	91.57	90.21

Table 5. For the ISIC 2017 skin lesion segmentation test set, we evaluated the performance of our proposed RL method and the state-of-the-art segmentation method. The statistical measures used are *Sen*, *Spe*, *Acc*, and the melanoma category is seborrheic keratosis (SK).

Method	Naevus			Melanoma			Seborrheic Keratosis			Overall		
	<i>Sen</i>	<i>Spe</i>	<i>Acc</i>	<i>Sen</i>	<i>Spe</i>	<i>Acc</i>	<i>Sen</i>	<i>Spe</i>	<i>Acc</i>	<i>Sen</i>	<i>Spe</i>	<i>Acc</i>
FCN-AlexNet [10]	82.44	97.58	94.84	72.35	96.23	87.82	71.70	97.92	89.35	78.86	97.37	92.65
FCN-32s [11]	83.67	96.69	94.59	74.36	96.32	88.94	75.80	96.41	89.45	80.67	96.72	92.72
FCN-16s [15]	84.23	96.91	94.67	75.14	96.27	89.24	75.48	96.25	88.83	81.14	96.68	92.74
FCN-8s [41]	83.91	97.22	94.55	78.37	95.96	89.63	69.85	96.57	87.40	80.72	96.87	92.52
DeepLabV3+ [51]	88.54	97.21	95.67	77.31	96.37	89.65	74.59	98.55	90.06	83.34	97.25	93.66
Mask R-CNN [48]	87.25	96.38	95.32	78.63	95.63	89.31	82.41	94.88	90.85	84.84	96.01	93.48
Ensemble-S [75]	84.74	97.98	95.58	73.35	97.30	88.40	71.80	98.58	89.91	80.58	97.94	93.33
Xie et al. [16]	90.93	95.74	95.51	83.40	95.00	90.61	85.81	94.74	91.34	88.70	95.45	93.93
Sarker et al. [42]	92.08	95.37	95.59	84.62	94.20	90.85	87.48	94.41	91.72	89.93	95.00	94.08
SLSNet [76]	86.23	94.22	93.61	85.94	93.65	92.52	84.18	94.21	93.81	85.45	94.02	93.44
Lina et al. [77]	87.22	94.25	93.14	85.56	93.57	92.58	86.38	94.12	91.22	86.38	93.98	92.31
Wang et al. [78]	63.54	93.25	86.54	66.51	94.31	85.62	68.05	93.72	84.33	66.03	93.76	85.49
Wibowo et al. [79]	86.25	95.29	92.56	87.12	94.32	91.29	86.32	93.25	90.98	86.56	94.28	91.61
Our RL algorithm	96.79	98.60	96.33	93.96	98.59	95.39	93.39	98.60	94.27	96.25	98.50	95.33

Table 5 shows the overall values for the various categories of the lesions are 96.25% for *Sen*, *Spe* as 94.71%, and *Acc* as 95.33%. Thus, we conclude that our RL algorithm on the benchmark datasets has enhanced the segmentation *Acc*, thus outperforming the state-of-the-art methods, increasing 7%, 8%, 9%, respectively, for Dice index, specificity, and Jaccard index, respectively. The other statistical measures are also higher for our RL methods. Consequently, the computational complexity of the one-initialized Q-learning

algorithm achieves a goal state with a target reward representation and finishes in less than $O(en)$ steps. S denotes a finite state set, and $G \in S$ is the goal states non-empty set. $A(s)$ is the actions finite set that are taken in execution such that $s \in S$. where $e = \sum_{s \in S} |A(s)|$. Our algorithm can solve complex problems of identifying skin lesions, which is very difficult using conventional techniques. The technique achieves long-term results and corrects the error that occurs during the training process. Once our model has corrected an error, there is less chance that it occurs again.

In case the training dataset is absent, the RL algorithm can learn from its experiences. It maintains a balance between exploitation and exploration, where exploration is the process of exploring the most rewarding ways to reach the target. In contrast, exploitation is searching for the most optimal solutions to see if they are better than the solution that has been tried before.

