

CU-Net: A New Improved Multi-Input Color U-Net Model for Skin Lesion Semantic Segmentation

LANIA RAMADAN¹ AND **SALEH ALY^{2,3}**, (Senior Member, IEEE)

¹Mathematics Department, Faculty of Science, Computer Science Section, Sohag University, Sohag 82524, Egypt

²Department of Information Technology, College of Computer and Information Sciences, Majmaah University, Al-Majmaah 11952, Saudi Arabia

³Department of Electrical Engineering, Faculty of Engineering, Aswan University, Aswan 81542, Egypt

Corresponding author: Saleh Aly (s.haridy@mu.edu.sa)

The work of Saleh Aly was supported by the Deanship of Scientific Research at Majmaah University under Project No. R-2022-27.

ABSTRACT Melanoma is considered one of the most dangerous skin cancer diseases that threaten human health and life. Early diagnosis of melanoma is a big challenge, especially with the presence of color variations across similar lesion types. Automatic skin lesion segmentation is an essential step to build a successful skin disease classification system. Recent deep learning architectures significantly improve the skin lesion segmentation results. Especially, U-Net deep convolutional neural network (CNN) is considered one of the state-of-the-art models with promising performance. Most deep CNNs and particularly U-Net model utilize a single input RGB color image for skin lesion semantic segmentation. However, RGB color space is not usually the best choice to represent the invariant characteristics of skin lesion chromatic information. The selection of the optimal color space significantly affects the performance of segmentation results. In this paper, three novel variants of U-Net model with single, dual, and triple inputs, namely, Single Input Color U-Net (SICU-Net), Dual Input Color U-Net (DICU-Net) and Triple Input Color U-Net (TICU-Net) are proposed. The structure of SICU-Net, DICU-Net, and TICU-Net contains single, dual, and triple encoder sub-networks connected with only a single decoder path. Each encoder sub-network is fed with different color space of the input image. A channel-wise attention module is utilized to fuse the contribution of the learned feature maps from each encoder sub-network which is fed to the decoder sub-network to generate segmented image map. Moreover, a composite loss function is designed to improve the performance of the proposed CU-Net models. Three public benchmark datasets, namely, International Skin Imaging Collaboration (ISIC 2017, ISIC 2018) and PH2 datasets, are utilized to evaluate the performance of the proposed models. Experimental results reveal that the proposed models significantly improve the performance of the original U-Net model and achieve comparable performance with other state-of-the-art methods.

INDEX TERMS Color U-Net, multi-input U-Net, combined loss function, melanoma, channel-wise attention, skin lesion semantic segmentation.

I. INTRODUCTION

Skin cancer occurs when some skin cells are grown abnormally. It is commonly appeared in areas of the body that are most exposed to ultraviolet (UV) radiation [1]. Skin cancer is divided into three major types, namely, squamous cell carcinoma, basal cell carcinoma, and melanoma [2]. Melanoma is the most dangerous type of skin cancer since it appears and grows in the melanocyte cells that produce melanin [3]–[5]. The main cause of melanoma is not known yet, but it

The associate editor coordinating the review of this manuscript and approving it for publication was Fahmi Khalifa¹.

is scientifically proven that direct exposure to ultraviolet rays from sunlight or tanning lamps and beds increases the risk of melanoma. Early detection and diagnosis are necessary to preserve human survival [6]. Skin cancer diagnosis, like many diseases, is vulnerable to human error or it may be costly. Thus, recent researches [7]–[10] tend to rely on Computer Aided Diagnosis (CAD) systems based on dermoscopic images to reduce the probability of errors. Skin lesion image analysis is an essential stage for early skin cancer disease diagnosis. As a crucial step, skin lesion image segmentation helps to distinguish lesions from the background by labeling each pixel in the image as healthy or Infected ones. Skin

Lesion segmentation is a challenging problem due to the low contrast of skin lesions images, irregular borders, and the existence of some extraneous elements such as hair, pen markers, oil drops, etc. Other factors such as the irregular color, size, scale, and precision of the lesion area add another challenge to the segmentation problem. Figure 1 shows some examples of skin lesion dermoscopy images with various variations. All those factors pose major obstacles to the success of many skin lesion segmentation methods. Increasing the number of collected skin lesion images makes it imperative to use computers to perform automatic segmentation easily and effectively. However, manually labeling such a large number of images by an expert is a tedious work and very costly.

A few research works have carried out to study the color contrast variations in skin lesion image segmentation [11]–[16]. Most of these works utilize color space conversion-based methods to efficiently represent the infected parts and discriminate them from background pixels. These methods convert the RGB input image to another color space like YCbCr, Lab, and HSV. A significant number of research applied a specialized pre-processing operation on the input image [11], [14], [17]–[19] while other works employed post-processing [11], [12], [17], [20], [21] to enhance the segmentation results of the proposed architectures. On the other hand, recent works using deep learning methods [22]–[25], [27]–[29] rely on extending the architecture of existing deep convolutional neural networks to improve the skin lesion image segmentation results.

One of the main issues in skin lesion segmentation is the color contrast variations among similar lesion types. Although previous works handled this issue by exploiting various color spaces using single input deep CNN, this paper investigates the effect of combining multiple color spaces of the input image using multi-input deep CNN. A new modified Color U-Net (CU-Net) model with multiple encoders and one decoder sub-network is developed to take advantages of combining multiple features extracted from various color space representations of the input image. The U-Net model [30] is used as a backbone architecture to build a three variants of CU-Net models. The proposed models depends on connecting multiple encoders into a single decoder sub-network to capture different color features of the skin lesions to overcome color variation problem of skin lesion image segmentation. Three different models are proposed based on U-Net architecture using single, dual, and triple inputs, namely, Single Input Color U-Net (SICU-Net), Dual Input Color U-Net (DICU-Net) and Triple Input Color U-Net (TICU-Net). The proposed DICU-Net and TICU-Net models comprise two and three separate encoder sub-networks based on the traditional U-Net structure, where each encoder path of the proposed network accepts different color space of the input image and interconnected with each other to learn more rich features. The learned feature maps from each encoder sub-network are fused through a Channel-wise Attention Network (CAN) and fed into the decoder sub-network to learn the contribution of

each color space into the segmentation results. Various combinations of color spaces are examined to find the optimum color spaces which achieve the best performance. Due to the significant effect of loss function in the segmentation results, a new hybrid binary-weighted loss function is proposed from the combination of three different loss functions, namely, cross-entropy, generalized dice, and sensitivity-specificity. We considered various performance metrics such as: accuracy, sensitivity, specificity, Jaccard index and Dice coefficient to evaluate the segmentation results. The contributions of the paper can be summarized as follows:

- Exploring the effect of changing input image color space on the performance of the new proposed color U-Net model segmentation results.
- Proposing two multi-input networks based on color U-Net architecture. These networks are composed of multiple encoder sub-networks and a single decoder to combine various color features of the input image.
- Using channel-wise attention module to interconnect the encoder and decoder paths of the proposed models.
- A new hybrid loss function is designed as a combination of cross-entropy, generalized dice, and sensitivity-specificity loss to improve the performance of the proposed models.
- Experiments are conducted using three standard skin lesion databases to validate the proposed models and a comparison of the results with other state-of-the-art methods is performed.

The rest of the paper is organized as follows: the related works are reviewed in section II. Our proposed models are detailed in section III. Then, we come to the experimental results section IV followed by the analysis of the obtained results. Discussion is presented in section V. Finally, the conclusion is given in section VI.

II. RELATED WORK

Several methods have been introduced to solve the automatic skin lesions segmentation problem over the last decade. In this section, we briefly explain the latest developments in skin lesion segmentation which are closely related to our work. Currently, recent skin lesion segmentation methods are transferred from adopting shallow hand-crafted feature extraction algorithms into deep feature learning using Convolutional Neural Network (CNN) architecture. These methods can be classified into three different categories: Traditional, deep learning, and color-based skin lesion segmentation methods.

A. TRADITIONAL SKIN LESION SEGMENTATION METHODS

An efficient skin lesion image segmentation technique helps to enhance the early detection and diagnosis of melanoma. Several approaches are adopted according to the nature of the segmented lesion. Region growing approaches [31], [32] and thresholding approaches [33]–[35] are used to find the fine border in lesion images. While clustering approaches [36], [37] are used to address the unclear border

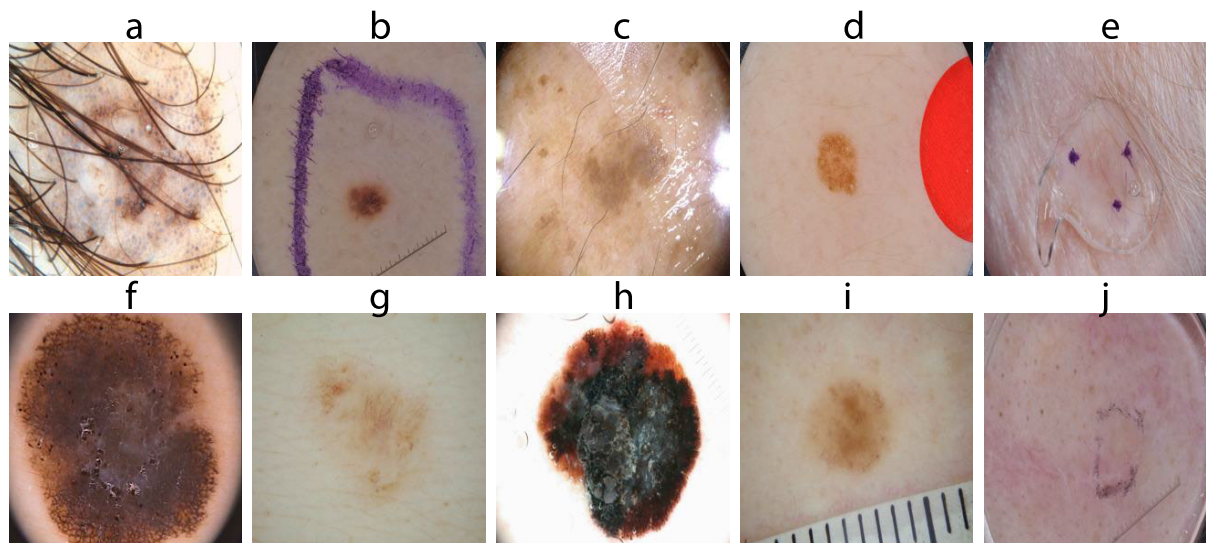


FIGURE 1. Image samples of appearance variations in ISIC 2017 skin lesion dataset due to: (a) hair, (b) pen markers, (c) oil, (d) markers, (e) pen marker and gel, (f) irregular boundaries, (g) light-colored lesion area, (h) dark-colored lesion area, (i) ruler sign, (j) unclear lesion.

lesions problem. Other approaches in [38], [39] are based on employing histogram-based cluster estimation to distinguish between healthy and infected parts in the images. Active contours based methods [40], [41] were among the proposed approaches that focused in the separation of the required pixels from the image for further processing and analysis by using energy forces with some constraints. In addition, edge-based methods [42] were suggested to identify the region of interest. All previously mentioned segmentation techniques are considered primitives as they depend only on low-level pixel-wise features. Therefore, they could not achieve the desired performance in comparison with deep learning methods.

B. DEEP LEARNING SKIN LESION SEGMENTATION METHODS

Recently, Deep Convolutional Neural Networks (DCNN) is a powerful technique that play a prominent role in developing new medical image segmentation methods [10], [17]–[19], [21], [43]–[48] due to the high-accuracy results in segmentation operations. In 2011, Yuan *et al.* [17] utilized a 19-layer deep convolutional neural network that is trained end-to-end. They did not have any prior knowledge of the data to present a fully automatic method for skin lesion segmentation. Furthermore, they proposed a novel loss function based on Jaccard distance to dispense with sample re-weighting. Their technique could eliminate the need for data re-balancing when the numbers of foreground and background pixels are un-balance.

