Karbala International Journal of Modern Science

Volume 8 | Issue 3 Article 20

Skin Lesion Segmentation based on U-Shaped Network

Muna Khalaf
University of Baghdad, munakd_comp@csw.uobaghdad.edu.iq

Ban N. Dhannoon
Al-Nahrain University

Skin Lesion Segmentation based on U-Shaped Network

Abstract

Skin lesion segmentation is an essential step toward accurate skin lesion diagnosis. The need to automate Skin lesion segmentation on the one hand, and the challenges it faces, on the other hand, have made it a growing area of research and focus. Automation of skin lesion segmentation helps reduce the effort and time needed for diagnosis and treatment and helps make better utilization of available data and shared experiences. The challenges faced by the automation of skin lesion segmentation can be broadly defined by (but not limited to); variations in texture, shape, and size for skin lesions and the low contrast between the lesion and surrounding skin.

The rise of deep learning has significantly improved the semantic segmentation results in medical imaging. U-Net structure with encoder and decoder approach is one of the most successful deep learning models for medical image segmentation. This paper introduces two models based on U-shaped structures: AlexUnet and AlexUnet+.

AlxUnet is a light U-Net model with an encoder based on pre-trained AlexNet on the ImageNet database. It significant-ly reduces memory consumption and the number of parameters, thus reducing the required FLOPS by eight times. In Alexunet+, another encoder was added to the AlxUnet structure that used pre-trained VGG11 on ImageNet. It is al-lowed to aggregate the feature maps obtained from two encoders to be used in the decoder.

AlxUnet and AlxUnet+ models were evaluated using three publicly available databases provided by the International Skin Imaging Collaboration, ISIC 2016, ISIC 2017, and ISIC 2018. Sensitivity, specificity, Jaccard similarity index, and dice similarity were used as performance metrics. Then, obtained structures were compared with U-Net, and many deep learning segmentation networks that were recently built for skin lesion segmentation. AlxUnet outperformed U-Net and produced acceptable results compared with the other networks. AlexUnet+ produced a more robust result and outperformed other networks.

Keywords

Deep Learning, Segmentation, skin Lesion, U-Net, Pre-trained

RESEARCH PAPER

Skin Lesion Segmentation Based on U-shaped Network

Muna Khalaf [a,*], Ban N. Dhannoon [b]

[a] University of Baghdad, Baghdad, Iraq
[b] Al-Nahrain University, Baghdad, Iraq

Abstract

Skin lesion segmentation is an essential step toward accurate skin lesion diagnosis. The need to automate Skin lesion segmentation on the one hand, and the challenges it faces, on the other hand, have made it a growing area of research and focus. Automation of skin lesion segmentation helps reduce the effort and time needed for diagnosis and treatment and helps make better utilization of available data and shared experiences. The challenges faced by the automation of skin lesion segmentation can be broadly defined by (but not limited to); variations in texture, shape, and size for skin lesions and the low contrast between the lesion and surrounding skin.

The rise of deep learning has significantly improved the semantic segmentation results in medical imaging. U-Net structure with encoder and decoder approach is one of the most successful deep learning models for medical image segmentation. This paper introduces two models based on U-shaped structures: AlexUnet and AlexUnet+.

AlxUnet is a light U-Net model with an encoder based on pre-trained AlexNet on the ImageNet database. It significantly reduces memory consumption and the number of parameters, thus reducing the required FLOPS by eight times. In Alexunet+, another encoder was added to the AlxUnet structure that used pre-trained VGG11 on ImageNet. It is allowed to aggregate the feature maps obtained from two encoders to be used in the decoder.

AlxUnet and AlxUnet + models were evaluated using three publicly available databases provided by the International Skin Imaging Collaboration, ISIC 2016, ISIC 2017, and ISIC 2018. Sensitivity, specificity, Jaccard similarity index, and dice similarity were used as performance metrics. Then, obtained structures were compared with U-Net, and many deep learning segmentation networks that were recently built for skin lesion segmentation. AlxUnet outperformed U-Net and produced acceptable results compared with the other networks. AlexUnet + produced a more robust result and outperformed other networks.

Keywords: Deep learning, Segmentation, Skin lesion, U-net, Pre-trained

1. Introduction

In several countries, skin cancer is one of the most common types of cancer, and its incidence rate has risen in recent years [1]. With these rising rates, it is become more crucial than ever to take preventive actions to lower the incidence and promote the early diagnosis abilities of these cancers [2]. An early diagnosis is essential in increasing survival rates, even with the most deadly form, melanoma [3] Computer-aided diagnosis helps in the early diagnosing of skin cancer [4].

Skin lesion segmentation is the primary step for Computer-aided diagnosis systems [4,5]. Accurate skin lesion segmentation from surrounding tissues and distinguishing lesions leads to more accurate skin lesion classification [1,6]. However, it is still a challenging process, mainly because skin lesions vary in colors, sizes, shapes, and texture. Meanwhile, there is low contrast between the lesion and surrounding tissue or the existence of hair in some cases, resulting in fuzzy and irregular borders during identification and diagnosis. Fig. 1 illustrates the main challenges of automatic skin lesion segmentation from the ISIC 2017 dataset [7].

Received 23 March 2022; revised 19 June 2022; accepted 20 June 2022.
Available online 1 August 2022

* Corresponding author.
E-mail addresses: munakd_comp@csw.uobaghdad.edu.iq (M. Khalaf), ban.n.dhannoon@nahrainuniv.edu.iq (B.N. Dhannoon).

