IEEE Access
Multidisciplinary | Rapid Review | Open Access Journal

Received April 8, 2022, accepted May 27, 2022, date of publication May 30, 2022, date of current version June 8, 2022.

Digital Object Identifier 10.1109/ACCESS.2022.3179390

Multiscale Attention U-Net for Skin Lesion Segmentation

MOHAMMAD D. ALAHMADI, (Member, IEEE)
Department of Software Engineering, College of Computer Science and Engineering, University of Jeddah, Jeddah 23890, Saudi Arabia
e-mail: mdalahmadi@uj.edu.sa

ABSTRACT Skin cancer is the most common type of cancer in the world and it is more treatable if diagnosed early. The diagnosis process usually starts with segmenting the skin lesion area and planning a follow-up treatment by the dermatologists. Thus, the segmentation process plays a critical role in the treatment process. In recent years, machine learning methods, especially deep convolutional neural networks are proposed to address the segmentation challenge. The common segmentation methods (e.g., U-Net) deploy a series of encoding blocks to model the local representation and subsequently a series of decoding blocks to capture the semantic relation. However, these structures are usually limited to model multi-scale objects with large variations in texture and shape. To address these limitations, we propose a Multi-Scale Attention U-Net (MSAU-Net) for skin lesion segmentation. In particular, we improve the typical U-net by inserting an attention mechanism at the bottleneck of the network to model the hierarchical representation. The attention module aggregates the multi-level representation in a non-linear fashion to selectively adjust the representative features. Then it deploys a Bidirectional Convolutional Long Short-term Memory (BDC-LSTM) structure to fetch the common discriminative features and suppress the less informative ones. We incorporate the resulted features in each block of the decoding path to highlight the important regions. We have evaluated our proposed network in three public skin lesion datasets, including ISIC 2017, ISIC 2018, and PH2 datasets. The experimental results demonstrate that the proposed pipeline outperforms the existing alternatives.

INDEX TERMS Attention mechanism, deep learning, U-net, skin cancer, segmentation.

I. INTRODUCTION

The skin is the largest organ in the body that plays important roles such as protecting the body from the outside environment, receiving sensory stimuli from the external environment, regulating body temperature through sweating, and highlighting hair growth when cold. When skin cells become disordered due to symptoms of the disease and grow out of control, they can turn into skin cancer and sometimes even spread to other parts of the body. Skin cancer is the most common type of cancer in the United States [1] and worldwide that threatens the lives of many people every year. Skin cancer can be divided into two groups, melanoma ,and non-melanoma types. Melanoma skin cancer is the most dangerous type of skin cancer and is reported as the most lethal skin cancer [2]. This type of skin cancer is the result of the unusual growth of melanocytes [3]. Melanocytes are cells located in the lower part of the skin epidermis and are responsible for making melanin pigments. Any change in the number of melanocytes or an increase or decrease in their activity causes disorders. Although melanoma skin cancer is not as common as other types of skin cancer, it is a too dangerous type of cancer due to its high spread rate to other parts of the, which holds a mortality rate of 1.62% [2]. According to the World Health Organization, approximately three million non-melanoma skin cancers and 132,000 melanoma skin cancers are recorded worldwide annually [4].

Like many cancers, the best treatment for melanoma is early detection since it is more treatable in the early stages of the disease. According to the studies [5], for the localized stage melanoma, the five-year relative survival rate is 98% which drops to about 14% in the latest stage. Therefore, rapid detection of melanoma or the suspected skin lesions is important and requires a method that can detect the disease as quickly as possible. In this regard, dermatologists use dermoscopic images to diagnose the disease. However, the examination of these images by dermatologists is not only

The associate editor coordinating the review of this manuscript and approving it for publication was Aasia Khanum.

This work is licensed under a Creative Commons Attribution 4.0 License. For more information, see https://creativecommons.org/licenses/by/4.0/

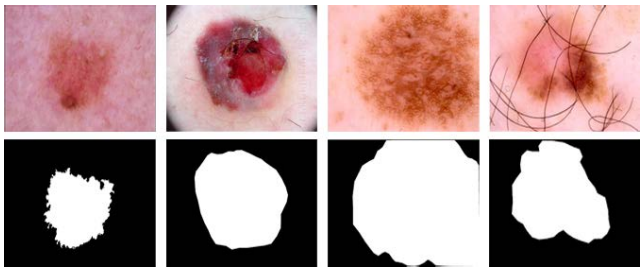

FIGURE 1. Some samples of skin lesion images along with the segmentation map generated by the deep learning model.

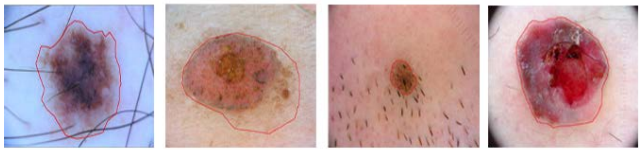

FIGURE 2. Typical challenging cases in dermoscopic images for skin lesion segmentation.

associated with a significant error rate but also is very time-consuming, and in some cases not enough specialists are available. In recent years, machine vision methods have had many applications in the examination of pathology images [6]. Among the many methods, automatic image segmentation is very useful and efficient for detecting disease [7]. In these methods, dermoscopic images are given to the deep learning model, and after processing these images by the network, places in these images that have a disease pattern appear as segmented in the output so that later dermatologists can focus directly on the disease areas and apply appropriate treatment methods. Figure 1 shows some examples of dermatology images, as inputs, and the skin lesion segmentation results generated by a deep segmentation model.

However, medical images segmentation, which separates the affected areas from other surrounding healthy tissues, is a challenging task due to some factors such as low contrast in medical images, the presence of multiple tissues that are similar, lesion sizes, color shift, and non-uniform lighting system between different laboratories. Moreover, in skin lesion segmentation,other obstacles such as body hair, air bubbles, blood vessels, ebony frames, color illumination, and patient-specific properties that may change skin colors make this task more complicated. Figure 2 shows some typical challenges in dermoscopic images [8].

Several methods have been proposed in the literature to address the semantic segmentation task in the medical domain. Among these approaches, deep-learning strategies have made significant advances in medicine, making them the best available methods for processing medical images. One of the first convolutional networks introduced for the image segmentation task is the fully convolutional network (FCN) [9]. This deep model is an end-to-end and pixels-to-pixels network for generating a semantic segmentation map through the input image. In FCNs, all fully connected layers are replaced with convolution and deconvolution layers to keep the original resolution. Ronneberger *et al.* [10] further extended the idea of FCN into a U-shape structure. This network architecture consists of symmetric encoding and decoding paths. The encoder reduces the dimensionality of input data and extracts a large number of feature maps. On the other hand, the decoder part applies a hierarchical series of up-convolutional layers to model the semantic information and produce the segmentation maps.

Many extensions of U-Net have been proposed [11]–[20] to improve its performance. These methods have tried to strengthen the original U-Net using techniques such as recurrent residual strategies, applying probabilistic functions to resolve uncertainty [21], inserting attention mechanisms, or using other non-linear functions in the convolutional layers.