In 2018, Poap *et al.* [47] presented a smart home system that using in-built sensors and proposed artificial intelligence methods to diagnose the skin health condition of the residents of the house. They compared the results of their proposed method with the results of some similar methods. In 2019,

Hashemi *et al.* [46] used an asymmetric similarity loss function to train a fully convolutional deep neural network in order to overcome the data imbalance issue and achieved a much better trade-off between precision and recall. Moreover, they developed a 3D fully convolutional densely connected network (FC-DenseNet) with large overlapping image patches as input and an asymmetric similarity loss layer based on Tversky index. They also used large overlapping image patches as inputs for intrinsic and extrinsic data augmentation, a patch selection algorithm, and a patch prediction fusion strategy using B-spline weighted soft voting to account for the uncertainty of prediction in patch borders.

An automatic semantic segmentation network for skin lesion segmentation named Dermoscopic Skin Network (DSNet) and a new loss function that combines a binary cross-entropy and intersection over-union are presented by Hasan *et al.* [10] in 2020. They succeeded to reduce the number of parameters and make the network lightweight. they used a depth-wise separable convolution instead of standard convolution to stand out discriminatory features in the pixel space at different steps of the encoder. Their proposed loss function played an important role in semantic segmentation and achieved higher true positive rates in the conducted experiments. While, in 2020, a novel segmentation methodology is proposed by Al-masni *et al.* [18]. They used Full resolution Convolutional Networks (FrCN) that do not require any pre- or post-processing operations but it learns the full resolution features of each individual pixel of the input data to improve pixel-wise segmentation performance. They achieved high evaluation metrics value in the tested datasets. Xie *et al.* [8] devised a high-resolution feature block with three branches, the first one is the main branch that accepts high-resolution feature maps to extract spatial details around boundaries. The second and the third branches are the spatial

attention and the channel-wise attention branches, which are used to enhance the discriminative features in the main branch regarding the spatial and channel-wise dimensions. Robust features with detailed spatial information were yielded and extracted, and accurate skin lesion boundaries have obtained by fusing the branch outputs.

In 2021, Khan *et al.* [44] proposed a fully automated computerized aided diagnosis system that is based on the deep learning framework. In their proposed scheme, they pre-processed the original dermoscopic images using the decorrelation formulation technique. Then, the resultant images are passed to the MASK-RCNN which is trained using the segmented RGB images generated from the ground truth images of the used datasets. Next, the resultant segmented images are passed to the DenseNet deep model for feature extraction. They combined the output of average pool and fully connected layers for feature extraction and the resultant vector is forwarded to the feature selection block for down-sampling using proposed entropy-controlled least square SVM. Their proposed model was the lowest in computational time compared with other works. Kadry *et al.* [45] extracted the skin melanoma section by employing the VGG-SegNet scheme to the digital dermoscopy image. Then, they executed a relative assessment between the segmented skin melanoma and the ground-truth, and computed the essential performance indices. Their proposed technique is significant in evaluating the clinical grader of digital dermoscopy image. Khan *et al.* [48] extended their previous works and proposed a fully automated approach for multi-class skin lesion segmentation and classification that used the most discriminant deep features. They used local color-controlled histogram to enhance the image intensity values. Then, they used a novel Deep Saliency Segmentation method that is a custom convolutional neural network (CNN) of ten layers to estimate Saliency. By using a thresholding function, they got a binary image from the generated heat map. They used the segmented color lesion images for feature extraction via a deep pre-trained CNN model. They implemented an Improved Moth Flame Optimization (IMFO) algorithm to select the most discriminant features to avoid the curse of dimensionality, and the resultant features are fused using a Multi set Maximum Correlation Analysis (MMCA) and classified using the Kernel Extreme Learning Machine (KELM) classifier. Their approach showed an improvement in accuracy but with high computational time.

Currently, several researches in skin lesion segmentation field employed a model of Fully Convolutional Network called U-Net [9], [49]–[54]. In 2019, a psoriasis lesion segmentation network (PsLSNet) was introduced by Dash *et al.* [52] as an automated method based on a modified U-Net architecture. The architecture has 29-layer deep fully convolutional network for extracting spatial information automatically. Their proposed structure accelerated the training by reducing the co-variate shift through the implementation of batch normalization and could segment the lesion even in challenging cases. Qamar *et al.* [49] combined

DenseNet [55] and ResNet [56] and presented an encoder-decoder-based CNN for skin lesion segmentation that is based on UNet architecture. Their aim was to improve the performance of skin lesion segmentation. In the encoder path, they used atrous spatial pyramid pooling (ASPP) to generate multi-scale features from different dilation rates and used dense skip connections to combine the encoder and decoder feature maps. They proposed a deep learning method to model lesion patterns to perform melanoma detection and lesion segmentation. They were able to exploit multi-scale contextual information, retrieve accurate information, and improve segmentation performance. Pham *et al.* [50] combined multiple hypotheses into a single decision point. For melanoma detection and seborrheic-keratosis classification, they trained Inception-v4 [57], ResNet-152, and DenseNet-161 [55]. While for lesion segmentation, U-Net and U-Net with VGG-16 Encoder [58] were trained to produce segmentation masks. Their model ranked 5th in classification and 8th in segmentation among 23 and 21 international teams, respectively. Also, Tang *et al.* [51] exploited the advantages of U-Net architecture and the separable convolutional block to propose a skin lesion segmentation method that is based on the separable-U-Net with stochastic weight averaging to get higher semantic feature information. They introduced a scheme based on stochastic weight averaging to obtain an optimum broader and better generalization. They enhanced the pixel-level discriminative representation capability.

Azad *et al.* [54] proposed a frequency re-calibration U-Net (FRCU-Net) for medical image segmentation to reduce the effect of texture bias, and get better generalization for a low data regime. They applied the Laplacian pyramid in the bottleneck stage in the U-structure. They used a channel-wise attention mechanism to capture the relationship between the channel features maps in a layer of the frequency pyramid. Then the extracted features of each level of the pyramid are combined through a non-linear function based on their impact on the final segmentation output. Their proposed net achieved state-of-the-art results, However, employing high frequency information can exaggerate the noise information exists in the skin images which may deteriorate the performance of the system in real-time scenarios. Alom *et al.* [59] utilized the power of U-Net to propose a Recurrent Convolutional Neural Network (RCNN) based on U-Net as well as a Recurrent Residual Convolutional Neural Network (RRCNN) based on U-Net models, and named RU-Net and R2U-Net respectively. All the proposed models had several advantages for segmentation tasks. Their proposed network showed better performance in segmentation tasks with the same number of network parameters when compared to existing methods including the U-Net and residual U-Net (or ResU-Net) models. However, computational Time for training and testing increased due to the recurrence operations. Asadi *et al.* [60] took the full advantages of U-Net, Squeeze and Excitation (SE) block, bi-directional ConvLSTM (BConvLSTM), and the mechanism of dense convolutions to create a new model called (MCGU-Net) as an extension of U-Net for

medical image segmentation. The network was able to capture more discriminative information and more precise segmentation results. However, using dense connections in the bottleneck increase the complexity of the network.

C. COLOR-BASED SKIN LESION SEGMENTATION METHODS

Few researchers [11], [12], [14], [61] have adopted the idea of using color spaces in the segmentation process that relying on convolutional neural networks. Schaefer *et al.* [19] presented an effective approach that combines an enhancement stage and two different segmentation algorithms. The enhancement stage was a pre-processing step to counter the weak contrast and lack of color calibration in the dermoscopy images. They proved that applying a color normalization technique was necessary to reduce color variations and enhance image contrast to improve skin lesions segmentation. Pour *et al.* [11] developed a segmentation model for skin lesion segmentation tasks and dermoscopic feature segmentation. They trained the network from scratch in spite of the limitation of data size without applying any data augmentation or any pre-processing techniques to remove artifacts or enhance images. Instead of that, they increased the depth of input to convolutional layers by using the efficient feature maps concatenation from transform domain in addition to using CIELAB colour space with RGB colour channels. Their proposed method improved the segmentation process in two ways: (1) concatenated feature maps to the network provided an excellent realization of the input to the model. (2) applying CIELAB colour space with RGB colour channels provides more information for the network. Likewise, a deep fully Convolutional Deconvolutional Neural Network (CDNN)- based framework is proposed by Yuan *et al.* [61] to automatically segment dermoscopic images of skin lesions. They did not use pre-or post-processing algorithms and even any hand-crafted features. They focused on designing a suitable network architecture and an effective training method such that it can handle images under various acquisition conditions. Besides using RGB color for the input image, they used HSV color space with L channel from Lab color space as the training input image. Their effective training strategies were able to handle images under different acquisition conditions.

De Angelo *et al.* [12] presented a methodology using a combination of deep learning, color space, and conditional random fields to segment a dataset created in partnership with a group of dermatologists. For this purpose, they developed an application to collect skin lesion images using smartphones' camera and created a new clinical dataset. They presented an investigation regarding the color spaces and post-processing that had enabled them to raise some important remarks about the ground truth images for skin lesions that affect the final segmentation results. Abbas *et al.* [14] enhanced the quality of dermoscopic images before segmentation by applying three pre-processing operations. Those operations include, firstly remove image noise by using median filter and morphological operations. Secondly, they

selected the green channel as the optimal color channel from RGB values. Thirdly, they utilized a combined Spline and B-spline to enhance the image before segmentation. After image pre-processing, they used the empirical threshold value of the optimal color channel to complete the lesion segmentation. Finally, post-processing (morphological operation) was utilized to fuse the smaller regions and the main lesion region and extract the lesion border. Although their approach achieved a good performance and obtained a high accuracy value, it depends on the quality of the pre- and post-processing results. Table 1 shows a summary of the most related methods listed in this section that concerned the skin lesion segmentation in terms of the advantages and the disadvantages the proposed methods with the detests employed in the testing.

III. PROPOSED METHOD

In this section, we explain the details of the three proposed color U-Net models which is based on the well-known U-Net architecture. Firstly, we explain the image pre-processing technique employed in this work. Then, we explain the structure of the proposed single input color U-Net (SICU-Net). Next, the other two proposed architectures, namely DICU-Net and TICU-Net are explained in details. Finally, the proposed composite loss function is presented.

A. IMAGES PRE-PROCESSING

The performance of the proposed skin lesion segmentation method can be improved by applying a simple pre-processing technique on the images before performing the training/testing process. Since the skin lesion images used for segmentation contain a lot of noise, complexities, and different textures, it was necessary to use a pre-processing to relieve the effects of these obstacles. We used two-steps pre-processing process including, color space transformation and image contrast normalization.

1) COLOR SPACE TRANSFORMATION

The captured skin lesion images are usually available as 3-channel images or tri-stimulus colors, red (R), green (G), and blue (B) called RGB color space. RGB color space is the most commonly used color space in the image processing field. Although RGB model is commonly used, it is a device-dependent color space and not uniform. Also, it is difficult to determine a specific color in that model and there are non-linear differences among similar colors. Figure 4 shows existing irregular color contrast representation in the ISIC dataset. The problem of color contrast lesion variations can be alleviated by selecting appropriate color space representation of the input image. Various color spaces are commonly used to solve skin lesion segmentation problems such as: Lab, HSV, YCbCr, XYZ, etc.

While RGB represents colors as a combination of red, green, and blue signals, the YCbCr color space is defined as a transformation from RGB color space. YCbCr represents

TABLE 1. Summary of the advantages and limitations of the most related methods concerned with skin lesion segmentation.