https://doi.org/10.33640/2405-609X.3248

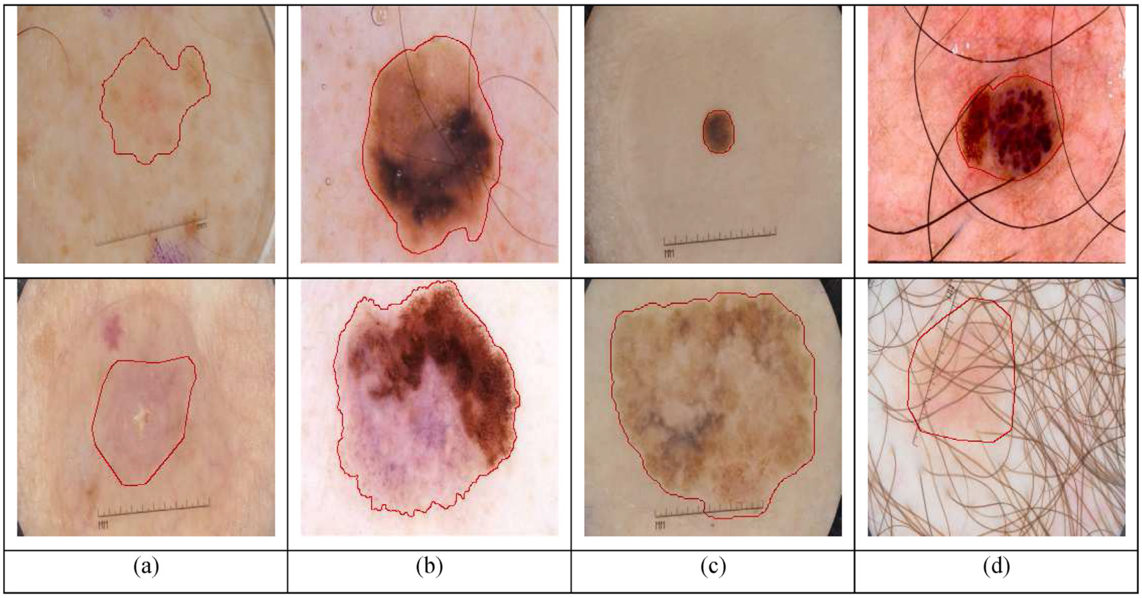

Fig. 1. Examples of main challenging cases of skin lesions: (a) low contrast, (b) variation in color, (c) variation in size, and (d) hair existence [7].

Deep learning made tremendous advancements in medical image processing during the 2000s [8]. The expression "deep" usually alludes to the number of hidden layers within the neural net, where each layer can be viewed as an individual algorithm [9].

The rapid growth in graphical processing units (GPUs) [10] on the one hand, and the available medical imaging.

Data sets for training [11], on the other hand, has led to enabling deep learning to form advanced techniques in medical image processing [12]. Medical image segmentation based on deep learning has recently proven successful by outperforming traditional techniques [13].

Since the U-Net appeared to the public in 2015 [14], it is considered one of the essential structures of deep learning in medical image segmentation, where it is the core segmentation structure in most medical image segmentation procedures and is used as the primary compression method [15].

U-Net, take its name from the "U" shaped structure, where it contains two parts; downsampling part encoder and upsampling part decoder, and there is a skip connection between them. Skip connection feeds the output of decoder layers as the input to corresponding decoder layers. Concatenation is the skip connection that is key to improving segmentation performance, that allows obtaining more information from feature maps, but at the same time, it is considered memory-consuming [14]. This paper presents two end-to-end deep learning structures that help in reducing memory consumption for skin lesion segmentation. They are based on a "U" shaped structure and use pre-trained decoders: AlxUnet and AlxUnet+.

AlxUnet consists of an encoder based on AlexNet [16] pre-trained on the ImageNet dataset and corresponding decoder. It benefits from the structure of AlexNet, which has eight layers with a large receptive field within the first and second layers [16]. This structure allowed AlxUnet to use fewer resources than U-Net by five times in terms of memory consumption, and by ten times in terms of the number of needed parameters, in the same time, it outperforms the U-Net performance in terms of skin lesion segmentation. Another encoder was added to the structure in AlexUnet+, which used pre-trained VGG11 [17] on ImageNet. We aggregated the feature maps from two encoders, which enables obtaining more information from feature maps. Alexunet + presents better accuracy in skin lesion segmentation when compared with many networks based on deep learning, and many variations of U-Net, as we will discuss later on within this paper.

2. Related works

Automated skin lesion segmentation methods can be categorized into two main categories: conventional methods and methods based on deep learning. Conventional methods like Histogram thresholding methods [17,18,19], edge-based methods [20], or Unsupervised clustering methods [21] use low-level features only to segment lesions [21,22,23]. These methods cannot handle nor process the low contrast and hair existence challenges, so they perform unacceptably [24]. Recently, methods based on deep learning have proven their success in skin lesion segmentation [25], as they have previously proven successful in different computer vision problems [25,26,27]. Its success came from the ability to extract high-level semantic features from images [28].

Several deep learning approaches for the segmentation of skin lesions were proposed by researchers, with significant contributions. Peng Tang et al. [5] proposed a U-Net model that utilized separable convolutional as the main block followed by a post-processing algorithm filling in the holes (FITH); separable convolution allows using fewer parameters with fewer computations resulting in shorter processing and diagnosis time. The stochastic weight averaging strategy is usually utilized to get more generalizations.

Manu Goyal et al. [24] presented the ensembles deep learning method based on MaskR-CNN was used in instance segmentation, and DeeplabV3+ was used in semantic segmentation, and they used pre and post-processing to enhance accuracy.

Md. Kamrul Hasan et al. [29] adopted the U-shaped approach with an encoder that mimics the DenseNet architecture. They use depth-wise separable convolution instead of standard convolution to reduce the number of parameters in their model.

Debesh Jha et al. [30] proposed the DoubleU-Net model, which integrates two U-Nets to capture more semantic segmentation. First U-Net uses pre-trained VGG-19 as the encoder, and the second U-Net adds to the bottom. They adopted Atrous Spatial Pyramid Pooling (ASPP) in each U-Net to capture contextual information that leads to enhanced segmentation performance.

Baiying Lei et al. [31] presented a generative adversarial network (GAN), a combination between two modules; dense convolution U-Net that captures fine-grained features and dual discrimination to enhance the reliability of the decision, both are used together to decide jointly.

Şaban Öztürk and Umut Özkaya [32], inspired by a fully convolutional network (FCN) used in semantic segmentation, presented an improved FCN (iFCN) with two parts encoder, which slowly downsamples features map and a decoder that use deconvolution to take the feature maps back to the actual input image size. They utilized different color spaces in their network by choosing the most efficient channel in each space for skin lesion segmentation. Tong et al. [33] proposed an improved U-Net by using three types of attention mechanisms; attention gate, spatial attention, and channel attention, which allows the model to capture more contextual information and spatial correlation between features to improve segmentation performance.