5. Conclusions

This paper proposes an effective multi-step approach for skin lesion segmentation based on a deep reinforcement-learning algorithm. The segmentation process is proposed as a Markov decision process and is solved by training an agent to segment the region of interest using a deep reinforcement-learning algorithm. The agent follows a set of serial actions for the region delineation. The action is defined as a set of continuous parameters. The segmentation accuracy is boosted further by using enhanced replay memory and the action bundle as a hyperparameter. The outcomes of the experiments demonstrate that the proposed reinforcement learning method yields good results. In the future, this method can be used for other medical image segmentation tasks and other forms of diagnostic imaging. The proposed approach can also detect small irregular-shaped objects or objects with no fixed geometry in the segmentation task. The statistical results infer the better performance of our reinforcement-learning algorithm on the datasets; thus, outperforming the state-of-the-art methods with an increase of 7%, 8%, 9%, respectively, for Dice index, specificity, and Jaccard index, respectively. The other statistical measures, such as accuracy and MCC, also rank higher than the other methods. Thus, the reinforcement-learning model proposed can learn with ease how to segment complex skin lesion images.

Author Contributions: Conceptualization: J.W., J.J. and I.A.A.; Methodology, Experiments: U.A.U. and A.R.; Writing—original draft preparation: U.A.U.; Writing—review and editing: J.W. and J.J.; supervision, project administration: J.W., J.J., I.A.A. and A.R. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by Yayasan UTP Prestigious Scholarship (YUTP) under Universiti Teknologi Petronas with cost centre 015LC0-281.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Not applicable.

Acknowledgments: This work was supported by Yayasan UTP Prestigious Scholarship (YUTP) under Universiti Teknologi Petronas.

Conflicts of Interest: The authors declare that they have no established conflicting financial interests or personal relationships that may have affected the work stated in this paper.

References

1. Siegel, R.L.; Miller, K.D.; Jemal, A. Cancer statistics, 2016. *CA Cancer J. Clin.* **2016**, *66*, 7–30. [\[CrossRef\]](#)
2. Esteva, A.; Kuprel, B.; Novoa, R.A.; Ko, J.; Swetter, S.M.; Blau, H.M.; Thrun, S. Dermatologist-level classification of skin cancer with deep neural networks. *Nature* **2017**, *542*, 115–118. [\[CrossRef\]](#)
3. Kroemer, S.; Frühauf, J.; Campbell, T.M.; Massone, C.; Schwantzer, G.; Soyer, H.P.; Hofmann-Wellenhof, R. Mobile teledermatology for skin tumour screening: Diagnostic accuracy of clinical and dermoscopic image tele-evaluation using cellular phones. *Br. J. Dermatol.* **2011**, *164*, 973–979. [\[CrossRef\]](#)