Reference	Advantages	Tested Dataset	Limitations
Arora et al. [9]	Using Group Normalization (GN) led to extract the feature maps efficiently, and using Attention Gates (AG) distinguish high dimensional information from low-level irrelevant background regions	ISIC 2018	This method relied on a specific pre-processing and post-processing techniques to improve the results.
Al-masni et al. [18]	Using the full spatial resolutions of the image enabled the proposed FrCNN to learn better features and improving the pixel-wise segmentation performance.	ISBI 2017 & PH2	Expensive computational cost due to the elimination of subsampling layers.
Unver et al. [21]	Combined a deep convolutional neural network named as You Only Look Once (YOLO) and the GrabCut algorithm allow segmentation images with higher resolution and dimension independent.	PH2 & ISBI 2017	The method has a less perfect segmentation than the deep learning-based methods. Also, it does not detect the lesion at a low contrast situation, and when the lesion occupies large area.
Yuan et al. [60]	Developed a deeper network architecture with smaller kernels to increase the discriminant capability of the network and combining information from multiple color spaces to improve segmentation performance.	ISBI 2017	Proposed method is evaluated using single dataset only.
Lin et al. [52]	Using U-Nets achieved a significantly higher Jaccard Index, and using the histogram equalization improved the U-Net segmentation results	ISIC 2017	The algorithm performs poorly when artifacts are present, and incorrectly identifies the black border, the medical gauze, and other dark objects as a lesion.
Reza et al. [61]	Proposed an extension of U-Net called Bi-directional ConvLSTM U-Net with Densely connected convolutions (BCDU-Net). They accelerated the convergence speed of the proposed network by employing batch normalization (BN).	Retinal blood vessel (DRIVE Dataset), ISIC 2018 Dataset, Lung Segmentation Dataset	The densely connected convolutional blocks highly increase the computational cost of the model.
Garcia et al. [62]	They developed a method based on fuzzy classification of pixels and histogram thresholding is efficient, and obtains very good results in terms of reliability, fast, and has low computational cost.	ISBI 2016 & ISBI 2017	The method can not handle severe imaging condition variations.

colors as a combination of luma component of the color signal and two chroma signals (blue-difference and red difference). HSV is a map of the RGB primary colors into easier and understandable dimensions for humans which are: hue, saturation, and value. HSV is a cylindrical color which is more natural than the additive (RGB) color components. XYZ color space is the conversion of the channels of the RGB image to CIE 1931 XYZ values. Y channel refers to luminance, Z channel is approximately equal to blue channel, and X is a mix of response curves chosen to be non negative and orthogonal to luminance. Lab is a color model that is introduced to obtain near uniform spacing to the differences of perceived color. The distance between the two points shows the difference of the colors in luminance, chroma, and hue. YIQ color space is the model that is used in NTSC color TV system. The Y channel represents the luma information or brightness of the image. While, I is the amount of blue or orange tones in the image. Q (Quadrature) is the amount of green or purple tones in the image. The gray-scale image is the image with pixels values equal to the amount of light for each pixel.

Each color space of the skin lesion image provides a different representation which help to capture the invariant color properties of the lesion. Figure 2 (left part) shows the appearance of the components of each color space. It is clear from the figure that some channels carry discriminative skin

lesion information with high contrast between lesion and normal skin such as: B, Y, V, X, Y and L channels in RGB, YCbCr, HSV, XYZ YIQ, and Lab color spaces, respectively. In addition, some channels like Cr, IQ, and ab have invariant characteristics of the lesion despite their low contrast appearance.

2) IMAGE CONTRAST NORMALIZATION AND ENHANCEMENT

In order to increase the contrast of skin images after converting them into another color space, we applied a simple image contrast enhancement technique to improve the contrast between background and foreground (skin lesion) in each channel. The contrast enhancement method maps the intensity values of each component of the input image to new values. The contrast enhancement technique also, control the relationship between the values in the input and the output images by defining the shape of the curve that describes this relationship to produce more brighter, darker, or linear mapping output values. The intensity of the normalized image have 1% values saturated at low and high intensities. In this paper, we chose the linear mapping method and investigated its performance on the proposed models. Figure 2 (right part) shows the appearance of each color space component after applying contrast normalization. It is clear

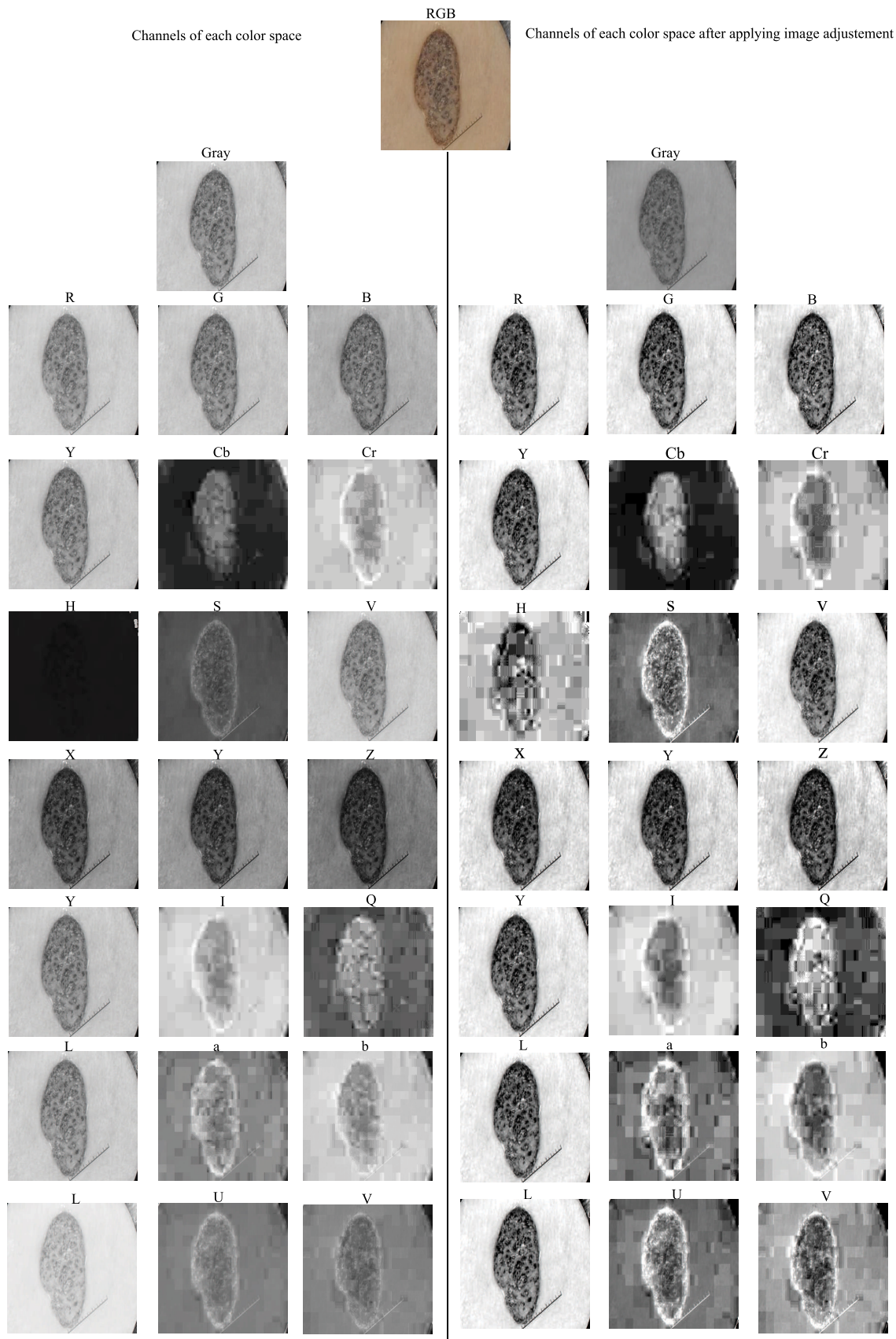


FIGURE 2. Different component representation of each color space. On the left without using image adjustment and on the right after using image contrast normalization and enhancement.

that, the distinction between foreground and background pixels is improved by applying the simple contrast enhancement technique.

B. SINGLE INPUT COLOR U-NET ARCHITECTURE

The use of convolutional neural networks (CNN) in the image classification field has been prevalent for a long time. CNN is a hierarchical model used to learn multi level features of the image and these feature maps are transformed into a vector which is then used for the classification task. However, with the presence of complex objects and the need to perform segmentation tasks which requires not only converting feature maps into a vector but also re-model the image from this vector, the use of traditional convolutional neural networks has become useless and can not achieve any acceptable progress. Therefore, the need for an alternative structure that can perceive these complex images has emerged and this is the beginning of the U-Net architecture.

The performance of classical U-Net is not satisfactory when applied to skin lesion segmentation due to the low contrast and color discrepancy of skin lesions. To overcome these problems, we propose three network architectures based on U-Net structure, which exploit different color spaces at each input path. The proposed networks enclose single, dual, and triple inputs and one output path. The inputs can be fed with any color space or any selected channel. Proposed networks are denoted as single input color U-Net (SICU-Net), dual input color U-Net (DICU-Net), and triple input color U-Net (TICU-Net).

U-Net was suggested by Ronneberger *et al.* [30] and designed to deal with biomedical image segmentation. It was focused on the image segmentation task where its input is an image and the output is a single label for each pixel in this image. U-Net is able to localize and recognize borders as it performs classification on every pixel, so the input and output images have the same size. The idea of U-Net is that the learned feature maps from the input image can be transformed again into another image instead of a vector of classes. Figure 3 shows the architecture of the proposed single input color U-Net (SICU-Net) model. The design of the SICU-Net looks like a 'U' shape, hence it gets its name. The SICU-Net structure consists of three parts: The contracting/down-sampling (encoder path), bottleneck, and the expanding/up-sampling (decoder) path. The encoder path consists of 4 blocks. Each block contains two 3×3 convolution layers followed by an activation function and a 2×2 Max Pooling layer. The number of feature maps is doubled after each pooling. The network begins with 64 feature maps in the first block, then it is doubled to 128 for the second, and so on. The function of the encoder path is to capture the context information of the input image, then producing feature maps. Those feature maps will be transferred to the up-sampling through a channel-wise attention module instead of the simple skip connections (*dashed gray arrows in Fig. 3*) in the classical U-Net model. The second part is the bottleneck which

is located between the contracting and expanding paths. The bottleneck is also built from two 3×3 CNN layers followed by a 2×2 up-convolution layer with dropout. The third part is the decoder path and again, is composed of 4 blocks each one of them consists of a deconvolution layer with stride 2, concatenated with the corresponding cropped feature map from the contracting path, and two 3×3 convolution layers with ReLU activation function. At the end of each block, the number of feature maps decreases by half to keep symmetry. The function of the decoder path enables precise localization using transposed convolutions. However, every time the input is appended by attentioned feature maps from the corresponding contraction layer. In this way, the features which are learned from contracting the input image will be passed through a Channel-wise attention module and used to reconstruct it. The number of expansion blocks is equal to the number of contraction blocks. In the end, the attentioned feature maps passes through another 1×1 convolutional layer with the number of feature maps equal to the number of classes desired to produce the segmentation map by assigning a label for each pixel in the image.

In the single input color U-Net (SICU-Net) structure, the color space/channels of the input image can be chosen according to the type of medical input image. The segmentation results can be improved if we choose an appropriate color space to represent the class information. Thus, SICU-Net model allows to manually choose the appropriate color space/channel to improve segmentation results. The contraction path of the SICU-Net network serves as a deep feature extractor which is connected to the expansive path through channel-wise attention modules. Expansive path enables precise localization information to be combined with contextual information coming from the contracting path. SICU-Net model inherits all advantages of traditional U-Net and provides a general (context and localization) information necessary to predict segmentation map. Input images can be of any size and color space/channels. The possibility of using data augmentation due to the limited number of annotated samples can further improve the results.