Lina Liu et al. [34] proposed a model that performed edge prediction as the auxiliary task to improve the main lesion segmentation task. They are performed simultaneously where The intermediate feature maps of each task are passed into the subblocks of the other task, guiding the model to focus on the segmentation task.

Shan and Yan. [35] proposed that spatial and channel attention network (SCA-Net) improves U-Net with two attention models that allow the capture of more contextual information from feature maps.

S. Chen et al. [36] proposed a more efficient and deeper U-Net called R2AU-Net, which utilizes recurrent residual convolutional as a basic block, and a robust skip connection by attention gets. This combination leads to obtaining more accurate segmentation.

Qamar et al. [37] presented a model based on a U-Shaped structure, which complained DenseNet, ResNet, and ASPP approaches to capture more contextual information, and at the same time, allow the decoder reconstructs fine-grained details.

Rania Ramadan and Salih Aly [38] proposed three novel U-Net versions, including single, dual, and triple encoder sub-networks coupled with a single decoder. Each encoder sub-network is assigned its own color space. To build a segmented image map, a channel-wise attention module combines the learned feature maps from each encoder sub-network.

Duwei Dai et al. [39] proposed a novel multi-scale residual network for efficiently obtaining reliable segments for a variety of skin lesion types. They used a multi-scale residual encoding and decoding fusion module that fuses multi-scale features adaptively.

The proposed methods in this paper share with many of the methods based on deep learning that we reviewed earlier in that they are based on the U-shaped network and build to skin lesion segmentation. For that, comparisons will be made between them to prove the effectiveness, efficiency, and feasibility of the network's performance we are proposing in this paper.

3. Materials and methods

This section will explain our methods and the databases we used to evaluate our methods with the Evaluation metrics and implementation details.

3.1. Databases

The most common, well-known, and publicly available three image datasets for Skin Lesion Segmentation were adopted; ISIC_ 2016 [40], ISIC_2017 [7], and ISIC_2018 [41]. These datasets were provided by the International Skin Imaging Collaboration (ISIC). Those were acquired from various devices used at prominent clinical centers on a global scale. Additionally, each image in the datasets has ground truth annotation by dermatologists. They usually contain artifacts such as ink markings, rulers, air bubbles, and ebony frames, which are challenging for the segmentation task [21,22,32]. Table 1 summarizes the training and testing size for the three datasets, with their resolution.

3.2. The proposed methods

Two methods for automatic skin lesion segmentation were proposed; AlexUnet and AlexUnet+. AlexUnet follows the U-Shaped structure. It consists of two paths: encoder and decoder with a skip connection between each layer in the encoder to its corresponding layer in the decoder, as shown in Fig. 2.

The first six layers were adopted from AlxNet [16] in the encoder and built its corresponding decoder. Encoder weights are initialized from pre-trained AlxNet on the ImageNet dataset, then fine-tuning weights using skin lesion datasets.

Each encoder layer includes four layers; convolution (Conv), rectified linear unit (ReLU) activation function, and max pooling (MP). Each decoder layer has transposed convolutions (TConv), which double the feature map size by concatenating it with the corresponding encoder layer. Afterward, this concatenation is followed by convolution, batch normalization (BN), and Relu layers.

The first layer is influential for the performance of AlexUnet; using a large kernel size (11) in the first layer allows for a larger effective receptive field, which is essential in classification and localization tasks that are needed in semantic segmentation.

At the same time, setting the stride to four in the convolution of the first layer reduced the resolution feature map to about a quarter (≈25%), thus, decreasing memory consumption in skip connection and decreasing the number of parameters needed in the subsequent layers.

AlexUnet outperforms U-Net and many of its variations. Although AlexUnet is considered a lightweight network compared to U-Net, it significantly decreases the number of parameters, memory consumption, and flops, as illustrated in Table 2.

In AlexUnet+, AlexUnet was modified to obtain better skin lesion segmentation by adding another coder based on the VGG11 network. AlexUnet + contains two encoders that feed into one decoder. Feature Aggregation (FA) block (illustrated in Fig. 3) was added to concatenate the feature maps from two encoders.

In the FA block at first, using 1 × 1 convolution (Conv1x1) to downsample the number of feature maps from the VGG11 encoder layer (CY) to equalize the number of features maps coming from the AlxNet encoder (CX) layer, then concatenating these feature maps (C2X), then using 1 × 1 convolution to downsample the number of concatenated feature maps to (CX). 1 × 1 convolution is used to reduce dimension, increase the depth of the network, and allow learnable interactions among channels in feature maps.

Table 1. Datasets information.

Dataset	ISIC 2016	ISIC 2017	ISIC 2018
Resolution	(556 × 679) To (4499 × 6748) Pixels		
Training Size	900	2000	2076
Testing Size	379	600	520

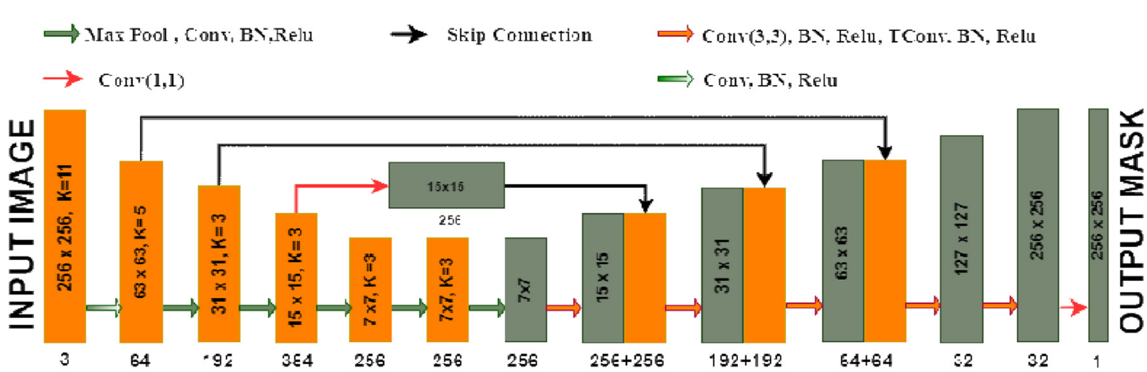

Fig. 2. The proposed AlexUnet architecture. Each box represents a multi-channel feature map. The kernels are shown below each box, while the resolution of kernels and kernel size (k) is shown inside the box (if needed).