Nevertheless, CNN facilitates the learning of representing abstract data, which robust the network to transfer local features. However, in semantic segmentation, the abstraction of spatial information may be undesirable. To address this issue, several methods have been proposed. Chen *et al.* [22] utilized ''Atrous spatial pyramid pooling'' (ASPP) and introduced Deeplab. This method uses several parallel ASPPs to capture contextual information at multiple scales [23]. Furthermore, The approach improved by utilizing the skip connection in the decoding path, similar to the U-Net approach. Although the pyramid representation improved the performance, it lacks to capture the common representation shared among the hierarchy of the deep model (no attention mechanism incorporated) to model robust and noise invariant features.

In recent years, attention-based techniques have been introduced to the deep models and have been widely used in various computer vision tasks [24]. Unlike conventional methods that use multiple similar feature maps, the attention strategy increases network performance, mostly in semantic segmentation tasks [24]–[27], by avoiding the use of similar feature maps and selecting the most informative features for a given task without additional supervision. In this paper, we propose a Multi-Scale Attention U-Net (MSAU-Net) for skin lesion segmentation. In particular, we improve U-net by inserting an attention mechanism at the bottleneck of the network. These attention modules aggregate the multi-level representation in a non-linear fashion to selectively adjust the representative features and finally deploy BDC-LSTM to fetch the common discriminative features and suppress the less informative ones. We incorporate the resulted features in each block of the decoding path.

We perform the attention mechanism in two steps process: firstly, to re-calibrate the feature map and pay more attention to more informative channels in each layer, we applied the channel-wise attention process. In other words, by assigning different weights to different channels of feature maps, the network focuses more on a channel with more discriminative information. In the second step, to aggregate the features extracted by the different blocks of the encoder module,

we apply the BDC-LSTM module. The objective of this module is to use the hierarchical representation to jointly encode both local and global representation into a unique transformed space, where the feature map can be used by the decoder path to effectively emphasize the informative regions. Furthermore, the hierarchical representation provided by the encoder module helps the BDC-LSTM layer for learning objects in multi-scale and multi-level. Thus, the resulted features are less sensitive to the variation in shape and texture. The main contributions of the paper are as follows:

- Multi-scale attention mechanism to capture hierarchical representation
- Including Bi-directional Convolutional LSTM module to capture discriminative features
- Significant improvement over the state-of-the-art methods

The rest of the paper is organized as follows. Section 2 reviews related work. The proposed network is presented in Section 3. The experimental results are described in Section 4. Finally, Section 5 concludes the paper.

II. RELATED WORK

The semantic segmentation task plays one of the most important roles in dermoscopic image processing. Numerous automatic and semi-automatic methods for skin segmentation have been proposed. Like other research lines in the computer vision field, skin lesion segmentation methods can be categorized into handcrafted and deep learning-based approaches. The earlier approaches focus on designing the specific feature to learn discriminative patterns from the image itself. Histogram thresholding methods [28]–[30] try to find a threshold that divides the images into two sections: skin lesions and adjacent tissues. Unsupervised color-based methods [31]–[33] try to use the color space properties of RGB dermoscopic images to determine a homogenous region for skin lesion areas and other tissues and perform segmentation accordingly. Region-merging-based approaches [34]–[36] compare neighboring regions and merge them if they are close enough in some properties. Active contour methods [37]–[39] segment lesion areas by utilizing algorithms like metaheuristic, genetic, and snake algorithms. Morphological operations-based methods [40], [41] rely on the relative ordering of pixel values for segmentation. However, these traditional image segmentation methods do not show satisfactory results and cannot overcome problems such as fuzzy lesion borders, hair artifacts, low contrast, and ebony frames.

In recent years, deep learning methods have returned to the field of artificial intelligence with more power and they have achieved outstanding results in many machine learning tasks [42], particularly semantic segmentation tasks. These deep learning methods, especially CNNs, have become standard baselines in many semantic segmentation problems. The majority of the CNNs breakthroughs are resulted from their capability of learning hierarchical as well as higher-level features that are more robust than normal raw image features. The state-of-the-art CNN segmentation architectures include but not limited to: Fully Convolutional Neural Network (FCN) [9], U-Net [10], SegNet [43], hourglass [44], and DeepLab [22]. Recently, many researchers have used CNN architecture for skin lesions semantic segmentation because of their high capability of learning diverse datasets. Some of the State-of-the-art methods based on CNNs are reviewed in the following.

Xie *et al.* [45] proposed MB-DCNN for improving skin lesion segmentation performance by using a collaboration between segmentation and classification. Each task facilitates the other in a bootstrapping way. This method mutually transfers coarse masks and location information between a coarse segmentation network (coarse-SN) and a mask-guided classification network (mask-CN). Maninis *et al.* [46] proposed a Deep Extreme Cut (DEXTR) model which combines original RGB images and extreme points (corner points on the contours) to feed the network's input. Although this method requires the input of extreme points in which their quality has an impact on the segmentation performance, they have shown this combination can improve the performance of instance segmentation. Abhishek *et al.* [47] designed a novel algorithm that improves skin lesion semantic segmentation by utilizing illumination invariant of different tissues. They combined information from illumination invariant grayscale images, specific color bands, and shading-attenuated images.

Based on the classical encoder-decoder architecture, Wu *et al.* [8] utilized a feature adaptive transformer network (FAT-Net) that effectively captures global context information and long-range dependencies by integrating an extra transformer branch. Their approach uses a feature adaptation module and a memory-efficient decoder to enhance the feature fusion between the adjacent-level features. In this regard, they activate the effective channels and restart the irrelevant background noise. The Laplacian Pyramid Super-Resolution Network (LapSRN) proposed by Lai *et al.* [48] is capable of progressively reconstructing the sub-band residuals of high-resolution images for image super-resolution. It predicts the high-frequency residuals by taking coarse-resolution feature maps as input.

Azad *et al.* [20] proposed a two-stages attention mechanism for skin lesion segmentation. They set a weight for each channel, which is determined by a set of feature maps to capture the relationship between the channels. Similar to the bi-directional strategies [14] this context gating mechanism network is capable of emphasizing more on the informative and meaningful channels. In addition, they use a second-level attention strategy to integrate the different layers of Atrous convolution, allowing the network to focus on a more goal-related field of view. Liu *et al.* [49] used auxiliary information based on the edge prediction technique for the skin lesion segmentation task. To make the network focuses on the boundary region of the segmentation task they used a cross-connection layer module. This module fed the intermediate feature maps of each task into the subblocks of the other task. They also used a multi-scale