C. PROPOSED DUAL INPUT CU-NET (DICU-NET) AND TRIPLE INPUT CU-NET (TICU-NET)

Similar to all convolution neural networks, SICU-Net architecture accepts only one input image which is processed by consecutive convolution and pooling layers. This paper extends the capabilities of the proposed SICU-Net architecture to accept multiple inputs. The proposed network architectures namely, dual input color U-Net (DICU-Net) and triple input color U-Net (TICU-Net) are adapted from the proposed SICU-Net structure with multiple encoders and single decoder path. The contracting path in each network has four (stages) comprising convolution blocks followed by a bottleneck stage. Each block contains two convolution layers with ReLU activation function followed by a max-pooling layer for down-sampling the input by a factor of two. Each encoder sub-network receives the same input image but with

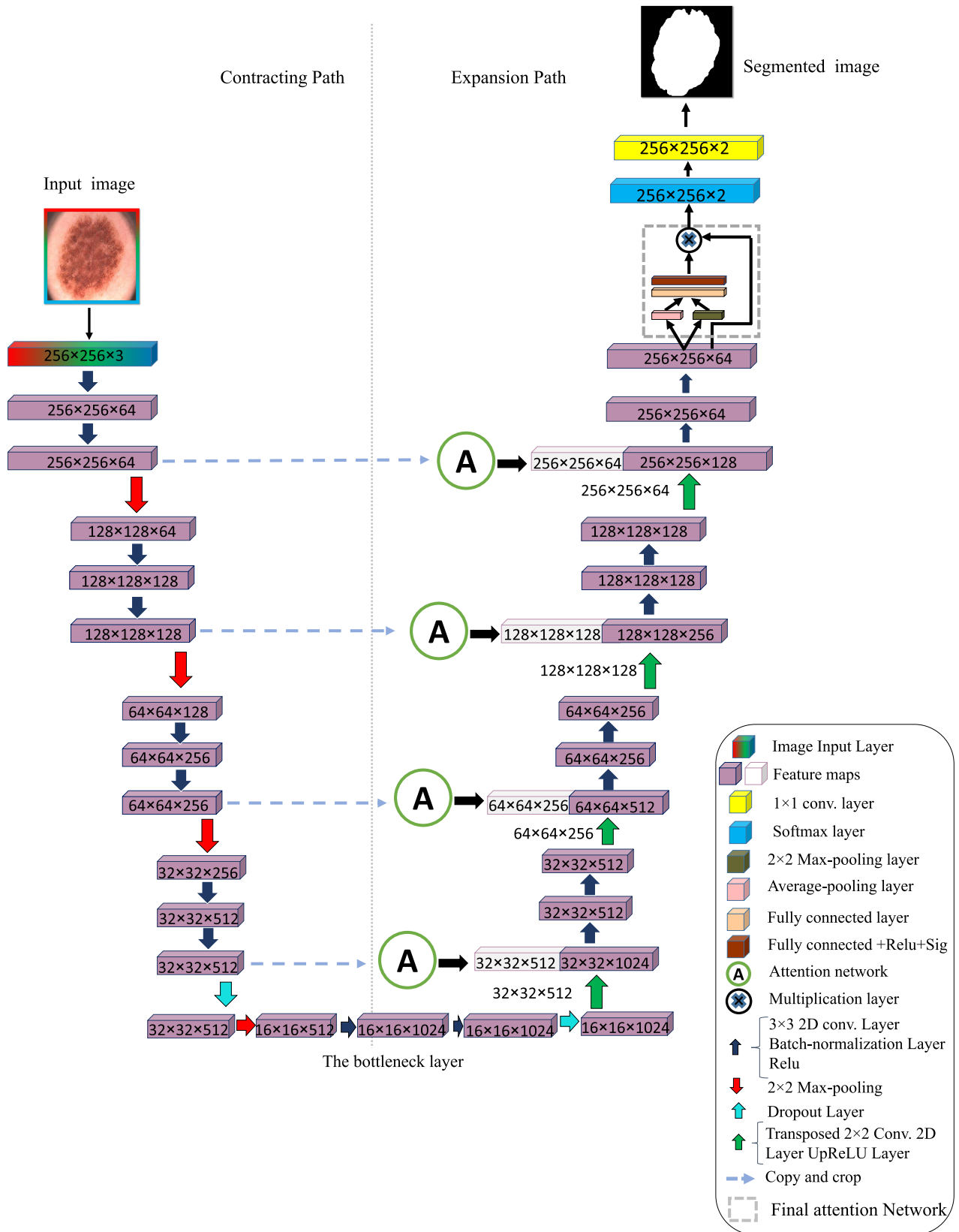


FIGURE 3. Single Input Color U-Net (SICU-Net) architecture.

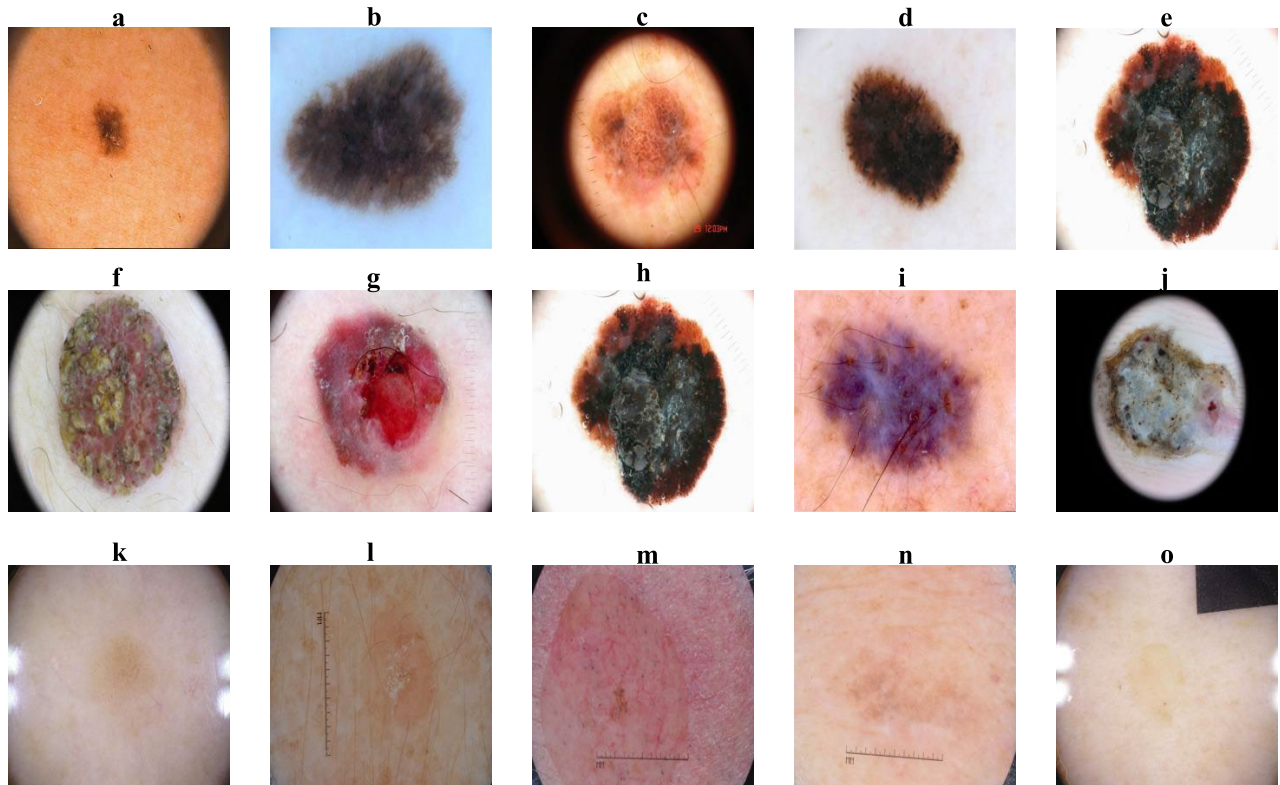


FIGURE 4. Color challenges in ISIC dataset. Top row (a-e) shows the lesion with different background colors. The middle row (f-j) shows the lesion itself with different colors. The bottom row (k-o) shows how the lesion and background are close in color.

different color spaces. The output from each max-pooling layer in each stage of the encoder sub-network is added to the output of the corresponding stage in the other encoder sub-network. Then, the output of each addition layer of each stage from every encoder sub-networks are passed regularly to the next stage. In contrast, the feature maps resulted from the encoding path in each stage from encoder sub-network are fed to a channel attention network in the expansive path. Then, the output is concatenated with output of each sub-network bottleneck stage passed through decoding blocks that begin with a de-convolution layer followed by two convolutional layers to decrease the number of feature map by a factor of two. Figure 5 and Figure 6 show in details the architecture of DICU-Net and TICU-Net, respectively. The extracted features in the decoding path are passed to the channel attention network at the end of the decoding path and fed to the final classification layer. The combination of various feature maps from different input color spaces leads to improve the performance of the segmentation results. Also, the entire context of the input images is kept due to the end-to-end training process from the multiple input images to produce its corresponding segmentation map. The structure of the attention network is shown in detail in section III-D.

The proposed DICU-Net and TICU-Net structures are fed with an RGB input image and followed by a distribution layer which convert the input image into another color space and fed it to each encoder path of multi input CU-Net. The structure of the proposed encoder sub-network including

the number of layers, size and number of convolutional filters are identical to each other. The feature maps from the last de-convolutional layer of the decoding path are fed to an attention network before entering the final 1×1 convolution layer, which produce the semantic segmentation map.

The main characteristics of the proposed SICU-Net, DICU-Net, and TICU-Net models can be summarized as follows:

- 1) At the beginning, the input image is resized and normalized then converted to the desired color spaces.
- 2) All the channels that are resulted from this conversion is stacked together one after another to form an input image with multiple channels from different color spaces.
- 3) The constructed multi-channel input image is sent to the distribution layer. This layer is responsible for dividing the input image channels into single, dual or triple subset(s) of channels according to the type of network.
- 4) In the single input color U-Net model (SICU-Net), all input channels are fed into the CU-Net.
- 5) In the dual input color U-Net model (DICU-Net), the first subset of channels is fed into the first encoder sub-network while the second subset is fed into the second encoder sub-network.
- 6) In the triple input color U-Net model (TICU-Net), the first subset of channels is fed into the first encoder sub-network, while the second and third subsets are

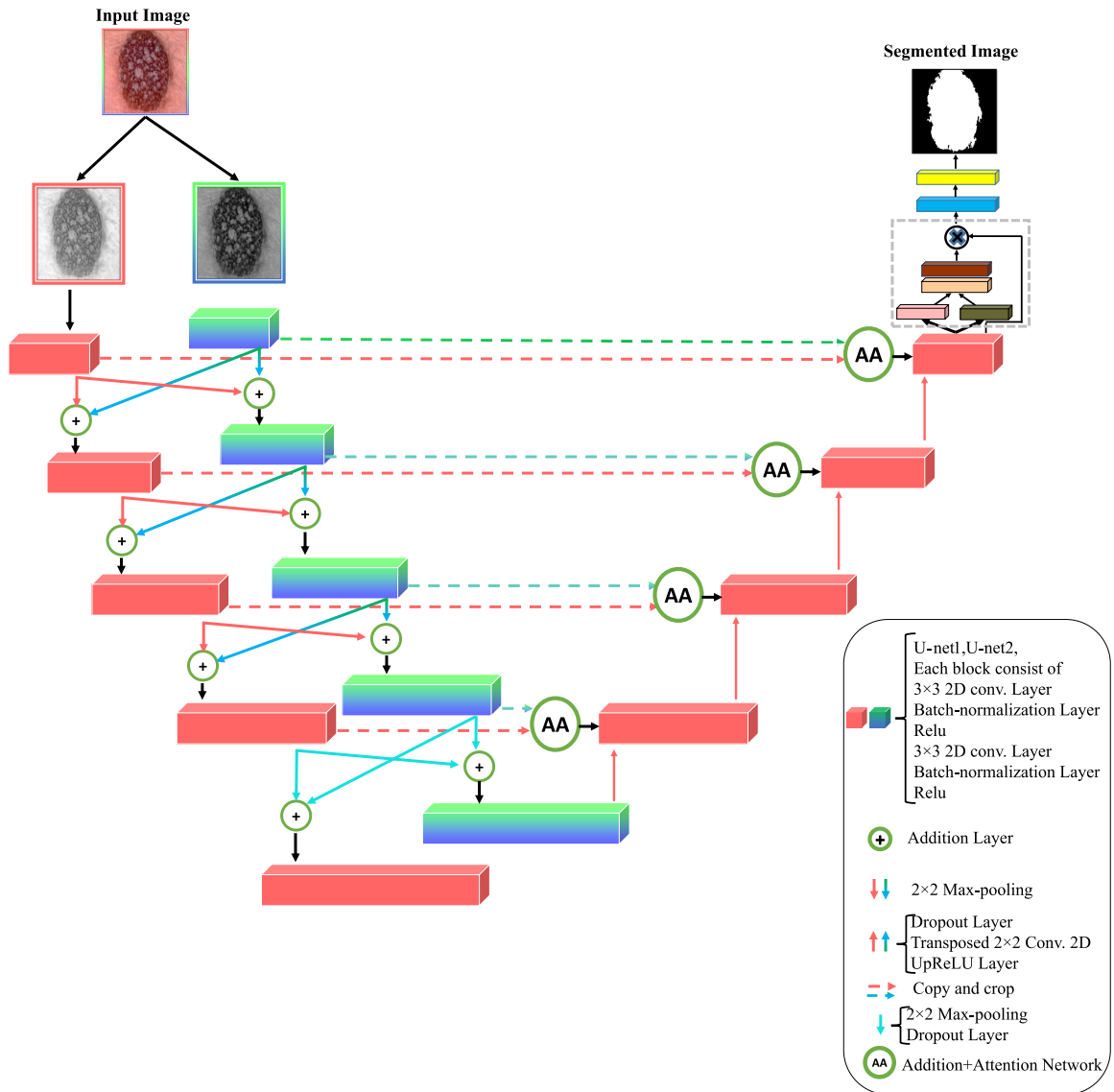


FIGURE 5. The proposed dual input color U-Net (DICU-Net) structure.

fed into the second and third encoder sub-networks, respectively.