Table 2. The resources used in the proposed networks compared to U-Net.

Model	*Memory consumption for Input size (MB): 0.75*	Number of FLOPS	Number of parameters
U_Net [14]	1008.62	40,160,985,088	17,266,241
AlexUnet	195.63	2,711,067,936	8,759,073
AlexUnet+	692.18	13,106,794,592	26,111,393

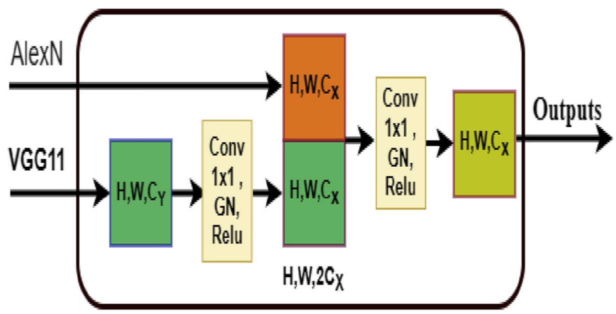

Fig. 3. Feature aggregation (FA), AlxN represents the feature map from the AlexNet encoder, and VGG11 represents the feature map from the VGG11 encoder. Hight, width, and number of channels are represented as H, W, and C, respectively.

Aggregated features from two encoders with different receptive fields capture more information that helps enhance the skin lesion segmentation performance. Encoder weights were initialized from pre-trained AlexNet and VGG11 on the ImageNet dataset, then fine-tuned them using skin lesion datasets. AlexUnet+ is faster than U-Net because it needs floating-point operations (flops) and several parameters, as illustrated in Table 2. AlexUnet + structure is illustrated in Fig. 4.

AlexNet and VGGNet are early convolutional neural network models that depend mainly on the traditional convolution layersDespite their simple architecture, they achieve high classification accuracy, which makes them ideal for our methods. They can also be useful for determining how pre-training affects the model's overall performance due to their simple designs.

3.3. Evaluation metrics

Overlap-based criteria were adopted to assess the performance of skin lesion segmentation. They are; Dice Score (DC), Jaccard Similarity (JS), Accuracy (AC), Sensitivity (SE), and Specificity (SP). Dice Score and Jaccard index are the primary metrics in skin lesion segmentation [7]. DC quantifies the similarity between predicted segmentation results and ground truth labels by computing the ratio between the size of their intersection over the average of their size, which is computed as in equation (1):

$$DC = \frac{2*TP}{2*TP + FN + FP} \quad (1)$$

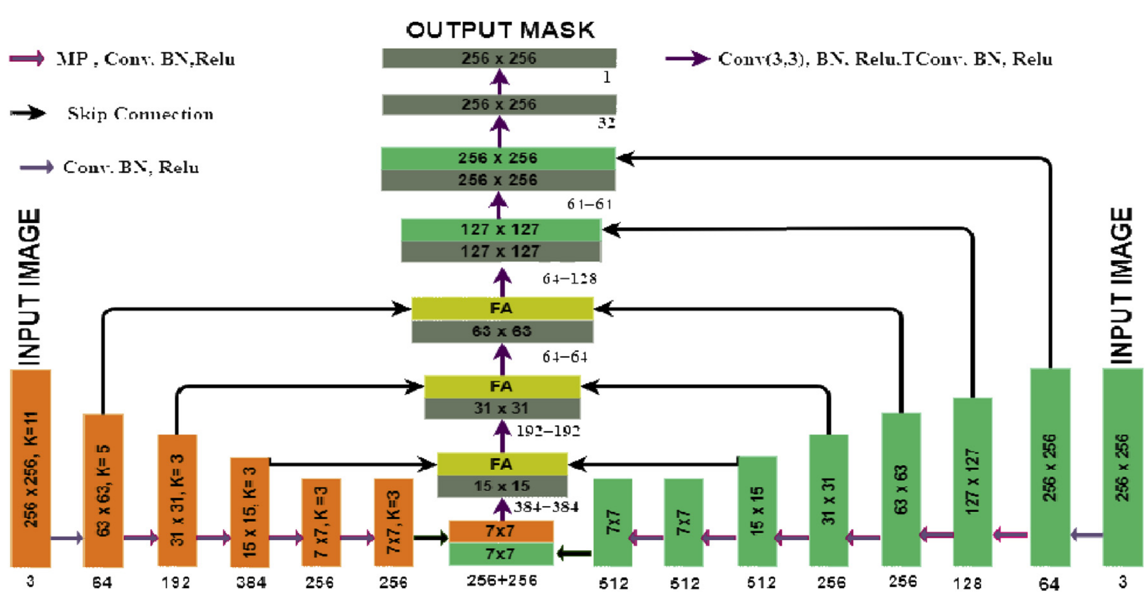

Fig. 4. An illustration of the proposed architecture AlxUnet+. Each box represents a multi-channel feature map. The number of channels is shown below each box. Resolution of the feature map is shown inside the box.

JS is the intersection of the predicted segmentation results and ground truth regions over their union, which can be computed as in equation (2):

$$JS = \frac{TP}{TP + FN + FP} \tag{2}$$

AC, SE, and SP are used as additional indicators. The formulas for these metrics criteria are:

$$AC = \frac{\mathrm{TP} + \mathrm{TN}}{\mathrm{TP} + \mathrm{FP} + \mathrm{TN} + \mathrm{FN}} \tag{3}$$

$$SP = \frac{\mathrm{TN}}{\mathrm{FP} + \mathrm{TN}} \tag{4}$$

$$SE = \frac{\mathrm{TP}}{\mathrm{TP} + \mathrm{FN}} \tag{5}$$

where TP, FP, TN, and FN represent the number of pixels in the predicted image that are True Positive, False Positive, True Negative, and False Negative, compared with the ground-truth image.

3.4. Implementation details

Our experiments were implemented utilizing the PyTorch framework [40] in Google's Colaboratory environment. For the ISIC-2017 and ISIC-2018 training sets, models were trained using batch size 16 for 50 epochs and scaled down the input image's resolution to 256 × 256 Pixels.