4. Alves, J.; Moreira, D.; Alves, P.; Rosado, L.; Vasconcelos, M.J.M. Automatic focus assessment on dermoscopic images acquired with smartphones. *Sensors* **2019**, *19*, 4957. [[CrossRef](#)]
5. Ngoo, A.; Finnane, A.; McMeniman, E.; Soyer, H.P.; Janda, M. Fighting melanoma with smartphones: A snapshot of where we are a decade after app stores opened their doors. *Int. J. Med. Inform.* **2018**, *118*, 99–112. [[CrossRef](#)] [[PubMed](#)]
6. Stolz, W. ABCD rule of dermatoscopy: A new practical method for early recognition of malignant melanoma. *Eur. J. Dermatol.* **1994**, *4*, 521–527.
7. Hazen, B.P.; Bhatia, A.C.; Zaim, T.; Brodell, R.T. The clinical diagnosis of early malignant melanoma: Expansion of the ABCD criteria to improve diagnostic sensitivity. *Dermatol. Online J.* **1999**, *5*, 3. [[PubMed](#)]
8. Argenziano, G.; Fabbrocini, G.; Carli, P.; Giorgi, V.D.; Sammarco, E.; Delfino, M. Epiluminescence microscopy for the diagnosis of doubtful melanocytic skin lesions: Comparison of the ABCD rule of dermatoscopy and a new 7-point checklist based on pattern analysis. *Arch. Dermatol.* **1998**, *134*, 1563–1570. [[CrossRef](#)] [[PubMed](#)]
9. Pehamberger, H.; Steiner, A.; Wolff, K. In vivo epiluminescence microscopy of pigmented skin lesions. I. Pattern analysis of pigmented skin lesions. *J. Am. Acad. Dermatol.* **1987**, *17*, 571–583. [[CrossRef](#)]
10. Yu, L.; Chen, H.; Dou, Q.; Qin, J.; Heng, P.A. Automated melanoma recognition in dermoscopy images via very deep residual networks. *IEEE Trans. Med. Imaging* **2016**, *36*, 994–1004. [[CrossRef](#)]
11. Liu, L.; Mou, L.; Zhu, X.X.; Mandal, M. Automatic skin lesion classification based on mid-level feature learning. *Comput. Med. Imaging Graph.* **2020**, *84*, 101765. [[CrossRef](#)] [[PubMed](#)]
12. Codella, N.C.; Gutman, D.; Celebi, M.E.; Helba, B.; Marchetti, M.A.; Dusza, S.W.; Kallou, A.; Liopyris, K.; Mishra, N.; Kittler, H. Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC). In Proceedings of the 2018 IEEE 15th International Symposium on Biomedical Imaging (ISBI 2018), Washington, DC, USA, 4–7 April 2018; pp. 168–172.
13. Li, Y.; Shen, L. Skin lesion analysis towards melanoma detection using deep learning network. *Sensors* **2018**, *18*, 556. [[CrossRef](#)] [[PubMed](#)]
14. Singh, V.K.; Abdel-Nasser, M.; Rashwan, H.A.; Akram, F.; Pandey, N.; Lalande, A.; Presles, B.; Romani, S.; Puig, D. FCA-net: Adversarial learning for skin lesion segmentation based on multi-scale features and factorized channel attention. *IEEE Access* **2019**, *7*, 130552–130565. [[CrossRef](#)]
15. Yang, X.; Zeng, Z.; Yeo, S.Y.; Tan, C.; Tey, H.L.; Su, Y. A novel multi-task deep learning model for skin lesion segmentation and classification. *arXiv* **2017**, arXiv:1703.01025.