- 7) The output feature maps from SICU-Net or from each encoder sub-networks of DICU-Net and TICU-Net models are fed to a simple channel-attention module until they reach the final classification layer.
- 8) The final layer is a convolutional layer with kernel size equal 1×1 to accomplish the semantic segmentation task.
- 9) The training process of the proposed models utilize backpropagation learning algorithm with gradient decent optimization and a composite loss function.

D. CHANNEL-WISE ATTENTION MODULE

The proposed multi-input deep convolutional neural network generates multiple feature maps from each input. It is

beneficial to fuse the extracted features by focusing the attention on significant feature maps to emphasize informative features and suppress redundant features. A channel-wise attention module is employed to draw the convolutional neural network attention towards important color features. Channel-wise attention modules are usually based on learning the interrelation between feature channels. The input to the channel-wise attention module is a set of feature maps $FM_c \in \mathbb{R}^{h \times w \times c}$ where h , w , and c are the number of rows, columns, and channels, respectively. The feature maps fed into the channel-wise attention module are generated by applying a set of convolutional and pooling operations on the contraction path of the input image with different color spaces. The channel attention module used in this work is inspired from [64] by applying global max pool and global average pool operations to get F_{max} and F_{avg} , respectively.

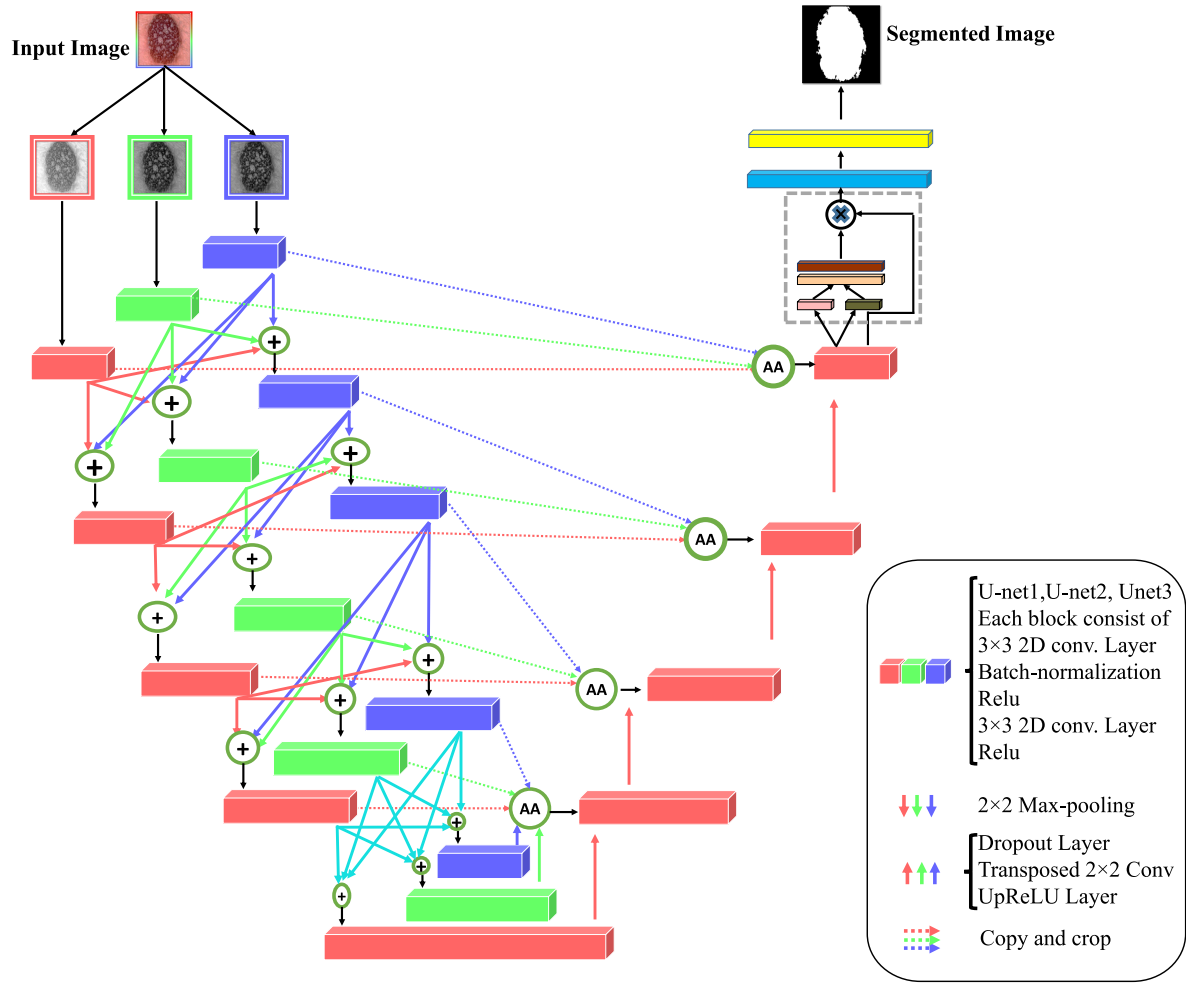


FIGURE 6. The proposed triple input color U-Net (TICU-Net) structure.

Then, the pooled features are concatenated and forwarded by two fully connected layers that reshape the output features into the required number of channels to be passed to the Relu and Sigmoid activation functions. The resulted channel attention map $M_c \in \mathbb{R}^{1 \times 1 \times c}$. where c represents the number of channels is multiplied with the input feature maps to get the required channel weighted map. Figure 7 shows the structure of the channel-wise attention network. The resulted feature map M_c is computed from the following equations:

$$M_c = \sigma(\text{Relu}(F_{c64}(\text{Relu}(F_{c32}(\text{concat}(F_{max}, F_{avg})))))) \quad (1)$$

where σ denotes sigmoid activation function. The final channel attention map CM_c is obtained from:

$$CM_c = FM_c \otimes M_c \quad (2)$$

where \otimes is the element-wise multiplication and $CM_c \in \mathbb{R}^{h \times w \times c}$.

E. PROPOSED COMPOSITE LOSS FUNCTION

Plenty of loss functions are used to improve the performance of deep convolution neural networks used to solve semantic

segmentation problem. The aim of the loss function is to evaluate the performance of the network and minimize the error resulted from the training. In this section, we will explain our proposed composite loss function, but before that, we will mention some parameters that are usually employed to formulate the metrics used to evaluate the network, namely, True Positive (TP), True Negative (TN), False Positive (FP), and False Negative (FN). Pixels in binary skin lesion image segmentation usually belong to one of the two-class objects: foreground (lesion skin) and background (normal skin). The group of pixels that belong to the foreground and are predicted correctly by the model is known as (TP). While the group of pixels that belong to the foreground but the model predicted them incorrectly are known as (FN). And finally, the pixels that do not belong to the foreground but the model predicted them incorrectly are (FP).

Now we can explain our novel proposed composite loss function called Binary Weighted Cross-entropy, Dice, Sensitivity-Specificity (BW-CE-Di-SS) loss function. It consists of three terms combined with binary weights where its components are defined as the following:

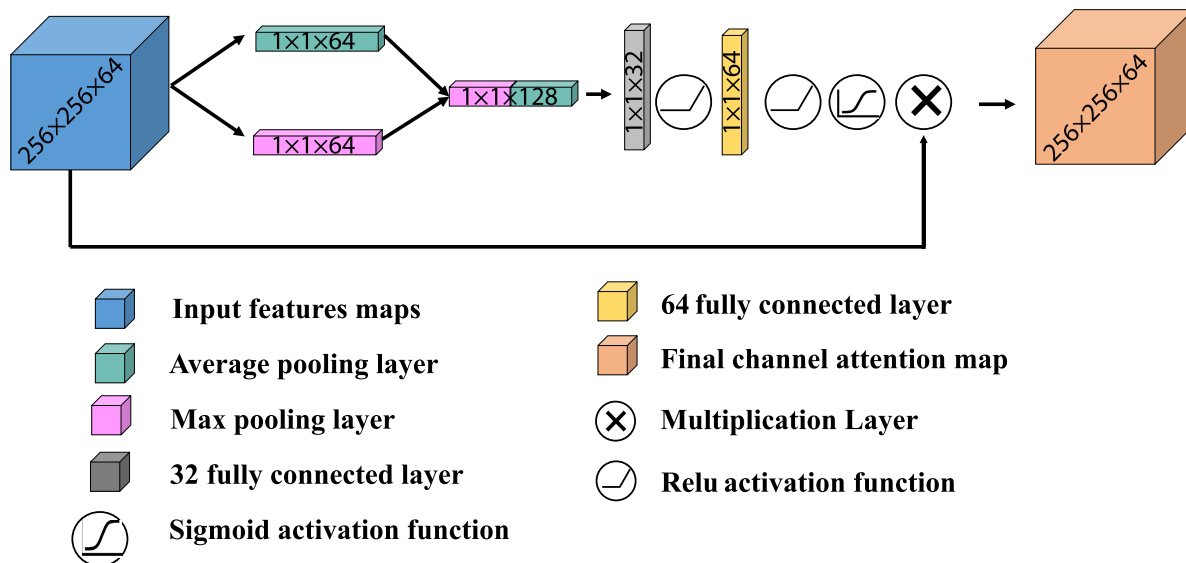


FIGURE 7. The structure of channel-wise attention module.

- Cross Entropy Loss:** cross-entropy is the most common effective metric in the binary segmentation tasks. It is derived from Kullback-Leibler (KL) divergence to measure the model performance whose output is a probability value between 0 and 1. Its value is increased when the predicted probability (Y) close to the actual label (T). It is defined as:

$$L_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{c=1}^C [T_{ic} \log(Y_{ic})] \quad (3)$$

where T_{ic} is the ground-truth label of the pixel i that belongs to class c and (Y_{ic}) is the predicted probability value for pixel i that belongs to the class c . N and C represent the total number of pixels and classes, respectively. Ronneberger *et al.* [30] introduced a Weighted Cross Entropy (WCE) formula to handle the class imbalance problem in binary segmentation tasks as the following equation:

$$L_{WCE} = -[wT \log(Y) + (1 - w)(1 - T) \log(1 - Y)] \quad (4)$$

where w is inversely proportional weight to the class frequencies and it is used to penalize majority classes. In our proposed loss, we assigned 0.7 for the weight w . However, due to high imbalance data in skin lesion segmentation, WCE can not overcome this problem on its own. Hence it is required to combine it with other complementary loss functions.