With the ISIC-2016, the same approach was used with 100 epochs. The Adam optimizer was used with its default parameters [41]. We started with a 0.001 learning rate and then reduced it if the Dise score metric stopped improving during seven epochs. The Generalized Dice Loss function given here [42] was used because of its suitability for an imbalanced dataset. During training, data augmentation was used by randomly flipping vertical and horizontal data.

4. Performance

In this section, the impact of pre-training on the proposed models' performance will be discussed. Also the comparison between them and other models for skin lesion segmentation based on deep learning.

4.1. Ablation study

An ablation study was performed on the proposed networks on the training and testing sets for the ISIC datasets, with pretraining and from scratch with the same number of epochs and the same hyperparameter. DS and JS were used for the evaluation testing set for each dataset because they are elementary metrics to represent the segmentation performance. The quantitative results presented in Table 3 illustrate that the proposed networks achieved a better score on all datasets with pretraining. The dice score curves are shown in Fig. 5, which illustrates the influence of pretraining on performance. Each curve represents the behavior of training and testing, with and without pretraining on each network.

It is noticeable in all curves that fast convergence occurred in the pretraining model, which is logical because the low-level feature captured from pretraining on the ImageNet dataset (containing a million images) is effective in the learning process.

Pretraining is a simple change that allows for better performance and enables the networks to achieve faster convergence. This influence is more evident with ISIC-2016 because its training set is smaller than other datasets.

4.2. Comparison with state-of-the-art methods

Proposed methods were compared with other deep learning methods recently built for skin lesion segmentation, using the ISIC-2016, ISIC-2017, and ISIC-2018 test data.

The quantitive results are summarized in Table 4 for ISIC-2016 shows that AlexUnet + outperforms other.

Methods in all metrics, except for a small ratio in SE, While AlexUnet outperformed the U-Net, although it needs remarkably fewer resources than U-Net, as mentioned in Table 1. AlexUnet + also outperformed the ASCU-Net, as both are based on the U-Shaped structure; however, ASCU-Net consists of a more complex design based on the attention techniques, which are memory consumption and computational cost.

Table 5 summarizes the performance of the proposed methods with the U-Net and some of the methods based on deep learning for skin lesion

Table 3. The performance for our networks with/without pretraining on ImageNet.

DataSet	Model	DS		JS	
		Pre-trained	Scratch	Pre-trained	Scratch
ISIC-2016	AlexUnet	0.919	0.89	0.8577	0.83
	AlexUnet+	0.94	0.91	0.892	0.85
ISIC-2017	AlexUnet	0.866	0.84	0.78	0.75
	AlexUnet+	0.89	0.872	0.809	0.78
ISIC-2018	AlexUnet	0.923	0.91	0.87	0.852
	AlexUnet+	0.934	0.92	0.88	0.86

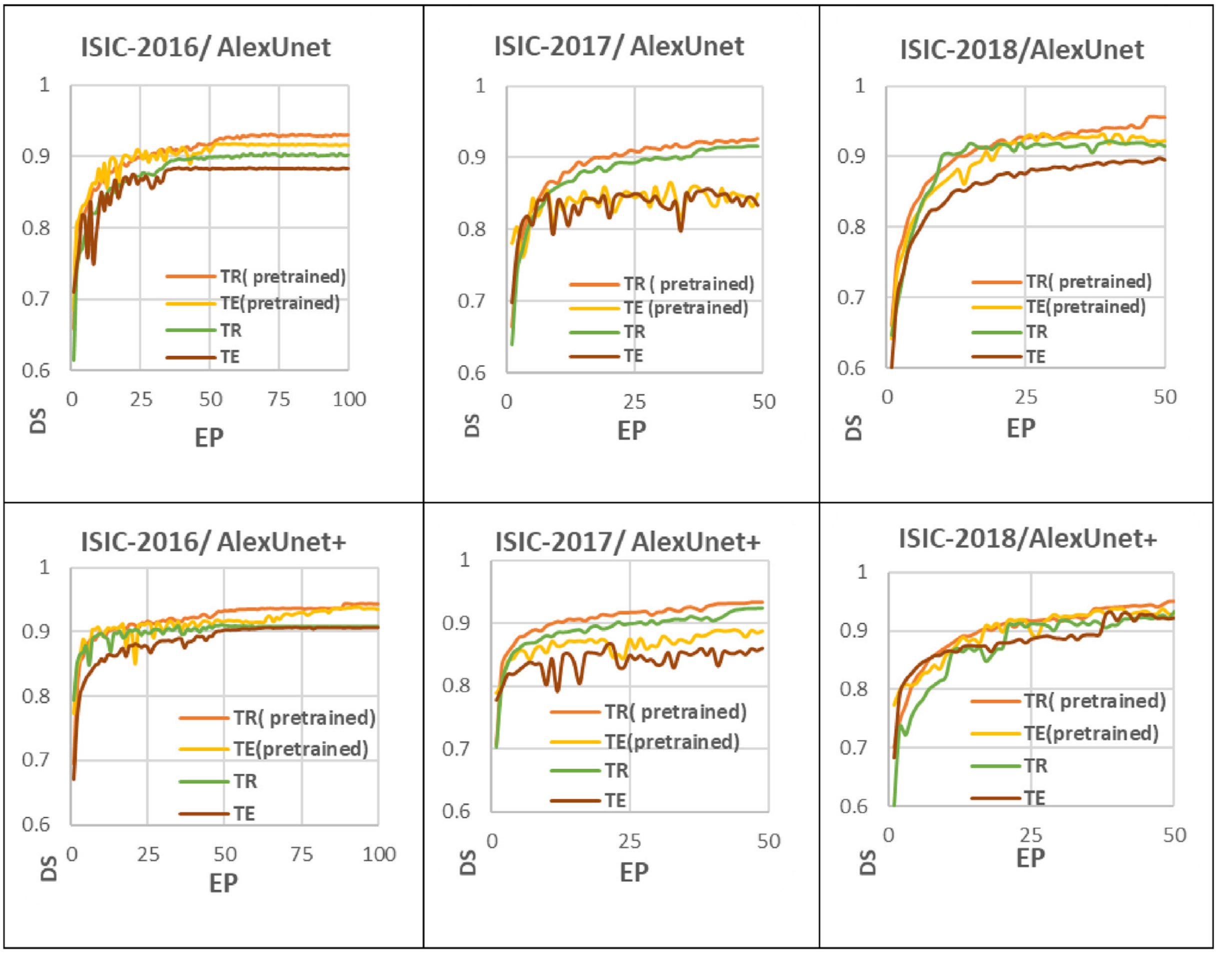

Fig. 5. Dice Score (DS) curves for ISIC datasets (ISIC-2017, ISIC-2016, and ISIC-2018) training on the proposed networks (AlexUnet and AlexUnet+) where TR represents the training set, TE represents the testing set, and EP represents the number of epoch.

segmentation that used the ISIC-2017 dataset. AlexUnet + outperformed the other networks in the most crucial segmentation metrics, DS and JS. AlexUnet presented an acceptable performance compared to the remaining networks with more complex structures. It also outperformed the other methods of the SP metric.