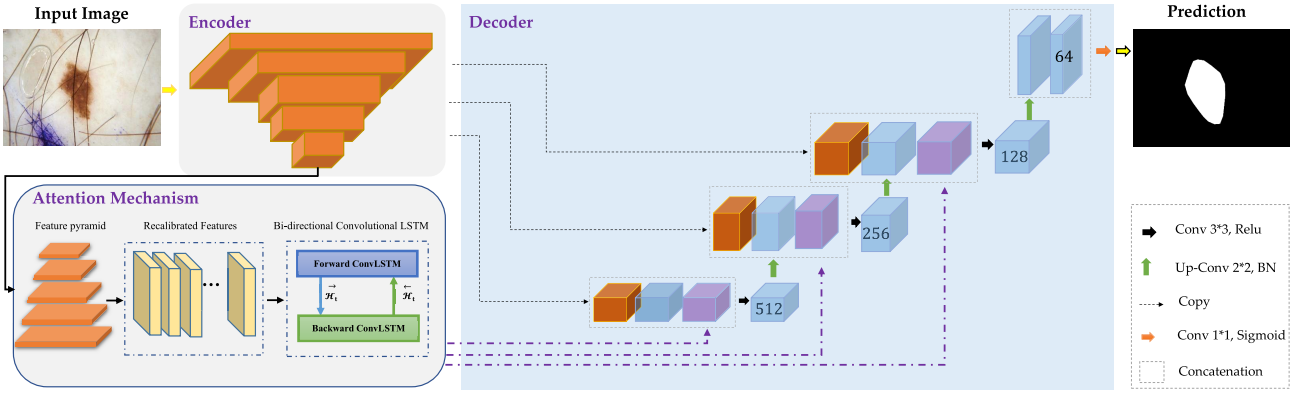

FIGURE 3. The structure of the proposed method for skin lesion segmentation. The proposed method incorporates the attention mechanism on top of the encoder module to learn hierarchical features.

feature aggregation module to increase network performance using different scale features. Dai *et al.* [50] segmented a variety of skin lesions by taking the advantage of multi-scale residual encoding and decoding fusion (MS RED) to fuse multi-scale features adaptively. Furthermore, they proposed a multi-resolution and multi-channel feature fusion module to enhance the capability of learning the feature representation. In the down-sampling stages, they used a new pooling module (Soft-pool) which retains more helpful information and enhances the segmentation performance. One central limitation of these multi-level fusion strategies is related to their poor aggregation strategies, which are not capable of combining different level features. To address this problem, we include the attention mechanism on top of the multi-level features to capture discriminative features.

III. PROPOSED METHOD

We propose MSAU-Net, attention incorporated U-Net model for skin lesion segmentation. The overview of our proposed network is shown in Figure 3. In our structure, we apply the encoder module to extract the hierarchical representation, then by utilizing the attention mechanism we perform the feature re-calibration process in a non-linear fashion. The description regarding each section of the proposed method is detailed in the following subsections.

A. ENCODER

Our proposed method utilizes a U-Net structure to model the segmentation problem. The U-Net model follows a symmetric structure and applies an encoder and decoder modules to learn the segmentation map [10]. Although the U-Net model is capable of capturing local information, its structure does not pay more attention to the boundary area [51], thus, it is less precise in separating skin lesions from the overlapped background. In other words, to accurately segment the skin lesion from other surrounding parts, both the local appearance and the entropy of the area should be learned through the training process. To model such region-sensitive representation, we include an attention mechanism on top of the encoder blocks. The purpose of the attention layer is to model the multi-scale representation and highlight the importance of each activated feature map during the recognition process [52]. The resulting feature map from the attention module can bring rich and scale-dependent descriptions, which is crucial for skin lesion segmentation tasks with various scales on the lesion patterns.

B. FEATURE RECALIBRATION

In conventional CNN networks, the resolution of the spatial feature is significantly reduced due to the use of a set of consecutive max-pooling and down-sampling functions. In addition, images can contain objects with different scales [53]. To diminish this problem, we propose to use multi-scale representation results from each block of the encoder module. In our design, we concatenate the different feature maps resulting from the encoder block to form a multi-scale representation. To scale the different feature maps into the same shape we use an Atrous convolution. To this end, on top of the last convolutional layer of each encoder block, we use Atrous operation to up-sample the representation filters. For up-sampling the filters, a hole convolutional filter applies to the full resolution image, i.e., inserting zeros between the filters' values. In this operation, the number of parameters stays constant due to the fact that non-zero filters' values are only considered in the calculations. The Atrous convolution provides a way to control the spatial resolution of feature responses. In addition, to calculate feature responses in each layer, we can enlarge the field of view of the filters, which results in a combination of larger context information. The Atrous convolution [22] for one-dimensional signal is calculated as:

$$x'[i] = \sum_{k} x[i + r.k]w[k] \tag{1}$$

where x is the input feature map, x' is the output feature map, i refers to a spatial location on y and w is a convolution filter. Moreover, r refers to the Atrous rate and determines the stride which we sample the input signal. By applying the Atrous convolution we build a feature pyramid to form a multi-scale representation (shown in Figure 4). To normalize the feature pyramid, we utilize the squeeze and excitation

module [54]. Using this strategy, the network uses the global information of the input data to selectively empathize the informative features and suppress the less useful ones. For producing each input channel's weight, the model exploits the global context information of the input features. Therefore, the global average pooling is calculated for each channel as:

$$z_f = \frac{1}{H \times W} \sum_{i}^{H} \sum_{j}^{W} x_f(i, j) \tag{2}$$

where $H \times W$ is the size of the channel, x_f is the f_{th} channel, and z_f is the output of the global average pooling. Moreover, we learn nonlinear interaction and also the non-mutually-exclusive relationship between channels at the next step. To capture the channel-wise dependencies two fully connected layers are then utilized. The output of these layers is calculated as:

$$s_f = \sigma\left(\mathbf{W}_2 \delta\left(\mathbf{W}_1 z_f\right)\right) \tag{3}$$

C. BI-DIRECTIONAL ConvLSTM

Standard LSTM uses full connections in input-to-state and state-to-state transitions which is its main disadvantage due to the fact that these networks do not consider the spatial correlation. ConvLSTM [55] has been proposed to address this problem. This method utilizes convolution operations into input-to-state and state-to-state transitions. An input gate i_t, an output gate o_t, a forget gate f_t, and a memory cell C_t form the ConvLSTM. Input, output and forget gates act as controlling gates to access, update, and clear memory cell. The ConvLSTM formula is written as follows, for simplicity we have avoided writing subscript and superscript.

$$\begin{aligned}
i_t &= \sigma\left(\mathbf{W}_{xi} * \mathcal{X}_t + \mathbf{W}_{hi} * \mathcal{H}_{t-1} + \mathbf{W}_{ci} * \mathcal{C}_{t-1} + b_i\right) \\
f_t &= \sigma\left(\mathbf{W}_{xf} * \mathcal{X}_t + \mathbf{W}_{hf} * \mathcal{H}_{t-1} + \mathbf{W}_{cf} * \mathcal{C}_{t-1} + b_f\right) \\
\mathcal{C}_t &= f_t \circ \mathcal{C}_{t-1} + i_t \tanh\left(\mathbf{W}_{xc} * \mathcal{X}_t + \mathbf{W}_{hc} * \mathcal{H}_{t-1} + b_c\right) \\
o_t &= \sigma\left(\mathbf{W}_{xo} * \mathcal{X}_t + \mathbf{W}_{ho} * \mathcal{H}_{t-1} + \mathbf{W}_{co} \circ \mathcal{C}_t + b_c\right) \\
\mathcal{H}_t &= o_t \circ \tanh\left(\mathcal{C}_t\right)
\end{aligned} \tag{4}$$

where $*$ states the convolution, and $\circ$ denotes Hadamard functions. H_t is the hidden sate tensor, and X_t is the input tensor. C_t indicates the memory cell tensor, and, W_{x*} and W_{h*} are 2D Convolution kernels corresponding to the input and hidden state, respectively. Finally, the bias terms are indicated with b_i, b_f, b_o, and b_c.