16. Xie, Y.; Zhang, J.; Xia, Y.; Shen, C. A mutual bootstrapping model for automated skin lesion segmentation and classification. *IEEE Trans. Med. Imaging* **2020**, *39*, 2482–2493. [[CrossRef](#)] [[PubMed](#)]
17. Humayun, J.; Malik, A.S.; Kamel, N. Multilevel thresholding for segmentation of pigmented skin lesions. In Proceedings of the IEEE International Conference on Imaging Systems and Techniques, Penang, Malaysia, 17–18 May 2011; pp. 310–314.
18. Mirikharaji, Z.; Hamarneh, G. Star shape prior in fully convolutional networks for skin lesion segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 737–745.
19. Kaymak, R.; Kaymak, C.; Ucar, A. Skin lesion segmentation using fully convolutional networks: A comparative experimental study. *Expert Syst. Appl.* **2020**, *161*, 113742. [[CrossRef](#)]
20. Rother, C.; Kolmogorov, V.; Blake, A. “GrabCut” interactive foreground extraction using iterated graph cuts. *ACM Trans. Graph.* **2004**, *23*, 309–314. [[CrossRef](#)]
21. Wang, G.; Li, W.; Zuluaga, M.A.; Pratt, R.; Patel, P.A.; Aertsen, M.; Doel, T.; David, A.L.; Deprest, J.; Ourselin, S. Interactive medical image segmentation using deep learning with image-specific fine tuning. *IEEE Trans. Med. Imaging* **2018**, *37*, 1562–1573. [[CrossRef](#)]
22. Boykov, Y.Y.; Jolly, M.P. Interactive graph cuts for optimal boundary & region segmentation of objects in ND images. In Proceedings of the Eighth IEEE International Conference on Computer Vision, Vancouver, BC, Canada, 7–14 July 2001; Volume 1, pp. 105–112.
23. Xie, N.; Zhao, T.; Tian, F.; Zhang, X.H.; Sugiyama, M. Stroke-based stylization learning and rendering with inverse reinforcement learning. In Proceedings of the Twenty-Fourth International Joint Conference on Artificial Intelligence, Buenos Aires, Argentina, 25–31 July 2015.
24. Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; Riedmiller, M. Playing atari with deep reinforcement learning. *arXiv* **2013**, arXiv:1312.5602.
25. Lillicrap, T.P.; Hunt, J.J.; Pritzel, A.; Heess, N.; Erez, T.; Tassa, Y.; Silver, D.; Wierstra, D. Continuous control with deep reinforcement learning. *arXiv* **2015**, arXiv:1509.02971.
26. Wong, A.; Scharcanski, J.; Fieguth, P. Automatic skin lesion segmentation via iterative stochastic region merging. *IEEE Trans. Inf. Technol. Biomed.* **2011**, *15*, 929–936. [[CrossRef](#)] [[PubMed](#)]
27. Riaz, F.; Naeem, S.; Nawaz, R.; Coimbra, M. Active contours-based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images. *IEEE J. Biomed. Health Inform.* **2018**, *23*, 489–500. [[CrossRef](#)] [[PubMed](#)]
28. Abbas, Q.; Fondón, I.; Sarmiento, A.; Celebi, M.E. An improved segmentation method for non-melanoma skin lesions using active contour model. In Proceedings of the International Conference Image Analysis and Recognition, Vilamoura, Portugal, 22–24 October 2014; pp. 193–200.