Table 6 summarizes the performance of the proposed methods with U-Net and other recently introduced methods based on deep learning for skin lesion segmentation that used the **ISIC-2018** dataset. Once again, AlexUnet + outperformed other methods by all metrics in the exception of the SE. AlexUnet performed well by outperforming many of

Table 4. Segmentation results compared with deep learning networks using ISIC-2016 test data.

ISIC-2016	DS	JS	AC	SP	SE
U-Net [14]	0.889	0.812	0.943	0.962	0.907
Separable-Unet [5]	0.93	**0.892**	0.971	0.956	**0.947**
DAGAN [31]	0.931	0.871	0.960	0.968	0.937
ASCU-Net [33]	0.908	0.845	0.954	0.961	0.927
Ms RED [39]	0.92	0.83	0.96	—	—
AlxUnet	0.919	0.858	0.958	0.965	0.93
AlxUnet+	**0.94**	**0.892**	**0.972**	**0.969**	0.946

Table 5. Segmentation results compared with deep learning networks using ISIC-2017 test data.

ISIC-2017	DS	JS	AC	SP	SE
U-Net [14]	0.765	0.62	0.91	0.973	0.68
Separable- Unet [5]	0.869	0.792	0.943	0.956	0.895
Ensemble DL [24]	0.871	0.79	0.94	0.95	**0.899**
DSNet [29]	—	0.775	—	0.955	0.875
DAGAN [31]	0.859	0.771	0.935	0.976	0.835
iFCN [32]	0.886	0.783	**0.953**	**0.98**	0.854
ASCU-Net [33]	0.830	0.742	0.926	0.965	0.825
DL_AT [34]	0.871	0.795	0.943	0.965	0.888
TICU-Net [38]	0.853	0.748	0.931	—	—
Ms RED [39]	0.86	0.78	0.94	—	—
AlxUnet	0.866	0.78	0.94	**0.98**	0.845
AlxUnet+	**0.89**	**0.809**	0.95	0.975	0.895

Table 6. Segmentation results compared with deep learning networks using ISIC-2018 test data.

ISIC-2018	DS	JS	AC	SP	SE
U-Net [14]	0.86	0.90	0.91	0.96	0.85
DAGAN [31]	0.885	0.824	0.92	0.91	0.953
DoubleU-Net [30]	—	0.878	—	—	0.861
Sca_net [35]	0.929	—	—	—	0.873
IBA-U-Net [42]	0.872	0.829	0.976	0.944	0.931
Dense En-De [37]	0.90	0.833	0.969	**0.97**	**0.965**
ASCU-Net [33]	0.830	0.825	0.965	0.926	0.742
R2AU-Net [36]	0.866	0.821	0.969	0.928	0.896
Ms RED [39]	0.89	0.83	0.96	0.90	0.91
AlxUnet	0.923	0.92	0.97	0.96	0.87
AlxUnet+	**0.934**	**0.94**	**0.98**	**0.97**	0.88

these methods, although it is considered a light network.

The quantitative results in results in Tables 4−6 show the effectiveness of the AlexUnet and AlexUnet + structures at skin lesion segmentation in three.

ISIC datasets. AlexUnet depends on merging large and small kernels to extract high-level semantic and low-level features. Their combination aids in semantic segmentation accuracy. AlexUnet + used the benefits of the AlexUnet structure and added another encoder to strengthen the encoder features before feeding to the decoder.

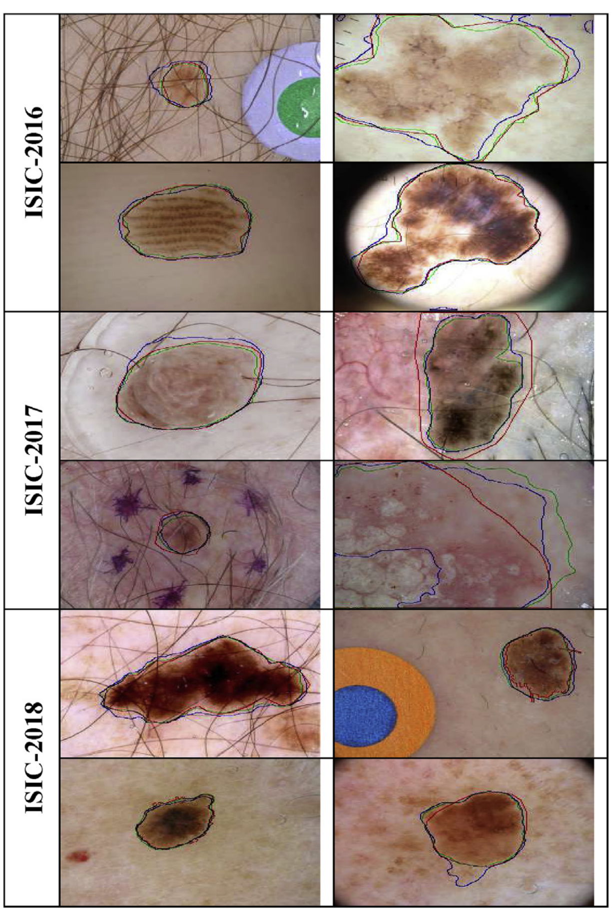

Fig. 6. Segmentation Results of some examples of skin lesions from ISIC-2016, ISIC-2017, and ISIC-2018 test set that are produced by AlexUnet(blue line), AlexUnet+(green line) compared with ground-truth mask(red line).

Figure 6 illustrates some qualitative examples of the performance of the proposed Networks in ISIC-2016, ISIC-2017, and ISIC-2018. It demonstrates their ability to overcome challenges found in each dataset, such as hair existence, dark corner artifacts, and low contrast.