In the proposed model, we utilize BConvLSTM [56] for encoding the recalibrated feature pyramid into a single multi-scale representation. In fact, BConvLSTM consists of two ConvLSTMs, one for processing input data in the forward path and the other for processing data in the backward path direction. Unlike a standard ConvLSTM that only processes the dependencies of the forward direction, the BConvLSTM considers data dependencies in both directions and makes a decision for the current input. Cui *et al.* [57] have proved that considering both forward and backward temporal perspectives boost the network performance. Since

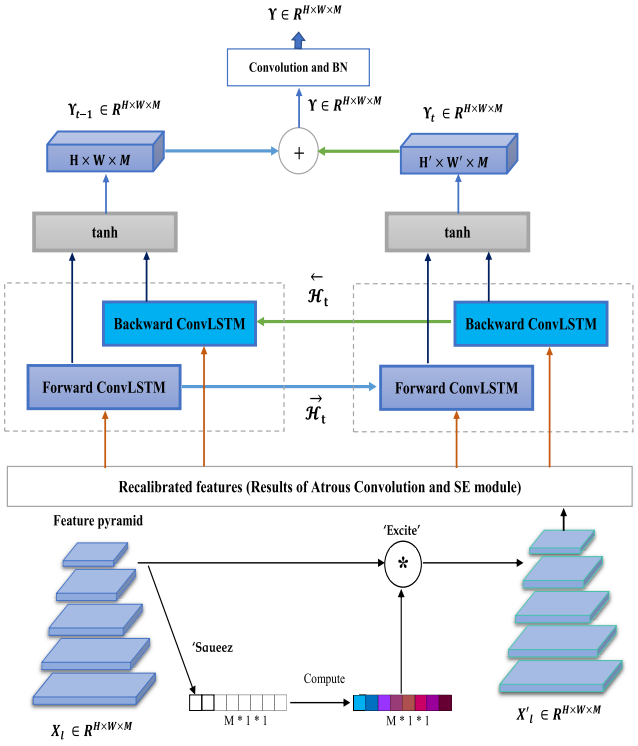

FIGURE 4. Attention mechanism proposed in our method to learn hierarchical representation. This attention mechanism applies the squeeze and excitation module to calibrate the feature pyramid based on the informative channels and then uses a bi-directional convolutional LSTM to aggregate different levels of the pyramid into a single representation.

the BConvLSTM consists of two standard ConvLSTM, we have two sets of parameters for backward and forward states. The output of the BConvLSTM is calculated as

$$\mathbf{Y}_t = \tanh\left(\mathbf{W}_y^{\overrightarrow{\mathcal{H}}} * \overrightarrow{\mathcal{H}}_t + \mathbf{W}_y^{\overleftarrow{\mathcal{H}}} \overleftarrow{\mathcal{H}}_t + b\right) \tag{5}$$

where H_t indicates the hidden state tensors for forward and H_t denotes the hidden state tensors for backward states. $Y_t \in R^{F_l \times W_l \times H_l}$ denotes the final output considering bidirectional Spatio-temporal information. b shows the bias term. We utilized hyperbolic tangent *tanh* for combining the output of both forward and backward states through a non-linear way. The detailed structure of the proposed mechanism is shown in Figure 4.

D. DECODER

In our proposed model, the decoder is implemented according to the regular U-Net. The features up-sampled from the previous decoder layer are concatenated with features that are imported directly from the encoder along with the multi-scale representation derived from the attention module. We use two Convolutional layers followed by the batch-normalization and activation layer in each block of the decoding path to learn the semantic representation. Finally, at the last decoding block, we deploy a softmax activation to produce the segmentation map.

IV. EXPERIMENTAL RESULT

In this section, we provide (i) details about the training process, (ii) the evaluation metrics we used to evaluate our approach, and (iii) a description of each dataset we used during our experimental evaluation.

A. TRAINING PROCESS

The proposed method is implemented in the Pytorch library and has been carried out on an NVIDIA RTX 3090GPU with a batch size of 8 without any data augmentation. We trained all the models with initial learning rate $1e-3$ and the decay rate $1e-4$ for 100 epochs. For model weight initialization we used a standard normal distribution, which provides a stable start point for the network. Furthermore, during the training process, in case the validation performance does not change in 10 consecutive epochs, we stop the training process. The baseline network in our experiments has the same structure as a U-Net model without the proposed attention mechanism.It is worthwhile to mention that during the training process on each dataset, the optimization algorithm steadily decreased the loss value on both train and validation sets and eventually converged to the optimal solution. Thus, we did not observe any instability during the training process.

B. EVALUATION METRICS

To experimentally evaluate our method performance, we have employed commonly well-known metrics including accuracy (AC), sensitivity (SE), specificity (SP), F1-Score, and Jaccard similarity (JS). The terminologies used to describe how metrics are calculated are given below.

True-Positive (TP) refers to the predicted label that is correctly predicted as a lesion class.

False-Positive (FP) refers to the predicted label that is falsely predicted as a lesion class.

True-Negative (TN) refers to the predicted label that is truly labelled as a background pixel.

False-Negative (FN) refers to the predicted label that is falsely labelled as a background pixel.

Accuracy shows the percentage of correct prediction,

$$ACC = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FP} + \text{FN}} \tag{6}$$

Specificity measures the proportion of FP that are correctly identified by model,

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} \tag{7}$$

Sensitivity measures the proportion of predicted TP that are correctly identified by model,

$$\text{Sensitivity/ Recall} = \frac{\text{TP}}{\text{TP} + \text{FN}} \tag{8}$$

F1 score also known as balanced F-score or F-measure, is a weighted average of the precision and recall,

$$\text{F1 score} = \frac{2 * \text{TP}}{2 * \text{TP} + \text{FP} + \text{FN}} \tag{9}$$

Jaccard similarity is also known as a mean intersection over union (mIoU) in segmentation tasks, measures the similarity between the predicted values $\hat{y}$ and real values y by comparing members of two sets to see which members are shared and which are distinct.

$$\text{Jaccard similarity} = \frac{|y \cap \hat{y}|}{|y| + |\hat{y}| - |y \cap \hat{y}|} \tag{10}$$

C. DATASETS

The proposed method was evaluated on three publicly available datasets ISIC 2017 [58], ISIC 2018 [59], PH^2 [60]. In the next subsection we will provide more details about each dataset.

1) ISIC 2017 DATASET

The International Skin Imaging Collaboration (ISIC) 2017 dataset is one of the most well-known datasets in skin cancer diagnosis. This dataset consists of 2,000 dermoscopic images of the skin taken using the technique of eliminating the surface reflection of skin that brings a deeper level of skin visualization [58]. For each instance, an expert clinician has annotated the ground-truth label, using either a semi-automated or manual process. The annotation data provides further information for three subtasks: lesion segmentation, localization, and skin disease classification. In this research work, we focus on the segmentation task. Following the literature work [13], we divided the original dataset into a training set with 1250 samples, validation sets consist of 150 samples, and a test set with 600 instances. Furthermore, we used a resize function to reduce the spatial dimension of the input data into 256*256 pixels.