- Dice Loss:** Dice loss is used as an alternative to cross-entropy to train 3D U-Net [65] and other network architectures. It is derived from *Sørensen-Dice coefficient*, which is a statistical metric used to evaluate the

similarity (overlap) between two samples. This metric ranges from 0 to 1 whereas assigning 1 to the Dice coefficient denotes perfect and complete overlap between the evaluated samples. Milletari *et al.* [65] utilizes *Dice coefficient* in computer vision and define it as the area of overlap divided by the mean of the total number of pixels in both images.

$$Dice = \frac{2|T \cap Y|}{|T| + |Y|} \quad (5)$$

where $|T \cap Y|$ is the common pixels between T and Y . While $|T| + |Y|$ are all pixels in T and all pixels in Y . In image segmentation problems, $|T \cap Y|$ is evaluated as element-wise multiplication between the predicted (Y) and the ground truth mask (T) and then sum the elements in the resulted matrix. On the other hand, to evaluate the denominator, some researchers use the regular sum whereas the other researchers prefer to use the squared sum. We tried both and found that the squared sum gave better results. By applying Dice definition (Eq. 5) to the Boolean data and employ the definition of $\{TP, FP, FN\}$, we can reformulate Eq. 5 into:

$$Dice = \frac{2TP}{2TP + FP + FN} \quad (6)$$

Dice loss weighs (FPs) and (FNs) equally to achieve a better trade-off between precision and recall. The value (2) in the numerator is due to the double count of the common pixels from the union calculation in the denominator, one from T and the other from Y . The dice loss function will be:

$$L_{Dice} = 1 - Dice \quad (7)$$

$$Dice = \frac{2 \sum_{i=1}^N \sum_{c=1}^C (T_{ic} Y_{ic})}{\sum_{i=1}^N \sum_{c=1}^C T_{ic}^2 + \sum_{i=1}^N \sum_{c=1}^C Y_{ic}^2} \quad (8)$$

$$L_{Dice} = 1 - \frac{2 \sum_{i=1}^N \sum_{c=1}^C (T_{ic} Y_{ic})}{\sum_{i=1}^N \sum_{c=1}^C T_{ic}^2 + \sum_{i=1}^N \sum_{c=1}^C Y_{ic}^2} \quad (9)$$

In which Y_{ic} and T_{ic} represent pairs of corresponding i pixel predicted and ground truth values at a specific class c , respectively. Dice loss can deal efficiently with situations where there is a great imbalance between the number of skin lesion and normal pixels and is one of the loss functions that achieved very good results in the semantic segmentation field. However, both WCE and DL can not directly improve the sensitivity and specificity performance metrics.

- **Sensitivity-Specificity Loss:** Due to the imbalance class representation problem in semantic segmentation, a loss function combining sensitivity and specificity is proposed by Brosch *et al.* [66]. *Sensitivity* and *Specificity* are used together to evaluate the classification performance for great unbalanced problems. Sensitivity refers to the ability to detect lesion pixels with the skin disease correctly.

$$Sensitivity = \frac{TP}{TP + FN} \quad (10)$$

Therefore, if a test skin image has 100% sensitivity, this means that the model correctly identifies all skin lesions from normal skins. On the other hand, the specificity refers to its ability to detect normal skin pixels correctly.

$$Specificity = \frac{TN}{TN + FP} \quad (11)$$

Therefore, a test skin image with 100% specificity means that all normal skin pixels are correctly identified. Sensitivity (true positive rate) measures the ratio of actual skin lesion pixels that are correctly classified. While Specificity (true negative rate) measures the ratio of actual normal skin pixels that are correctly classified. By combining them, the final error will measure a weighted sum of the mean squared difference of the lesion pixels (sensitivity) and normal pixels (specificity). The final error was formulated in the following form

$$L_{SS} = \frac{\sum_{i=1}^N \sum_{c=1}^C (T_{ic} - Y_{ic})^2 T_{ic}}{\sum_{i=1}^N \sum_{c=1}^C T_{ic} + \epsilon}$$

TABLE 2. Distribution of ISIC 2017 dataset classes.

Dataset distribution	Clinical diagnosis			Total
	Benign	Melanoma	SK	
Training	1372	374	254	2000
Validation	78	30	42	150
Test	393	117	90	600

$$+ \frac{\sum_{i=1}^N \sum_{c=1}^C (T_{ic} - Y_{ic})^2 (1 - T_{ic})}{\sum_{i=1}^N \sum_{c=1}^C (1 - T_{ic}) + \epsilon} \quad (12)$$

The first term in Eq. (12) is the Sensitivity and the second one refers to specificity.

Due to the class imbalance in the skin lesion images, a binary weighted composite loss function is proposed in this paper in the following way.

$$L_{BW-CE-D-SS} = 1 * L_{WCE} + 2 * L_{Dice} + 4 * L_{SS} \quad (13)$$

L_{WCE} , L_{Dice} , and L_{SS} are defined in Eqs. 4, 9, and 12 respectively. Each term in the proposed loss function has contributed to balance the results of accuracy, true positive and true negative rates. Whereas Dice gave additional weight to false positives and false negatives, which boosted the performance of segmentation results. The specificity term further enhances the true negative rate. Our proposed loss function works in harmony with the proposed multi-input CU-Net to segment color skin lesion images in different color spaces.

IV. EXPERIMENTAL RESULTS

In this section, the segmentation performance of the proposed CU-Net architectures are evaluated using three data sets, namely, ISIC 2017, ISIC 2018 and PH2. International Skin Imaging Collaboration ISIC dataset 2017¹ contains three subsets, the first training subset has 2000 lesion images in JPG format divided into 3 subcategories, 1372 images for benign images, 374 for Melanoma images, and 254 for Seborrheic_Keratosis (SK) and their 2000 corresponding binary mask images in PNG format. The second testing subset includes 600 lesion images for testing in jpg format, 393 benign images, 117 Melanoma images, and 90 Seborrheic_Keratosis (SK) and their 600 corresponding binary mask images in PNG format. The third validation subset includes 150 lesion images for testing in JPG format, 78 benign images, 30 Melanoma images, and 42 Seborrheic_Keratosis (SK) with their 150 corresponding binary mask images. To evaluate our proposed models on this dataset, all images are scaled to (160×224) Figure 8 shows samples from ISIC 2017 dataset. In addition, the class distribution of each skin lesion type is shown in Table 2.

¹<https://challenge2017.isic-archive.com/>

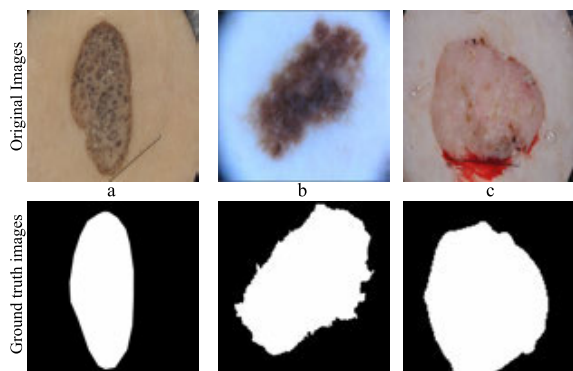


FIGURE 8. Three image samples from each ISIC 2017 dataset class (top row) with its corresponding ground truth (bottom row) (a) Melanoma class, (b) Benign class, and (c) Seborrheic_Keratosis class.

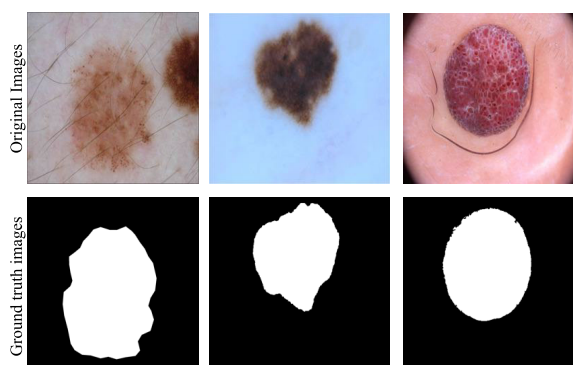


FIGURE 9. Three image samples from each ISIC 2018 dataset class (top row) with its corresponding ground truth (bottom row).

The ISIC 2018² dataset was published by the International Skin Imaging Collaboration (ISIC) as a large-scale dataset of dermoscopy images. The training dataset of ISIC 2018 contains 2594 RGB dermoscopic images that have a resolution ranged from (576×768) to (6748×4499) in jpg format. We divided the training images into 80% (2076) images for training and 20% (518) images for testing with size (256×256). Figure 9 shows samples from ISIC 2018 dataset.

PH2 dataset was created at the Hospital Pedro Hispano with the help of the research group Universidade do Porto, Técnico Lisboa in Matosinhos, Portugal. The PH2 dataset consists of 200 skin lesions in total, 80 of them including Common Nevus cases, 80 for Atypical Nevus, and 40 for Melanoma cases. Images in the PH2 dataset were captured under the same conditions, and the size of all images in this dataset was 768 × 560 pixels. Segmentation masks of this dataset were drawn by expert dermatologists. In this study, the PH2 dataset is used only for testing due to its small size for deep CNN training, thus we utilized ISIC 2017 dataset in the model training and apply PH2 for testing. All images are re-scaled to (160×224). The class distribution of lesion types

TABLE 3. Distribution of PH2 dataset classes,(CN) is Common Nevus, and (AN) is Atypical Nevus.

Dataset distribution	Clinical diagnosis			Total
	CN	AN	Melanoma	
PH2	80	80	40	200

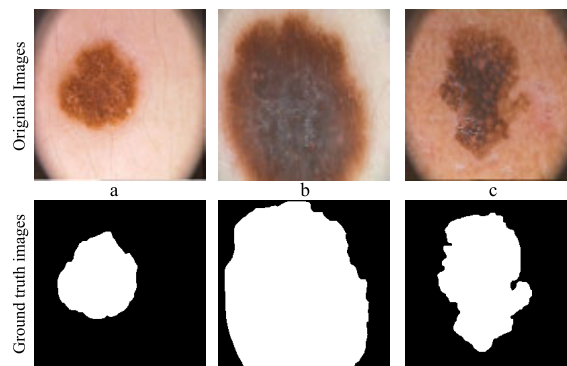


FIGURE 10. Three image samples from each PH2 dataset class (top row) and their corresponding ground truth (bottom row) (a) Common Nevus class, (b) Atypical Nevus class, and (c) Melanoma class.

TABLE 4. Effect of changing input color space on the SICU-Net performance without using any attention modules or image adjustment for ISIC 2017 dataset.

Color space	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB	89.45	88.14	89.87	80.28	67.06
Gray	91.67	78.43	95.96	82.17	69.73
YCbCr	91.88	76.80	96.74	82.17	69.74
YIQ	90.38	65.22	98.48	76.76	62.29
XYZ	92.07	73.63	98.02	81.91	69.36
HSV	89.66	61.07	98.88	74.22	59.01
Lab	88.42	54.94	99.21	69.81	53.62
LUV	91.60	75.68	96.76	81.52	68.80

TABLE 5. Effect of changing input color space on SICU-Net performance without using attention module or image adjustment for PH2 dataset.

Color space	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB	89.94	81.82	93.80	83.98	72.39
Gray	89.92	90.52	81.64	85.01	73.93
YCbCr	91.18	89.42	92.02	86.73	76.56
YIQ	90.23	82.96	93.68	84.55	73.23
XYZ	89.73	83.37	92.76	83.96	72.35
HSV	91.30	87.06	93.31	86.57	76.32
Lab	90.23	81.62	94.33	84.34	72.92
LUV	91.68	87.99	93.43	87.21	77.31

for PH2 dataset is shown in Table 3. Figure 10 shows some image samples from each class of PH2 dataset.

In order to make a fair comparison with other state-of-the-art deep learning models, we used the standard testing protocol (train/valid/test) that commonly utilized for ISIC 2017 datasets. The validation dataset of the ISIC 2017 database is utilized to adjust the hyperparameters of the proposed models while we used the test dataset to calculate the evaluation metrics of the proposed models.