5. Conclusion

The diagnosis is greatly aided by accurate segmentation of skin lesions. AlexUnet and AlexUnet+, two end-to-end deep learning networks for skin lesion segmentation, have been suggested and implemented in this research. Both used a U-shaped structure and used the ImageNet dataset to pre-train their encoders. AlexUnet is a light network that uses the AlexNet encoder. Another encoder from the VGG11 network was incorporated (to AlexUnet) to boost the fusion of multi-scale features in AlexUnet+. ISIC-2016, ISIC-2017, and ISIC-2018 are three well-known demanding datasets that were used to evaluate the performance of our networks. AlexUnet provided an acceptable outcome and outperformed many more difficult deep learning models, while AlexUnet + outperformed other state-of-the-art models, according to the findings. To improve semantic segmentation, our methods rely on a combination of large kernel size and multi-scale feature fusion. Furthermore, these approaches can be improved using fine-tuning weights rather than starting from scratch. We believe that this evidence will be useful in additional medical image segmentation tasks in the future.

Conflicts of interest

There is no conflict of interest.

References

[1] R.B. Oliveira, J.P. Papa, A.S. Pereira, J.M.R.S. Tavares, Computational methods for pigmented skin lesion classification in images: review and future trends, Neural Comput Appl. 29 (2018) 613−636, https://doi.org/10.1007/s00521-016-2482-6.

[2] M.H. Trager, L.J. Geskin, F.H. Samie, L. Liu, Biomarkers in melanoma and non-melanoma skin cancer prevention and risk stratification, Exp Dermatol. 31 (2022) 4−12, https://doi.org/10.1111/exd.14114.

[3] R.L. Siegel, K.D. Miller, A. Jemal, Cancer statistics, CA A Cancer J Clin. 68 (2018) 284−296, https://doi.org/10.3322/caac.21442.

[4] P. Mehta, B. Shah, Review on techniques and steps of computer aided skin cancer diagnosis, Procedia Comput Sci. 85 (2016) 309−316, https://doi.org/10.1016/j.procs.2016.05.238.

[5] P. Tang, Q. Liang, X. Yan, S. Xiang, W. Sun, D. Zhang, G. Coppola, Efficient skin lesion segmentation using separable-Unet with stochastic weight averaging, Comput Methods Progr Biomed. 178 (2019) 289−301, https://doi.org/10.1016/j.cmpb.2019.07.005.

[6] M.K. Hasan, M.T.E. Elahi, M.A. Alam, M.T. Jawad, R. Martí, DermoExpert: skin lesion classification using a hybrid convolutional neural network through segmentation, transfer learning, and augmentation, Inform Med Unlock. 28 (2022), 100819, https://doi.org/10.1016/j.imu.2021.100819.

[7] N.C.F. Codella, D. Gutman, M.E. Celebi, B. Helba, M.A. Marchetti, S.W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, A. Halpern, 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: IEEE 15th int. Symp. Biomed. Imaging (ISBI 2018), IEEE, 2018: pp. 168−172, https://doi.org/10.1109/ISBI.2018.8363547.

[8] G. Litjens, T. Kooi, B.E. Bejnordi, A.A.A. Setio, F. Ciompi, M. Ghafoorian, J.A.W.M. van der Laak, B. van Ginneken, C.I. Sánchez, A survey on deep learning in medical image analysis, Med Image Anal. 42 (2017) 60−88, https://doi.org/10.1016/j.media.2017.07.005.

[9] Z.K. Hussien, B.N. Dhannoon, Anomaly detection approach based on deep neural network and dropout, Baghdad Sci J. 17 (2020) 701−709, https://doi.org/10.21123/bsj.2020.17.2(SI).0701.

[10] S. Mittal, S. Vaishay, A survey of techniques for optimizing deep learning on GPUs, J Syst Architect. 99 (2019), 101635, https://doi.org/10.1016/j.sysarc.2019.101635.

[11] G. Langs, A. Hanbury, B. Menze, H. Müller, VISCERAL: towards large data in medical imaging - challenges and directions, in: MICCAI int. Work. Med. Content-based retr. Clin. Decis. Support, Springer, Berlin, Heidelberg, 2012: pp. 92−98, https://doi.org/10.1007/978-3-642-36678-9_9.

[12] M.J. Willemink, W.A. Koszek, C. Hardell, J. Wu, D. Fleischmann, H. Harvey, L.R. Folio, R.M. Summers, D.L. Rubin, M. P. Lungren, Preparing medical imaging data for machine learning, Radiology. 295 (2020) 4−15, https://doi.org/10.1148/radiol.2020192224.

[13] R. Wang, T. Lei, R. Cui, B. Zhang, H. Meng, A.K. Nandi, Medical image segmentation using deep learning: a survey, IET Image Process. 16 (2022) 1243−1267, https://doi.org/10.1049/ipr2.12419.

[14] W. Weng, X. Zhu, U-Net, Convolutional networks for biomedical image segmentation, in: Int. Conf. Med. Image comput. Comput. Interv., Springer, Cham, 2015: pp. 234−241, https://doi.org/10.1109/ACCESS.2021.3053408.

[15] L. Liu, J. Cheng, Q. Quan, F.X. Wu, Y.P. Wang, J. Wang, A survey on U-shaped networks in medical image segmentations, Neurocomputing. 409 (2020) 244−258, https://doi.org/10.1016/j.neucom.2020.05.070.

[16] A. Krizhevsky, I. Sutskever, G.E. Hinton, ImageNet classification with deep convolutional neural networks, Adv Neural Inf Process Syst. 25 (2012), https://doi.org/10.1145/3383972.3383975.

[17] K. Simonyan, A. Zisserman, Very deep convolutional networks for large-scale image recognition, in: ICLR 2015, 2015: pp. 1−14, https://doi.org/10.48550/arXiv.1409.1556.

[18] M. Emre Celebi, Q. Wen, S. Hwang, H. Iyatomi, G. Schaefer, Lesion border detection in dermoscopy images using ensembles of thresholding methods, Skin Res Technol. 19 (2013) e252−e258, https://doi.org/10.1111/j.1600-0846.2012.00636.x.