2) ISIC 2018 DATASET

The ISIC 2018 dataset, like the former ISIC datasets, includes a large collection of quality-controlled dermoscopic images of skin lesions, introduced by an international collaboration to improve melanoma diagnosis [59]. This dataset contains 2594 images, each of which is accompanied by a corresponding grand truth mask. Similar to ISIC 2017, this dataset also defines three sub-tasks: lesion segmentation, lesion attribute detection, and disease classification. We have categorized the dataset into three sub-sections: train data with 1815 images, evaluation data with 259 images, and test data with 520 images. Furthermore, to reduced the computational and network training cost, we have resized the input images from 2016×3024 pixels to 256×256 pixels.

3) PH^2

The PH^2 dataset consists of 200 dermoscopic images of skin lesions region, acquired at the dermatology services of Pedro Hispano Hospital, Matosinhos, Portugal. The main objective of this dataset is to enable future researches on classification and segmentation of cancerous regions in dermoscopic images. Similar to [13], we have randomly divided the dataset into two categories of 100 instances, one

TABLE 1. Performance comparison of the proposed method vs the SOTA approaches on the ISIC 2017 dataset.

Methods	F1	SE	SP	AC	JS
Lesion Analysis [61]	0.8840	0.8250	0.9750	0.9340	0.9365
R2U-net [12]	0.8920	0.9414	0.9425	0.9424	0.9421
MCGU-Net [13]	0.8927	0.8502	0.9855	0.9570	0.9570
Baseline	0.8682	0.9479	0.9263	0.9314	0.9314
Proposed Method	**0.9032**	**0.8870**	**0.9714**	**0.9576**	**0.9576**

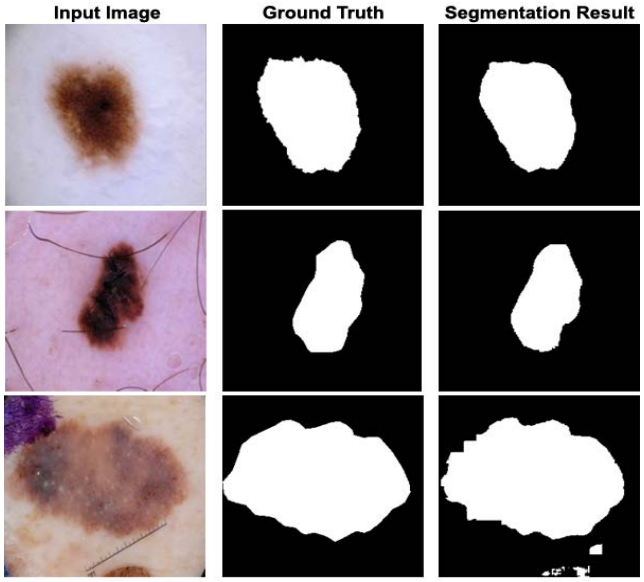

FIGURE 5. Segmentation results of the proposed method on ISIC 2017. The proposed method produces smooth segmentation result on the boundary area and separates the lesion area from the overlapped background.

of which is used as training data and the other set for the evaluation purpose.

D. RESULTS

The quantitative results of the proposed method on ISIC 2017 are illustrated in Table 1. The results show that the proposed method outperforms the state-of-the-art (SOTA) methods in almost all metrics. Compared to the recent MCGU-Net model which utilizes an attention mechanism inside the network, our strategy produces a better segmentation map which further proves the effectiveness of our method. In Figure 5, we depicted some visualization results of the proposed method on the ISIC 2017.

We further evaluated our method on ISIC 2018 to compare the results with SOTA approaches. As clear from Table 2, our method marginally increases the performance compared to the counterpart approaches. On the other hand, incorporating the attention mechanism proposed in our paper increases the U-Net F1 score by 0.25 as it is shown in Table 2. To further demonstrate the effectiveness of the proposed method from a qualitative perspective, we provide Figure 6.

During the third experiment, we evaluated our approach on the PH^2 dataset. Obtained results compared to the SOTA strategies are shown in Table 3. We can observe that our

TABLE 2. Performance comparison on the the ISIC 2018 dataset.

Methods	F1	SE	SP	AC	PC	JS
Att U-net [26]	0.665	0.717	0.967	0.897	0.787	0.566
R2U-net [12]	0.679	0.792	0.928	0.880	0.741	0.581
Att R2U-Net [12]	0.691	0.726	0.971	0.904	0.822	0.592
MCGU-Net [13]	0.895	0.848	0.986	**0.955**	0.947	0.955
Baseline	0.647	0.708	0.964	0.890	0.779	0.549
Proposed Method	**0.896**	**0.841**	**0.979**	0.949	**0.956**	**0.956**

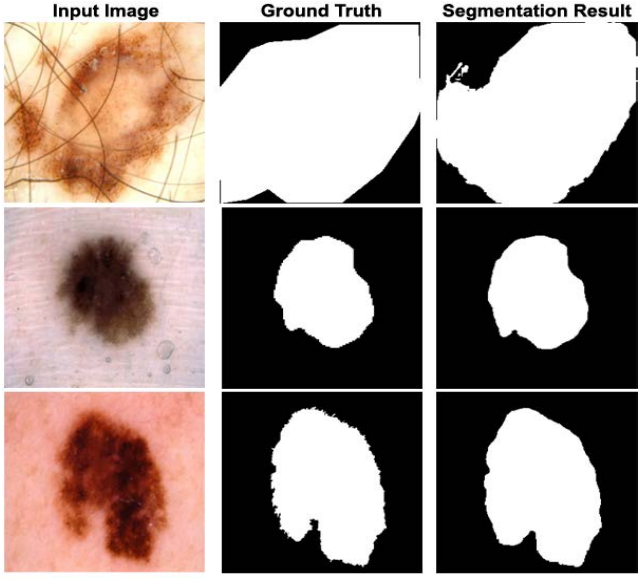

FIGURE 6. Segmentation results of the proposed method on ISIC 2018. The visualization shows that the proposed method learns the complex pattern of the lesion and precisely segments the abnormal regions.

TABLE 3. Performance comparison on the PH^2 dataset.

Methods	DIC	SE	SP	AC	JS
FCN [62]	0.8903	0.9030	0.9402	0.9282	0.8022
SegNet [43]	0.8936	0.8653	0.9661	0.9336	0.8077
FrCN [61]	0.9177	0.9372	0.9565	0.9508	0.8479
MCGU-Net [13]	0.9263	0.8322	0.9714	0.9537	0.9537
Baseline	0.8761	0.8163	0.9776	0.9255	0.7795
Proposed Method	**0.9377**	**0.943**	**0.9698**	**0.9617**	**0.9617**

method significantly improves (DSC 0.937) the performance compared to both baseline (DSC 0.867) and recent MCGU-Net [13] approaches (DSC 0.926). We also provide Figure 7 to represent some segmentation results of the proposed method. As it can be seen from the visual results, our network produced a smooth segmentation output on the boundary area, which is remarkably useful from a clinical perspective.

To further compare the qualitative results of the proposed method to the SOTA approaches, we visualized a sample of the segmentation results in Figure 8, achieved by applying different methods on ISIC 2018 dataset. It is crystal clear that our proposed method pays more attention to the boundary area compared to the U-Net model and outperform this approach. Additionally, compared to the BCDU-Net method, our network produces a smooth segmentation boundary without an extra noisy area.