29. Tang, J. A multi-direction GVF snake for the segmentation of skin cancer images. *Pattern Recognit.* **2009**, *42*, 1172–1179. [[CrossRef](#)]
30. Jafari, M.H.; Samavi, S.; Soroushmehr, S.M.R.; Mohaghegh, H.; Karimi, N.; Najarian, K. Set of descriptors for skin cancer diagnosis using non-dermoscopic color images. In Proceedings of the 2016 IEEE International Conference on Image Processing (ICIP), Phoenix, AZ, USA, 25–28 September 2016; pp. 2638–2642.
31. Ali, A.R.; Couceiro, M.S.; Hassenian, A.E. Melanoma detection using fuzzy C-means clustering coupled with mathematical morphology. In Proceedings of the 14th International Conference on Hybrid Intelligent Systems, Mubarak Al-Abdullah, Kuwait, 14–16 December 2014; pp. 73–78.
32. He, K.; Zhang, X.; Ren, S.; Sun, J. Deep residual learning for image recognition. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 770–778.
33. Huang, G.; Liu, Z.; Maaten, L.V.D.; Weinberger, K.Q. Densely connected convolutional networks. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 4700–4708.
34. Maninis, K.K.; Caelles, S.; Pont-Tuset, J.; Gool, L.V. Deep extreme cut: From extreme points to object segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 616–625.
35. Jafari, M.H.; Karimi, N.; Nasr-Esfahani, E.; Samavi, S.; Soroushmehr, S.M.R.; Ward, K.; Najarian, K. Skin lesion segmentation in clinical images using deep learning. In Proceedings of the 23rd International Conference on Pattern Recognition (ICPR), Cancun, Mexico, 4–8 December 2016; pp. 337–342.
36. Ronneberger, O.; Fischer, P.; Brox, T. U-net: Convolutional networks for biomedical image segmentation. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Munich, Germany, 5–9 October 2015; pp. 234–241.
37. Berseth, M. ISIC 2017-skin lesion analysis towards melanoma detection. *arXiv* **2017**, arXiv:1703.00523.
38. Chang, H. Skin cancer reorganization and classification with deep neural network. *arXiv* **2017**, arXiv:1703.00534.
39. Liu, L.; Mou, L.; Zhu, X.X.; Mandal, M. Skin Lesion Segmentation based on improved U-net. In Proceedings of the IEEE Canadian Conference of Electrical and Computer Engineering (CCECE), Edmonton, AB, Canada, 5–8 May 2019; pp. 1–4.
40. Abhishek, K.; Hamarneh, G.; Drew, M.S. Illumination-based transformations improve skin lesion segmentation in dermoscopic images. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops, Seattle, WA, USA, 14–19 June 2020; pp. 728–729.
41. Yuan, Y. Automatic skin lesion segmentation with fully convolutional-deconvolutional networks. *arXiv* **2017**, arXiv:1703.05165.
42. Al-Masni, M.A.; Al-Antari, M.A.; Choi, M.T.; Han, S.M.; Kim, T.S. Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks. *Comput. Methods Programs Biomed.* **2018**, *162*, 221–231. [[CrossRef](#)]
43. Bi, L.; Kim, J.; Ahn, E.; Kumar, A.; Feng, D.; Fullham, M. Step-wise integration of deep class-specific learning for dermoscopic image segmentation. *Pattern Recognit.* **2019**, *85*, 78–89. [[CrossRef](#)]
44. Sarker, M.M.K.; Rashwan, H.A.; Akram, F.; Banu, S.F.; Saleh, A.; Singh, V.K.; Chowdhury, F.U.; Abdulwahab, S.; Romani, S.; Radeva, P. SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks. In Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Granada, Spain, 16–20 September 2018; pp. 21–29.
45. Long, J.; Shelhamer, E.; Darrell, T. Fully convolutional networks for semantic segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Boston, MA, USA, 7–12 June 2015; pp. 3431–3440.
46. Karthik, R.; Gupta, U.; Jha, A.; Rajalakshmi, R.; Menaka, R. A deep supervised approach for ischemic lesion segmentation from multimodal MRI using Fully Convolutional Network. *Appl. Soft Comput.* **2019**, *84*, 105685. [[CrossRef](#)]
47. Milletari, F.; Navab, N.; Ahmadi, S.A. V-net: Fully convolutional neural networks for volumetric medical image segmentation. In Proceedings of the Fourth International Conference on 3D Vision (3DV), Stanford, CA, USA, 25–28 October 2016; pp. 565–571.
48. Cheng, T.; Wang, X.; Huang, L.; Liu, W. Boundary-preserving mask r-cnn. In Proceedings of the European Conference on Computer Vision, Glasgow, UK, 23–28 August 2020; pp. 660–676.
49. Kim, M.; Woo, S.; Kim, D.; Kweon, I.S. The devil is in the boundary: Exploiting boundary representation for basis-based instance segmentation. In Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision, Waikola, HI, USA, 5–9 January 2021; pp. 929–938.
50. Castrejon, L.; Kundu, K.; Urtasun, R.; Fidler, S. Annotating object instances with a polygon-rnn. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Honolulu, HI, USA, 21–26 July 2017; pp. 5230–5238.
51. Chen, L.C.; Zhu, Y.; Papandreou, G.; Schroff, F.; Adam, H. Encoder-decoder with atrous separable convolution for semantic image segmentation. In Proceedings of the European Conference on Computer Vision (ECCV), Munich, Germany, 8–14 September 2018; pp. 801–818.
52. Dai, J.; He, K.; Sun, J. Instance-aware semantic segmentation via multi-task network cascades. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 3150–3158.
53. Lu, H.; Kondo, M.; Li, Y.; Tan, J.; Kim, H.; Murakami, S.; Aoki, T.; Kido, S. Supervoxel graph cuts: An effective method for ggo candidate regions extraction on CT images. *IEEE Consum. Electron. Mag.* **2019**, *9*, 61–66. [[CrossRef](#)]
54. Lu, H.; Li, B.; Zhu, J.; Li, Y.; Li, Y.; Xu, X.; He, L.; Li, X.; Li, J.; Serikawa, S. Wound intensity correction and segmentation with convolutional neural networks. *Concurr. Comput. Pract. Exp.* **2017**, *29*, e3927. [[CrossRef](#)]