[19] R. Garnavi, M. Aldeen, M.E. Celebi, G. Varigos, S. Finch, Border detection in dermoscopy images using hybrid thresholding on optimized color channels, Comput Med Imag Graph. 35 (2011) 105−115, https://doi.org/10.1016/j.compmedimag.2010.08.001.

[20] J. Yasmin, M. Sathik, An improved iterative segmentation algorithm using canny edge detector for skin lesion border detection, Int Arab J Inf Technol. 12 (2015) 325−332, https://doi.org/10.5120/7779-0865.

[21] F. Xie, A.C. Bovik, Automatic segmentation of dermoscopy images using self-generating neural networks seeded by genetic algorithm, Pattern Recogn. 46 (2013) 1012−1019, https://doi.org/10.1016/j.patcog.2012.08.012.

[22] M. Emre Celebi, Q. Wen, H. Iyatomi, K. Shimizu, H. Zhou, G. Schaefer, A state-of-the-art survey on lesion border detection in dermoscopy images, Dermosc Image Anal. 10 (2015) 97−129, https://doi.org/10.1201/b19107.

[23] M.E. Celebi, H. Iyatomi, G. Schaefer, W.V. Stoecker, Lesion border detection in dermoscopy images, Comput Med Imag Graph. 33 (2009) 148−153, https://doi.org/10.1016/j.compmedimag.2008.11.002.

[24] M. Goyal, A. Oakley, P. Bansal, D. Dancey, M.H. Yap, Skin lesion segmentation in dermoscopic images with ensemble deep learning methods, IEEE Access. 8 (2019) 4171−4181, https://doi.org/10.1109/ACCESS.2019.2960504.

[25] A. Adegun, S. Viriri, Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art, Artif Intell Rev. 54 (2021) 811−841, https://doi.org/10.1007/s10462-020-09865-y.

[26] D.D. Himabindu, S.P. Kumar, A survey on computer vision architectures for large scale image classification using deep learning, Int J Adv Comput Sci Appl. 12 (2021) 105−120, https://doi.org/10.14569/IJACSA.2021.0121013.

[27] K. Bayoudh, R. Knani, F. Hamdaoui, A. Mtibaa, A survey on deep multimodal learning for computer vision: advances, trends, applications, and datasets, Vis Comput. (2021) 1−32, https://doi.org/10.1007/s00371-021-02166-7.

[28] J. Chai, H. Zeng, A. Li, E.W.T. Ngai, Deep learning in computer vision: a critical review of emerging techniques and application scenarios, Mach Learn Appl. 6 (2021), 100134, https://doi.org/10.1016/j.mlwa.2021.100134.

[29] M.K. Hasan, L. Dahal, P.N. Samarakoon, F.I. Tushar, R. Martí, DSNet: automatic dermoscopic skin lesion segmentation, Comput Biol Med. 120 (2022) 1−25, https://doi.org/10.1016/j.compbiomed.2020.103738.

[30] D. Jha, M.A. Riegler, D. Johansen, P. Halvorsen, H.D. Johansen, DoubleU-Net, A deep convolutional neural network for medical image segmentation, in: 2020 IEEE 33rd int. Symp. Comput. Med. Syst., IEEE, 2020: pp. 558−564, https://doi.org/10.1109/CBMS49503.2020.00111.

[31] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, Y. Xu, J. Qin, S. Chen, T. Wang, S. Wang, Skin lesion segmentation via generative adversarial networks with dual discriminators, Med Image Anal. 64 (2020), 101716, https://doi.org/10.1016/j.media.2020.101716.

[32] Ş. Öztürk, U. Özkaya, Skin lesion segmentation with improved convolutional neural network, J Digit Imag. 33 (2020) 958−970, https://doi.org/10.1007/s10278-020-00343-z.

[33] X. Tong, J. Wei, B. Sun, S. Su, Z. Zuo, P. Wu, ASCU-net : attention gate , spatial and channel attention U-net for skin lesion segmentation, Diagnostics. 11 (2021) 501, https://doi.org/10.3390/diagnostics11030501.

[34] L. Liu, Y.Y. Tsui, M. Mandal, Skin lesion segmentation using deep learning with auxiliary task, J Imag. 7 (2021) 67, https://doi.org/10.3390/jimaging7040067.

[35] T. Shan, J. Yan, SCA-Net, A spatial and channel attention network for medical image segmentation, IEEE Access. 9 (2021) 160926−160937, https://doi.org/10.1109/access.2021.3132293.

[36] Q. Zuo, S. Chen, Z. Wang, R2AU-Net: attention recurrent residual convolutional neural network for multimodal medical image segmentation, Secur Commun Network. 2021 (2021), https://doi.org/10.1155/2021/6625688.

[37] S. Qamar, P. Ahmad, L. Shen, Dense encoder-decoder−based architecture for skin lesion segmentation, Cognit Comput. 13 (2021) 583−594, https://doi.org/10.1007/s12559-020-09805-6.

[38] R. Ramadan, S. Aly, CU-net: a new improved multi-input color U-net model for skin lesion semantic segmentation, IEEE Access. 10 (2022) 15539−15564, https://doi.org/10.1109/ACCESS.2022.3148402.

[39] D. Dai, C. Dong, S. Xu, Q. Yan, Z. Li, C. Zhang, N. Luo, Ms RED: a novel multi-scale residual encoding and decoding network for skin lesion segmentation, Med, Image Anal. 75 (2022), 102293, https://doi.org/10.1016/j.media.2021.102293.
[40] D. Gutman, N.C.F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, A. Halpern, Skin lesion analysis toward melanoma detection: a challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC) 2016, Eprint ArXiv1605, 2016: pp. 3–7, 01397, http://arxiv.org/abs/1605.01397.
[41] N. Codella, V. Rotemberg, P. Tschandl, M.E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, A. Halpern, Skin lesion analysis toward melanoma detection 2018: a challenge hosted by the international skin imaging collaboration (ISIC), ArXiv Prepr. ArXiv1902, 2019, 03368, http://arxiv.org/abs/1902.03368.
[42] S. Chen, Y. Zou, P.X. Liu, IBA-U-Net: Attentive BConvLSTM U-Net with Redesigned Inception for medical image segmentation, Comput Biol Med. 135 (2021), 104551, https://doi.org/10.1016/j.compbiomed.2021.104551.