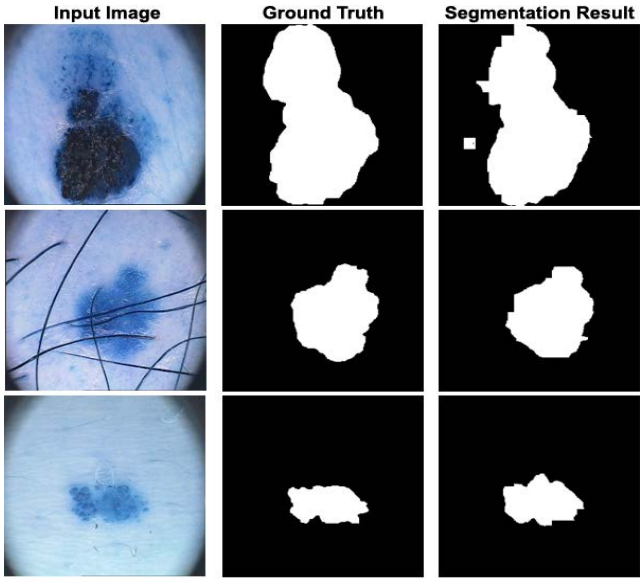

FIGURE 7. Segmentation results of the proposed method on PH^2. The segmentation results illustrate that the proposed method accurately segmented the skin lesion area from the surrounding tissue region.

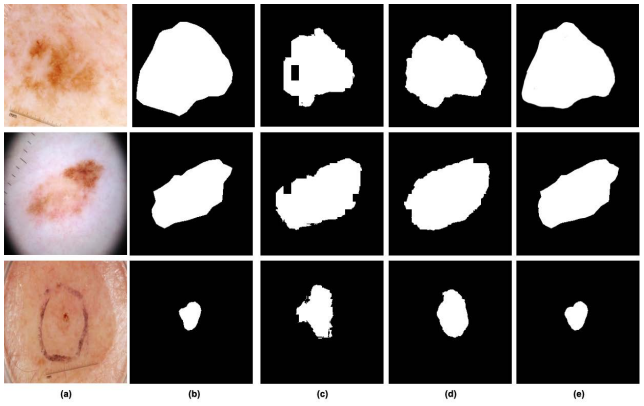

FIGURE 8. Comparing results of proposed method with other state-of-the-art methods on the ISIC 2018 database [59]. (a) shows the original input image, (b) indicates ground truth mask, (c) shows segmentation results of U-Net method, (d) indicates the BCDU-Net method's segmentation results, and finally (e) shows segmentation results of the proposed method.

E. ABLATION STUDY

This section provides an ablation study regarding the effect of the proposed modules. To analyze the contribution of modules individually we experimented with different settings. In our settings, we designed a possible combination of the proposed modules to provide a clear picture of how these modules can effectively be incorporated to increase the model generalization performance on a skin lesion segmentation task. We further included the one-direction version of the ConvLSTM to show the capability of the bidirectional form on encoding a stronger representation and consequently boosting the model performance. Our founding indicates that each module contributes to the model performance and together they provide a strong features representation for the network. Table 4 shows the obtained results.

TABLE 4. Performance comparison on the ISIC 2018 dataset.

Methods	F1	SE	SP	AC	PC
Baseline	0.647	0.708	0.964	0.890	0.779
Baseline+one-directional ConvLSTM	0.751	0.781	0.930	0.911	0.880
Baseline+bi-directional ConvLSTM	0.792	0.852	0.927	0.924	0.912
Baseline+SE block	0.783	0.843	0.916	0.923	0.909
Baseline+bi-directional ConvLSTM+SE block (proposed method)	**0.896**	**0.841**	**0.979**	**0.949**	**0.956**

The conducted experiments (Table 4) show that adding the ConvLSTM module on top of the hierarchical features provided by the seminal U-Net (baseline) model helps the model to learn a rich and generic multi-scale representation and increases the performance considerably. In addition, modifying the direction of the ConvLSTM into a bi-directional further enhances the generalization performance. This fact is in line with the previous research work [14] which included the bi-directional ConvLSTM in the skip connection of the U-Net model and obtained a significant improvement. Besides the ConvLSTM module, we can observe that incorporating the SE block inside the proposed pipeline also increases the model performance. Finally, the combination of these modules with the U-Net model provides a strong feature learning strategy for the medical image segmentation task, which is novel and unique in its design. It is also worthwhile to mention that the processing time for each batch of the eight samples in our pipeline only takes four seconds which demonstrates the suitability of the suggested network for real-time and commercial application.

F. DISCUSSION

The proposed method has been evaluated using both quantitative and qualitative studies to demonstrate its capability in learning rich and generic representation for skin lesion segmentation tasks. The contribution of each proposed components are also evaluated to ensure the effectivness of the suggested design. Although our pipeleine uses U-Net based model, the entire proposed strategy does not have any restrictions on the selection of the segmentation network (e.g., U-Net) and it can be incorporated into any segmentation network, which further supports our contribution in terms of the generalizability and scalability of the network design. Moreover, in Figure 9, we provided sample results of the proposed method where the model fails to segment the skin lesion area. The model performance is largely impacted by the accurate annotation of the images. Therefore, noisy annotation, which is common in the clinical domain, degrades the training performance. Our model cannot detect the inaccurate annotation, and consider all images even if they have inaccurate annotation. As we discussed in the ablation study, the proposed method contains several components which gradually increase the overall performance of the model. One drawback of these modules is their need for computational resources. Hence, although these modules increase the model generalization performance, in the meanwhile they increase

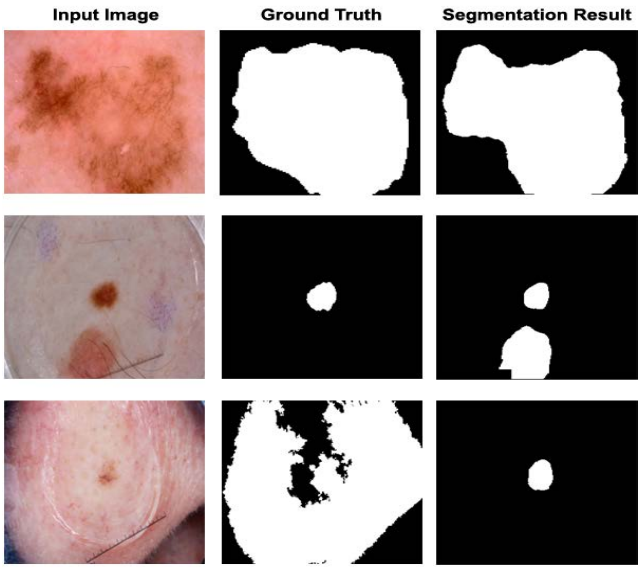

FIGURE 9. Some poor segmentation results of the proposed method on the ISIC 2018 dataset. Visualization shows that the proposed method shows poor performance in cases where the input images have structural complexity and the related annotation mask are inappropriately provided by the specialist (noisy annotation).

the number of parameters and consequently require more computational powers. In this case, there is a trade-off between the performance and the complexity of the model.