55. Yoshino, Y.; Miyajima, T.; Lu, H.; Tan, J.; Kim, H.; Murakami, S.; Aoki, T.; Tachibana, R.; Hirano, Y.; Kido, S. Automatic classification of lung nodules on MDCT images with the temporal subtraction technique. *Int. J. Comput. Assist. Radiol. Surg.* **2017**, *12*, 1789–1798. [[CrossRef](#)]
56. Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; Klimov, O. Proximal policy optimization algorithms. *arXiv* **2017**, arXiv:1707.06347.
57. Jagadeesan, S.; Subbiah, J. Real-time personalization and recommendation in Adaptive Learning Management System. *J. Ambient. Intell. Humaniz. Comput.* **2020**, *11*, 4731–4741. [[CrossRef](#)]
58. Kim, D.; Lee, T.; Kim, S.; Lee, B.; Youn, H.Y. Adaptive packet scheduling in IoT environment based on Q-learning. *Procedia Comput. Sci.* **2018**, *141*, 247–254. [[CrossRef](#)]
59. Alansary, A.; Oktay, O.; Li, Y.; Folgoc, L.L.; Hou, B.; Vaillant, G.; Kamnitsas, K.; Vlontzos, A.; Glocker, B.; Kainz, B. Evaluating reinforcement learning agents for anatomical landmark detection. *Med. Image Anal.* **2019**, *53*, 156–164. [[CrossRef](#)]
60. Caicedo, J.C.; Lazebnik, S. Active object localization with deep reinforcement learning. In Proceedings of the IEEE International Conference on Computer Vision, Santiago, Chile, 7–13 December 2015; pp. 2488–2496.
61. Zhu, Y.; Mottaghi, R.; Kolve, E.; Lim, J.J.; Gupta, A.; Fei-Fei, L.; Farhadi, A. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In Proceedings of the IEEE International Conference on Robotics and Automation (ICRA), Marina Bay Sands, Singapore, 29 May–3 June 2017; pp. 3357–3364.
62. Liu, F.; Li, S.; Zhang, L.; Zhou, C.; Ye, R.; Wang, Y.; Lu, J. 3DCNN-DQN-RNN: A deep reinforcement learning framework for semantic parsing of large-scale 3D point clouds. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 5678–5687.
63. Rao, Y.; Lu, J.; Zhou, J. Attention-aware deep reinforcement learning for video face recognition. In Proceedings of the IEEE International Conference on Computer Vision, Venice, Italy, 22–29 October 2017; pp. 3931–3940.
64. Sahba, F.; Tizhoosh, H.R.; Salama, M.M. Application of reinforcement learning for segmentation of transrectal ultrasound images. *BMC Med. Imaging* **2008**, *8*, 8. [[CrossRef](#)]
65. Watkins, C.J.; Dayan, P. Q-learning. *Mach. Learn.* **1992**, *8*, 279–292. [[CrossRef](#)]
66. Wang, Z.; Sarcar, S.; Liu, J.; Zheng, Y.; Ren, X. Outline objects using deep reinforcement learning. *arXiv* **2018**, arXiv:1804.04603.
67. Song, G.; Myeong, H.; Lee, K.M. Seednet: Automatic seed generation with deep reinforcement learning for robust interactive segmentation. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Salt Lake City, UT, USA, 18–23 June 2018; pp. 1760–1768.
68. Tschandl, P.; Rosendahl, C.; Kittler, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Sci. Data* **2018**, *5*, 180161. [[CrossRef](#)]
69. Finlayson, G.D.; Trezzi, E. Shades of gray and colour constancy. In Proceedings of the Color and Imaging Conference, Scottsdale, AZ, USA, 9–12 November 2004; Volume 2, pp. 37–41.
70. Silver, D.; Lever, G.; Heess, N.; Degris, T.; Wierstra, D.; Riedmiller, M. Deterministic policy gradient algorithms. In Proceedings of the International conference on machine learning, Beijing, China, 21–26 June 2014; pp. 387–395.
71. Sutton, R.S.; McAllester, D.A.; Singh, S.P.; Mansour, Y. Policy gradient methods for reinforcement learning with function approximation. In Proceedings of the Advances in Neural Information Processing Systems, Virtual, 6–12 December 2000; pp. 1057–1063.
72. Xu, X.; Lu, H.; Song, J.; Yang, Y.; Shen, H.T.; Li, X. Ternary adversarial networks with self-supervision for zero-shot cross-modal retrieval. *IEEE Trans. Cybern.* **2019**, *50*, 2400–2413. [[CrossRef](#)]
73. Ioffe, S.; Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. In Proceedings of the International Conference on Machine Learning, Lille, France, 6–11 July 2015; pp. 448–456.
74. Shi, W.; Caballero, J.; Huszár, F.; Totz, J.; Aitken, A.P.; Bishop, R.; Rueckert, D.; Wang, Z. Real-time single image and video super-resolution using an efficient sub-pixel convolutional neural network. In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, Las Vegas, NV, USA, 27–30 June 2016; pp. 1874–1883.
75. Arora, R.; Raman, B.; Nayyar, K.; Awasthi, R. Automated skin lesion segmentation using attention-based deep convolutional neural network. *Biomed. Signal Process. Control.* **2021**, *65*, 102358. [[CrossRef](#)]
76. Sarker, M.M.K.; Rashwan, H.A.; Akram, F.; Singh, V.K.; Banu, S.F.; Chowdhury, F.U.; Choudhury, K.A.; Chambon, S.; Radeva, P.; Puig, D. SLSNet: Skin lesion segmentation using a lightweight generative adversarial network. *Expert Syst. Appl.* **2021**, *183*, 115433. [[CrossRef](#)]
77. Liu, L.; Tsui, Y.Y.; Mandal, M. Skin lesion segmentation using deep learning with auxiliary task. *J. Imaging* **2021**, *7*, 67. [[CrossRef](#)] [[PubMed](#)]
78. Wang, X.; Jiang, X.; Ding, H.; Zhao, Y.; Liu, J. Knowledge-aware Deep Framework for Collaborative Skin Lesion Segmentation and Melanoma Recognition. *Pattern Recognit.* **2021**, *120*, 108075. [[CrossRef](#)]
79. Wibowo, A.; Purnama, S.R.; Wirawan, P.W.; Rasyidi, H. Lightweight encoder-decoder model for automatic skin lesion segmentation. *Inform. Med. Unlocked* **2021**, *25*, 100640. [[CrossRef](#)]