²<https://challenge2018.isic-archive.com/task1/>

TABLE 6. Effect of changing input color space on the DICU-Net performance without using attention module or image adjustment using ISIC 2017 dataset.

Color space	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-Gray	90.98	86.59	92.40	82.45	70.14
RGB-YCbCr	91.81	84.87	94.06	83.53	71.72
RGB-HSV	91.86	78.46	96.20	82.50	70.22
RGB-YIQ	91.48	82.05	94.54	82.50	70.21
RGB-XYZ	90.37	88.12	91.10	81.75	69.13
RGB-Lab	91.63	81.12	95.04	82.59	70.35
RGB-LUV	90.54	87.93	91.38	81.98	69.46
Gray-YCbCr	90.85	86.67	92.20	82.26	69.86
Gray-YIQ	92.36	81.22	95.98	83.89	72.24
Gray-XYZ	92.09	78.12	96.61	82.85	70.73
Gray-HSV	92.04	77.06	96.89	82.57	70.31
Gray-Lab	91.45	83.43	94.05	82.69	70.49
Gray-LUV	91.72	87.33	79.84	84.89	57.51
XYZ-YCbCr	91.35	82.01	94.38	82.27	69.89
XYZ-YIQ	90.71	66.61	98.52	77.83	63.70
XYZ-HSV	92.36	82.64	95.51	84.11	72.58
XYZ-Lab	90.69	88.57	91.38	82.32	69.95
XYZ-LUV	91.91	74.73	97.48	81.89	69.33

TABLE 7. Effect of changing input color space on DICU-Net performance without using attention module or image adjustment using PH2 dataset.

Color space	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-Gray	90.52	87.33	92.04	85.58	74.80
RGB-YCbCr	90.93	90.92	90.93	86.59	76.36
RGB-HSV	91.49	88.25	93.03	86.99	76.97
RGB-YIQ	91.19	85.91	93.71	86.28	75.87
RGB-XYZ	90.24	78.51	95.82	83.84	72.17
RGB-Lab	90.80	91.20	90.62	86.47	76.17
RGB-LUV	90.57	88.36	91.62	85.80	75.12
Gray-YCbCr	90.88	91.20	90.73	86.57	76.32
Gray-HSV	91.79	90.43	92.43	87.65	78.01
Gray-YIQ	92.18	88.15	94.09	87.89	78.40
Gray-XYZ	90.51	81.41	94.84	84.69	73.44
Gray-Lab	90.74	91.99	90.14	86.49	76.19
Gray-LUV	90.46	86.59	92.30	85.40	74.52
LUV-YCbCr	90.98	92.29	90.36	86.83	76.72
LUV-HSV	91.28	89.95	91.92	86.93	76.88
LUV-YIQ	90.38	84.10	93.37	84.93	73.80
LUV-Lab	91.08	89.04	92.04	86.54	76.27
LUV-XYZ	89.87	83.90	92.71	84.22	72.75
Lab-YCbCr	91.36	89.19	92.39	86.93	76.89
Lab-HSV	91.13	90.81	91.29	86.84	76.75
Lab-YIQ	90.43	87.30	91.92	85.46	74.62
Lab-XYZ	89.55	84.48	91.97	83.90	72.27

The experiments conducted in this paper are implemented using Matlab 2020a running on Windows 10 using PC with Intel core i7 processor, 32 GB RAM, and (NVIDIA GeForce RTX 2080 Ti). Regarding the overfitting issue, we employed two techniques to avoid overfitting (1) using L2 regularization in the loss function, and (2) using dropout layers. As for training stopping we tried different numbers of training epochs and chose the one that give the best results. The number of epochs was 30, SGDM optimizer was used to train all proposed models with the following training hyperparameters: the learning rate was 0.05, and the mini-batch size was set as 4. The following experiments are conducted to investigate different aspects of our proposed method.

The evaluations of the proposed models are presented using common performance metrics such as: true positive rate/sensitivity (SEN) defined in Eq.10, and true negative

TABLE 8. Effect of changing input color space on TICU-Net performance without using attention module or image adjustment using ISIC 2017 dataset.

Color space	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-YCbCr-Gray	91.60	86.21	93.34	83.39	71.52
RGB-YCbCr-YIQ	91.96	85.00	94.22	83.81	72.13
RGB-YCbCr-XYZ	91.68	85.46	93.69	83.40	71.53
RGB-YCbCr-LUV	92.38	76.99	97.36	83.18	71.20
RGB-LUV-Gray	92.01	78.57	96.36	82.79	70.64
RGB-LUV-YIQ	90.35	87.98	91.12	81.70	69.06
RGB-LUV-XYZ	91.95	76.74	96.88	82.35	69.99
RGB-XYZ-Gray	92.06	80.76	95.72	83.27	71.33
RGB-XYZ-YIQ	91.88	76.04	97.02	82.09	69.63
RGB-YIQ-Gray	92.11	82.70	95.16	83.69	71.95
Gray-YCbCr-LUV	91.96	83.86	94.59	83.62	71.85
Gray-YCbCr-YIQ	92.34	83.63	95.16	84.23	72.76
Gray-YCbCr-XYZ	92.38	83.63	95.21	84.30	72.86
Gray-LUV-YIQ	92.01	81.22	95.50	83.26	71.33
Gray-LUV-XYZ	92.13	76.03	97.34	82.54	70.27
XYZ-HSV-Gray	90.28	88.49	90.86	81.67	69.02
XYZ-HSV-Gray	86.26	92.18	84.34	76.65	62.14
XYZ-HSV-YCbCr	91.53	85.39	93.52	83.14	71.15
XYZ-HSV-Lab	92.00	77.81	96.59	82.63	70.41
XYZ-HSV-LUV	91.77	83.58	94.43	83.26	71.32

TABLE 9. Effect of changing input color space on TICU-Net performance without using attention module or image adjustment using PH2 dataset.

Color space	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-Lab-Gray	89.79	94.49	87.56	85.65	74.90
RGB-LUV-Gray	90.06	93.48	88.44	85.84	75.19
RGB-LUV-Lab	89.91	93.75	88.09	85.69	74.97
Lab-LUV-Gray	90.42	93.21	89.10	86.25	75.83
RGB-XYZ-Gray	90.26	83.32	93.56	84.65	73.38
RGB-XYZ-Lab	90.43	93.32	89.06	86.27	75.85
RGB-XYZ-LUV	91.80	90.65	92.34	87.69	78.08
RGB-XYZ-YIQ	91.85	90.34	92.57	87.72	78.13
RGB-XYZ-YCbCr	90.56	91.76	89.99	86.24	75.81
Gray-YIQ-RGB	89.17	94.42	86.68	84.90	73.76
Gray-YIQ-lab	89.98	90.79	89.60	85.38	74.49
Gray-YIQ-LUV	90.57	89.59	91.03	85.96	75.37
Gray-YIQ-XYZ	92.17	86.61	94.81	87.69	78.08
Gray-YIQ-YCbCr	89.61	94.39	87.34	85.41	74.54
LUV-YCbCr-RGB	90.69	93.07	89.56	86.57	76.32
LUV-YCbCr-Gray	89.06	94.37	86.54	84.76	73.55
LUV-YCbCr-Lab	90.26	93.87	88.54	86.13	75.64
LUV-YCbCr-XYZ	90.10	93.94	88.28	85.94	75.35
LUV-YCbCr-YIQ	90.57	88.00	91.80	85.75	75.05

rate/specificity (SPE) defined in Eq. 11, Dice coefficient (DIC) defined in Eq. 6. While the Jaccard index (JAC) and Accuracy (ACC) metrics are defined as:

$$JAC = \frac{TP}{TP + FP + FN} \quad (14)$$

$$ACC = \frac{TP + TN}{TP + FP + FN + TN} \quad (15)$$

where, TP, TN, FP, and FN have been clarified in subsection III-E

A. EFFECT OF CHANGING COLOR SPACE ON PROPOSED CU-NET MODELS

The first experiment evaluates the contribution of each color space without using any attention modules or image enhancement. All three models are trained and tested using ISIC 2017 and PH2 datasets. Tables 4 and 5 show the effect

TABLE 10. Effect of using channel-attention modules with SICU-Net using ISIC 2017 dataset.

Color space	Attention Module	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB	with	91.22	69.42	98.29	79.47	65.93
	without	89.45	88.14	89.87	80.28	67.06
Gray	with	92.03	73.98	97.88	81.95	69.43
	without	91.67	78.43	95.96	82.17	69.73
YCbCr	with	91.32	69.29	98.42	79.56	66.06
	without	91.88	76.80	96.74	82.17	69.74
YIQ	with	92.53	84.15	95.24	84.65	73.38
	without	90.38	65.22	98.48	76.76	62.29
XYZ	with	92.48	75.64	97.91	83.06	71.03
	without	92.07	73.63	98.02	81.91	69.36
HSV	with	92.01	76.75	96.95	82.46	70.16
	without	89.66	61.07	98.88	74.22	59.01
Lab	with	91.73	71.75	98.17	80.88	67.90
	without	88.42	54.94	99.21	69.81	53.62
LUV	with	92.02	78.47	96.42	82.80	70.65
	without	91.60	75.68	96.76	81.52	68.80

TABLE 11. Effect of using channel attention modules with SICU-Net using PH2 dataset.

Color space	Attention Module	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB	With	91.28	89.66	92.05	86.89	76.82
	without	89.94	81.82	93.80	83.98	72.39
Gray	with	91.02	93.07	90.05	86.98	76.96
	without	89.92	90.52	81.64	85.01	73.93
YCbCr	with	92.85	93.01	92.77	89.34	80.74
	without	91.18	89.42	92.02	86.73	76.56
YIQ	with	91.81	92.91	91.29	87.97	78.52
	without	90.23	82.96	93.68	84.55	73.23
XYZ	with	91.57	89.46	92.58	87.25	77.38
	without	89.73	83.37	92.76	83.96	72.35
HSV	with	92.51	88.85	94.25	88.43	79.27
	without	91.30	87.06	93.31	86.57	76.32
Lab	with	91.88	88.37	93.55	87.52	77.82
	without	90.23	81.62	94.33	84.34	72.92
LUV	with	92.46	91.49	92.92	88.66	79.63
	without	91.68	87.99	93.43	87.21	77.31

of changing the input color space on the SICU-Net performance by using ISIC 2017 and PH2 datasets, respectively. For ISIC 2017 dataset, Table 4 shows that using XYZ color space achieves the highest ACC while RGB achieved the highest TPR and Lab color space achieved the highest TNR. The best values for DIC and JAC coefficients are accomplished using YCbCr. Results in Table 5 for PH2 dataset show that using LUV color space for the input image achieved the best value for ACC, DIC and JAC coefficients. While the gray image accomplished the best value for TPR, and Lab achieved best TNR.

For DICU-Net which accepts two different color spaces, Tables 6 and 7 summarize the results of changing the input color spaces using ISIC 2017 and PH2 datasets, respectively. In this experiment, we tried many color space combinations to find the best ones. For ISIC 2017, Table 6 shows that the XYZ color space achieves superiority over other color spaces when it is combined with other color spaces. While for PH2 dataset, Table 7 shows that the combination of gray and YIQ color spaces give better results than using each of them separately.

TABLE 12. Effect of using channel attention modules with DICU-Net using ISIC 2017 dataset.

Color space	Attention Module	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-YCbCr	with	92.11	84.51	94.57	83.98	72.38
	without	91.81	84.87	94.06	83.53	71.72
RGB-XYZ	with	91.48	80.68	94.98	82.26	69.86
	without	90.37	88.12	91.10	81.75	69.13
RGB-HSV	with	91.92	78.82	96.16	82.68	70.47
	without	92.36	82.64	95.51	84.11	72.58
Gray-YIQ	with	92.15	76.00	97.38	82.58	70.32
	without	92.36	81.22	95.98	83.89	72.24
Gray-HSV	with	92.24	75.80	97.57	82.70	70.51
	without	92.04	77.06	96.89	82.57	70.31
Gray-LUV	with	92.14	77.93	96.74	82.91	70.80
	without	91.72	87.33	79.84	84.89	57.51
XYZ-YIQ	with	92.10	77.09	96.96	82.68	70.48
	without	90.71	66.61	98.52	77.83	63.70
XYZ-HSV	with	92.28	77.50	97.07	83.09	71.07
	without	92.36	82.64	95.51	84.11	72.58
XYZ-Lab	with	91.98	76.84	96.88	82.42	70.10
	without	90.69	88.57	91.38	82.32	69.95

TABLE 13. Effect of using channel attention modules with DICU-Net using PH2 dataset.