V. CONCLUSION

In this paper, we proposed a multi-scale attention mechanism to learn a hierarchical representation. Our attention module receives multi-level feature maps from the encoding model and applies a channel-wise normalization method to recalibrate the feature vectors based on their contribution to the object recognition level, then it utilizes a bi-directional ConvLSTM to learn a hierarchical non-linear representation. By including the resulted feature in each block of the decoding path we incorporate the scale-invariant features inside the network to further boost the performance. The experiment results described throughout the paper proved the effectiveness of our proposal. One possible direction for future work to extend our idea is to model the underlying uncertainty in the skin lesion annotation task. More specifically, with precise modelling the weak annotation of the skin lesion during the training process, the model can further increase its performance.

REFERENCES

[1] G. P. Guy, S. R. Machlin, D. U. Ekwueme, and K. R. Yabroff, "Prevalence and costs of skin cancer treatment in the U.S., 2002–2006 and 2007–2011," *Amer. J. Preventive Med.*, vol. 48, no. 2, pp. 183–187, Feb. 2015.

[2] T. Traver, "Cancer facts & figures," *J. Consum. Health Internet*, vol. 16, no. 3, pp. 366–367, 2012.

[3] J. Feng, N. Isern, S. Burton, and J. Hu, "Studies of secondary melanoma on C57BL/6J mouse liver using 1H NMR metabolomics," *Metabolites*, vol. 3, no. 4, pp. 1011–1035, Oct. 2013.

[4] *Radiation: Ultraviolet (UV) Radiation and Skin Cancer*. Accessed: Sep. 16, 2021. [Online]. Available: https://who.int/news-room/q-a-detail/radiation-ultraviolet-(uv)-radiation-and-skin-cancer

[5] R. L. Siegel, K. D. Miller, and A. Jemal, "Cancer statistics, 2018," *CA, Cancer J. Clin.*, vol. 68, no. 1, pp. 7–30, Jan. 2018.

[6] A. Bozorgpour, R. Azad, E. Showkatian, and A. Sulaiman, "Multi-scale regional attention Deeplab3+: Multiple myeloma plasma cells segmentation in microscopic images," 2021, *arXiv:2105.06238*.

[7] A. R Feyjie, R. Azad, M. Pedersoli, C. Kauffman, I. B. Ayed, and J. Dolz, "Semi-supervised few-shot learning for medical image segmentation," 2020, *arXiv:2003.08462*.

[8] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and Z. Wen, "FAT-Net: Feature adaptive transformers for automated skin lesion segmentation," *Med. Image Anal.*, vol. 76, Feb. 2022, Art. no. 102327.

[9] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2015, pp. 3431–3440.

[10] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2015, pp. 234–241.

[11] S. A. Alryalat, M. Al-Antary, Y. Arafa, B. Azad, C. Boldyreff, T. Ghnaimat, N. Al-Antary, S. Alfegi, M. Elfalah, and M. Abu-Ameerh, "Deep learning prediction of response to anti-VEGF among diabetic macular edema patients: Treatment response analyzer system (TRAS)," *Diagnostics*, vol. 12, no. 2, p. 312, Jan. 2022.

[12] M. Z. Alom, M. Hasan, C. Yakopcic, T. M. Taha, and V. K. Asari, "Recurrent residual convolutional neural network based on U-Net (R2U-Net) for medical image segmentation," 2018, *arXiv:1802.06955*.

[13] M. Asadi-Aghbolaghi, R. Azad, M. Fathy, and S. Escalera, "Multi-level context gating of embedded collective knowledge for medical image segmentation," 2020, *arXiv:2003.05056*.

[14] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Bi-directional ConvLSTM U-Net with Densley connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis. Workshop (ICCVW)*, Oct. 2019, pp. 1–10.

[15] D.-P. Fan, G.-P. Ji, T. Zhou, G. Chen, H. Fu, J. Shen, and L. Shao, "PraNet: Parallel reverse attention network for polyp segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Cham, Switzerland: Springer, 2020, pp. 263–273.

[16] J. Schlemper, O. Oktay, M. Schaap, M. Heinrich, B. Kainz, B. Glocker, and D. Rueckert, "Attention gated networks: Learning to leverage salient regions in medical images," *Med. Image Anal.*, vol. 53, pp. 197–207, Apr. 2019.

[17] R. Azad, A. Bozorgpour, M. Asadi-Aghbolaghi, D. Merhof, and S. Escalera, "Deep frequency re-calibration U-Net for medical image segmentation," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2021, pp. 3274–3283.

[18] R. Azad, A. R. Fayjie, C. Kauffmann, I. B. Ayed, M. Pedersoli, and J. Dolz, "On the texture bias for few-shot CNN segmentation," in *Proc. IEEE Winter Conf. Appl. Comput. Vis. (WACV)*, Jan. 2021, pp. 2674–2683.

[19] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh, and J. Liang, "UNet++: A nested U-Net architecture for medical image segmentation," 2018, *arXiv:1807.10165*.

[20] R. Azad, M. Asadi-Aghbolaghi, M. Fathy, and S. Escalera, "Attention deeplabv3+: Multi-level context attention mechanism for skin lesion segmentation," in *Proc. Eur. Conf. Comput. Vis.* Cham, Switzerland: Springer, 2020, pp. 251–266.

[21] S. Kohl, B. Romera-Paredes, C. Meyer, J. De Fauw, J. R. Ledsam, K. Maier-Hein, S. Eslami, D. J. Rezende, and O. Ronneberger, "A probabilistic U-Net for segmentation of ambiguous images," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 31, 2018, pp. 6965–6975.

[22] L. C. Chen, G. Papandreou, and I. Kokkinos, "DeepLab: Semantic image segmentation with deep convolutional nets, Atrous convolution, and fully connected CRFs," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 40, no. 4, pp. 834–848, Jun. 2016.

[23] L.-C. Chen, G. Papandreou, F. Schroff, and H. Adam, "Rethinking atrous convolution for semantic image segmentation," 2017, *arXiv:1706.05587*.

[24] A. Sinha and J. Dolz, "Multi-scale self-guided attention for medical image segmentation," 2019, *arXiv:1906.02849*.

[25] R. Azad, N. Khosravi, and D. Merhof, "SMU-Net: Style matching U-Net for brain tumor segmentation with missing modalities," 2022, *arXiv:2204.02961*.

[26] O. Oktay, J. Schlemper, L. Le Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y Hammerla, B. Kainz, B. Glocker, and D. Rueckert, "Attention U-Net: Learning where to look for the pancreas," 2018, *arXiv:1804.03999*.

[27] H. Zhao, Y. Zhang, S. Liu, J. Shi, C. C. Loy, D. Lin, and J. Jia, "PSANet: Point-wise spatial attention network for scene parsing," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 267–283.

[28] J. L. Garcia-Arroyo and B. Garcia-Zapirain, "Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding," *Comput. Methods Programs Biomed.*, vol. 168, pp. 11–19, Jan. 2019.

[29] P. M. Pereira, L. M. Tavora, R. Fonseca-Pinto, R. P. Paiva, P. A. A. Assunção, and S. M. M. de Faria, "Image segmentation using gradient-based histogram thresholding for skin lesion delineation," in *Proc. 12th Int. Joint Conf. Biomed. Eng. Syst. Technol.*, 2019, pp. 84–91.

[30] M. E. Yüksel and M. Borlu, "Accurate segmentation of dermoscopic images by image thresholding based on type-2 fuzzy logic," *IEEE Trans. Fuzzy Syst.*, vol. 17, no. 4, pp. 976–982, Aug. 2009.

[31] S. Kockara, M. Mete, V. Yip, B. Lee, and K. Aydin, "A soft kinetic data structure for lesion border detection," *Bioinformatics*, vol. 26, no. 12, pp. i21–i28, Jun. 2010.

[32] A. S. Ashour, A. R. Hawas, Y. Guo, and M. A. Wahba, "A novel optimized neutrosophic k-means using genetic algorithm for skin lesion detection in dermoscopy images," *Signal, Image Video Process.*, vol. 12, no. 7, pp. 1311–1318, 2018.

[33] R. Azad, E. Ahmadzadeh, and B. Azad, "Real-time human face detection in noisy images based on skin color fusion model and eye detection," in *Intelligent Computing, Communication and Devices*. New Delhi, India: Springer, 2015, pp. 435–447.

[34] A. Wong, J. Scharcanski, and P. Fieguth, "Automatic skin lesion segmentation via iterative stochastic region merging," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 6, pp. 929–936, Nov. 2011.

[35] O. Salih and S. Viriri, "Skin lesion segmentation using stochastic region-merging and pixel-based Markov random field," *Symmetry*, vol. 12, no. 8, p. 1224, Jul. 2020.

[36] M. E. Celebi, H. A. Kingravi, H. Iyatomi, Y. A. Aslandogan, W. V. Stoecker, R. H. Moss, J. M. Malters, J. M. Grichnik, A. A. Marghoob, H. S. Rabinovitz, and S. W. Menzies, "Border detection in dermoscopy images using statistical region merging," *Skin Res. Technol.*, vol. 14, no. 3, pp. 347–353, Aug. 2008.

[37] F. Riaz, S. Naeem, R. Nawaz, and M. Coimbra, "Active contours based segmentation and lesion periphery analysis for characterization of skin lesions in dermoscopy images," *IEEE J. Biomed. Health Informat.*, vol. 23, no. 2, pp. 489–500, Mar. 2019.

[38] J. Tang, "A multi-direction GVF snake for the segmentation of skin cancer images," *Pattern Recognit.*, vol. 42, no. 6, pp. 1172–1179, Jun. 2009.

[39] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. Marçal, T. Mendonça, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 35–45, Feb. 2009.

[40] A.-R. Ali, M. S. Couceiro, and A. E. Hassenian, "Melanoma detection using fuzzy C-means clustering coupled with mathematical morphology," in *Proc. 14th Int. Conf. Hybrid Intell. Syst.*, Dec. 2014, pp. 73–78.

[41] J. Burdick, O. Marques, J. Weinthal, and B. Furht, "Rethinking skin lesion segmentation in a convolutional classifier," *J. Digit. Imag.*, vol. 31, no. 4, pp. 435–440, Aug. 2018.

[42] *A Brief History of Deep Learning*. Accessed: Feb. 4, 2022. [Online]. Available: https://dataversity.net/brief-history-deep-learning/

[43] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.

[44] R. Azad, L. Rouhier, and J. Cohen-Adad, "Stacked hourglass network with a multi-level attention mechanism: Where to look for intervertebral disc labeling," in *Proc. Int. Workshop Mach. Learn. Med. Imag.* Cham, Switzerland: Springer, 2021, pp. 406–415.

[45] Y. Xie, J. Zhang, Y. Xia, and C. Shen, "A mutual bootstrapping model for automated skin lesion segmentation and classification," *IEEE Trans. Med. Imag.*, vol. 39, no. 7, pp. 2482–2493, Dec. 2020.

[46] K.-K. Maninis, S. Caelles, J. Pont-Tuset, and L. Van Gool, "Deep extreme cut: From extreme points to object segmentation," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 616–625.

[47] A. Kumar, G. Hamarneh, and M. S. Drew, "Illumination-based transformations improve skin lesion segmentation in dermoscopic images," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit. Workshops (CVPRW)*, Jun. 2020, pp. 728–729.

[48] W.-S. Lai, J.-B. Huang, N. Ahuja, and M.-H. Yang, "Deep Laplacian pyramid networks for fast and accurate super-resolution," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 624–632.

[49] L. Liu, Y. Y. Tsui, and M. Mandal, "Skin lesion segmentation using deep learning with auxiliary task," *J. Imag.*, vol. 7, no. 4, p. 67, Apr. 2021.

[50] D. Dai, C. Dong, S. Xu, Q. Yan, Z. Li, C. Zhang, and N. Luo, "Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation," *Med. Image Anal.*, vol. 75, Jan. 2022, Art. no. 102293.

[51] K. D. Foote. (Feb. 2022). *A Brief History of Deep Learning*. [Online]. Available: https://www.dataversity.net/brief-history-deep-learning/#

[52] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, "Attention is all you need," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 30, 2017, pp. 6000–6010.

[53] S. Hao, Y. Cui, and J. Wang, "Segmentation scale effect analysis in the object-oriented method of high-spatial-resolution image classification," *Sensors*, vol. 21, no. 23, p. 7935, Nov. 2021.

[54] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE/CVF Conf. Comput. Vis. Pattern Recognit.*, Jun. 2018, pp. 7132–7141.

[55] X. Shi, Z. Chen, H. Wang, D.-Y. Yeung, W.-K. Wong, and W.-C. Woo, "Convolutional LSTM network: A machine learning approach for precipitation nowcasting," in *Proc. Adv. Neural Inf. Process. Syst.*, vol. 28, 2015, pp. 802–810.

[56] H. Song, W. Wang, S. Zhao, J. Shen, and K.-M. Lam, "Pyramid dilated deeper ConvLSTM for video salient object detection," in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, 2018, pp. 715–731.

[57] Z. Cui, R. Ke, Z. Pu, and Y. Wang, "Deep bidirectional and unidirectional LSTM recurrent neural network for network-wide traffic speed prediction," 2018, *arXiv:1801.02143*.

[58] 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, and 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 *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.

[59] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," 2019, *arXiv:1902.03368*.

[60] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH^2—A dermoscopic image database for research and benchmarking," in *Proc. 35th Annu. Int. Conf. IEEE Eng. Med. Biol. Soc. (EMBC)*, Jul. 2013, pp. 5437–5440.

[61] M. A. Al-masni, M. A. Al-antari, M.-T. Choi, S.-M. Han, and T.-S. Kim, "Skin lesion segmentation in dermoscopy images via deep full resolution convolutional networks," *Comput. Methods Programs Biomed.*, vol. 162, pp. 221–231, Aug. 2018.

[62] H. Noh, S. Hong, and B. Han, "Learning deconvolution network for semantic segmentation," in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, Dec. 2015, pp. 1520–1528.

MOHAMMAD D. ALAHMADI (Member, IEEE) received the M.Sc. and Ph.D. degrees from Florida State University, in 2018 and 2020, respectively. He is currently an Assistant Professor with the Software Engineering Department, University of Jeddah. He has published in top journals and conferences on a wide variety of topics in his research areas. His research interests include software engineering, computer vision, and machine learning.