Color space	Attention Module	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-YCbCr	with	91.99	94.88	90.62	88.42	79.25
	without	90.93	90.92	90.93	86.59	76.36
RGB-XYZ	with	92.06	93.79	91.24	88.39	79.19
	without	90.24	78.51	95.82	83.84	72.17
RGB-HSV	with	91.27	91.85	90.99	87.15	77.22
	without	91.49	88.25	93.03	86.99	76.97
RGB-Lab	with	91.64	94.98	90.06	87.99	78.55
	without	90.80	91.20	90.62	86.47	76.17
Gray-YIQ	with	92.42	92.20	92.53	88.69	79.68
	without	92.18	88.15	94.09	87.89	78.40
Gray-Lab	with	91.84	93.03	91.28	88.02	78.61
	without	90.74	91.99	90.14	86.49	76.19
Gray-XYZ	with	91.89	90.55	92.53	87.80	78.25
	without	90.51	81.41	94.84	84.69	73.44
Gray-LUV	with	91.86	90.18	92.66	87.72	78.13
	without	90.46	86.59	92.30	85.40	74.52
LUV-YCbCr	with	92.12	90.65	92.82	88.12	78.76
	without	91.28	89.95	91.92	86.93	76.88
LUV-YIQ	with	92.67	94.28	91.91	89.24	80.57
	without	90.38	84.10	93.37	84.93	73.80
Lab-YCbCr	with	91.67	93.34	90.87	87.83	78.31
	without	91.36	89.19	92.39	86.93	76.89
Lab-HSV	with	92.47	92.70	92.37	88.81	79.87
	without	91.13	90.81	91.29	86.84	76.75

For the proposed TICU-Net, a combination of three different color spaces is fed to the network. Table 8 shows the effect of changing the three color space combination on the evaluation metric values. For ISIC 2017 data set, using ‘RGB-YCbCr-LUV’ color spaces achieved the best ACC and TNR values, while using the ‘Gray-YCbCr-XYZ’ combination achieved the best ACC, DIC, and JAC values. The best TPR is obtained using ‘XYZ-HSV-Gray’ color spaces combination. For PH2 dataset, Table 9 shows that the ‘Gray-YIQ-XYZ’ color spaces combination accomplished the best ACC, SPE, DIC, and JAC values, while using ‘RGB-Lab-Gray’ combination achieved the best value for

TABLE 14. Effect of using channel attention modules with TICU-Net using ISIC 2017 dataset.

Color space	Attention Module	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-YCbCr-YIQ	with	92.46	85.01	94.87	84.65	73.39
	without	91.96	85.00	94.22	83.81	72.13
RGB-LUV-Gray	with	92.12	80.87	95.77	83.40	71.52
	without	92.01	78.57	96.36	82.79	70.64
RGB-YCbCr-LUV	with	91.79	79.99	95.62	82.67	70.46
	without	92.38	76.99	97.36	83.18	71.20
RGB-YIQ-Gray	with	92.83	81.32	96.57	84.74	73.52
	without	92.11	82.70	95.16	83.69	71.95
RGB-XYZ-Gray	with	92.26	82.19	95.53	83.87	72.22
	without	92.06	80.76	95.72	83.27	71.33
Gray-YCbCr-LUV	with	92.37	82.17	95.67	84.05	72.49
	without	91.96	83.86	94.59	83.62	71.85
Gray-YCbCr-YIQ	with	91.95	86.20	93.82	83.98	72.39
	without	92.34	83.63	95.16	84.23	72.76
Gray-YCbCr-XYZ	with	91.92	82.35	95.02	83.30	71.38
	without	92.38	83.63	95.21	84.30	72.86
XYZ-HSV-Gray	with	92.59	80.15	96.62	84.12	72.59
	without	86.26	92.18	84.34	76.65	62.14
XYZ-HSV-LUV	with	92.27	79.35	96.46	83.40	71.53
	without	91.77	83.58	94.43	83.26	71.32
Gray-LUV-YIQ	with	92.27	84.55	94.77	84.26	72.80
	without	92.01	81.22	95.50	83.26	71.33

TABLE 15. Effect of using channel attention modules with TICU-Net using PH2 dataset.

Color space	Attention Module	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-LUV-Gray	with	92.24	93.04	91.86	88.54	79.43
	without	90.06	93.48	88.44	85.84	75.19
Lab-LUV-Gray	with	93.34	91.76	94.09	89.88	81.62
	without	90.42	93.21	89.10	86.25	75.83
RGB-Lab-Gray	with	92.07	91.39	92.39	88.14	78.79
	without	89.79	94.49	87.56	85.65	74.90
RGB-XYZ-Gray	with	92.62	92.05	92.89	88.94	80.07
	without	90.26	83.32	93.56	84.65	73.38
RGB-XYZ-Lab	with	92.33	94.70	91.21	88.84	79.92
	without	90.43	93.32	89.06	86.27	75.85
RGB-XYZ-LUV	with	91.80	93.04	91.21	87.97	78.52
	without	91.80	90.65	92.34	87.6	78.08
RGB-XYZ-YIQ	with	91.07	93.39	89.96	87.08	77.11
	without	91.85	90.34	92.57	87.72	78.13
Gray-YIQ-XYZ	with	92.82	90.70	93.83	89.06	80.27
	without	92.17	86.61	94.81	87.69	78.08
Gray-YIQ-RGB	with	92.42	90.21	93.47	88.47	79.32
	without	89.17	94.42	86.68	84.90	73.76
Gray-YIQ-YCbCr	with	92.38	92.65	92.25	88.68	79.67
	without	89.61	94.39	87.34	85.41	74.54
LUV-YCbCr-RGB	with	92.81	93.17	92.64	89.31	80.68
	without	90.69	93.07	89.56	86.57	76.32
LUV-YCbCr-Gray	with	93.06	93.90	92.66	89.70	81.33
	without	89.06	94.37	86.54	84.76	73.55
LUV-YCbCr-YIQ	with	92.70	92.15	92.96	89.06	80.27
	without	90.57	88.00	91.80	85.75	75.05

TPR, also good results for DICE, and JAC is achieved using ‘RGB-XYZ-LUV’.

B. EFFECT OF USING CHANNEL ATTENTION MODULES ON MULTI-INPUT CU-NET

This experiment is conducted by employing same color spaces that are used in previous experiments but by adding all channel attention modules in the proposed CU-Net models.

TABLE 16. Effect of normalizing input image using image adjustment in SICU-Net performance for ISIC 2017.

Color spaces	Normalizing image	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB	with	92.52	81.11	96.22	84.15	72.64
	without	92.03	73.98	97.88	81.95	69.43
YCbCr	with	92.31	77.77	97.02	83.18	71.21
	without	91.32	69.29	98.42	79.56	66.06
YIQ	with	92.29	80.84	96	83.69	71.96
	without	92.53	84.15	95.24	84.65	73.38
XYZ	with	92.63	86.37	94.65	85.15	74.13
	without	92.48	75.64	97.91	83.06	71.03
HSV	with	90.73	68.07	98.06	78.22	64.24
	without	92.01	76.75	96.95	82.46	70.16
Lab	with	91.46	71.51	97.93	80.39	67.21
	without	91.73	71.75	98.17	80.88	67.90
LUV	with	92.21	76.37	97.34	82.75	70.57
	without	92.02	78.47	96.42	82.80	70.65

TABLE 17. Effect of normalizing input image in SICU-Net performance using PH2 dataset.

Color spaces	Normalizing image	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB	with	91.46	95.24	89.67	87.79	78.24
	without	91.28	89.66	92.05	86.89	76.82
Gray	with	91.51	86.81	93.75	86.82	76.72
	without	91.02	93.07	90.05	86.98	76.96
YCbCr	with	92.42	85.25	95.83	87.88	78.38
	without	92.85	93.01	92.77	89.34	80.74
YIQ	with	84.17	56.54	97.30	69.71	53.50
	without	91.81	92.91	91.29	87.97	78.52
XYZ	with	91.75	92.78	91.25	87.87	78.37
	without	91.57	89.46	92.58	87.25	77.38
HSV	with	84.80	62.29	95.50	72.53	56.90
	without	92.51	88.85	94.25	88.43	79.27
Lab	with	91.41	82.06	95.86	86.03	75.48
	without	91.88	88.37	93.55	87.52	77.82
LUV	with	88.54	73.28	95.80	80.48	67.33
	without	92.46	91.49	92.92	88.66	79.63

TABLE 18. Effect of normalizing input image in DICU-Net performance for ISIC 2017.

Color spaces	Normalizing image	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-YCbCr	with	92.35	82.32	95.60	84.04	72.48
	without	92.11	84.51	94.57	83.98	72.38
RGB-XYZ	with	92.12	80.82	95.78	83.38	71.50
	without	91.48	80.68	94.98	82.26	69.86
Gray-HSV	with	92.49	85.38	94.79	84.77	73.56
	without	92.24	75.80	97.57	82.70	70.51
XYZ-Lab	with	92.27	82.44	95.46	83.93	72.31
	without	91.98	76.84	96.88	82.42	70.10
XYZ-YIQ	with	92.17	83.17	95.09	83.87	72.22
	without	92.10	77.09	96.96	82.68	70.48

Table 10 shows the comparison between the SICU-Net segmentation results with and without using channel-wise attention modules for ISIC 2017 dataset. In general, adding channel attention modules improves the value of ACC, DIC, and JAC coefficients in the vast majority of color spaces except some exclusions of Gray and YCbCr color spaces. YIQ color space accomplished the best values for the

TABLE 19. Effect of normalizing input image in DICU-Net performance for PH2 dataset.

Color spaces	Normalizing image	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-Lab	with	90.58	91.95	89.94	86.29	75.88
	without	91.64	94.98	90.06	87.99	78.55
Gray-YIQ	with	92.30	92.89	92.01	88.60	79.53
	without	92.42	92.20	92.53	88.69	79.68
LUV-YCbCr	with	93.23	93.47	93.12	89.90	81.65
	without	92.76	93.59	92.37	89.29	80.65
LUV-HSV	with	91.92	90.17	92.75	87.79	78.24
	without	92.12	90.65	92.82	88.12	78.76
LUV-YIQ	with	93.06	92.98	93.11	89.63	81.20
	without	92.67	94.28	91.91	89.24	80.57

TABLE 20. Effect of normalizing input images on TICU-Net performance using ISIC 2017 dataset.

Color spaces	Normalizing image	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
RGB-YCbCr-YIQ	with	93.21	82.35	96.73	85.58	74.80
	without	92.46	85.01	94.87	84.65	73.39
RGB-YIQ-Gray	with	89.21	91.88	88.34	80.64	67.57
	without	92.83	81.32	96.57	84.74	73.52
Gray-YCbCr-YIQ	with	92.97	79.11	97.46	84.64	73.37
	without	91.95	86.20	93.82	83.98	72.39
XYZ-HSV-Gray	with	93.13	83.64	96.21	85.63	74.88
	without	92.59	80.15	96.62	84.12	72.59

evaluation metrics. Moreover, for PH2 dataset, Table 11 shows the SICU-Net segmentation evaluation metrics. The performance of the SICU-Net is improved by adding the attention modules in all color spaces. YCbCr color space achieves the best values of ACC, DICE, and JAC evaluation metrics. Tables 12 and 13 show the effect of using attention modules in DICU-Net model with the ISIC 2017 and PH2 dataset, respectively. For ISIC 2017 dataset, using attention modules improves the values of ACC, DIC, and JAC when it used with ‘RGB-YCbCr’, ‘RGB-XYZ’, ‘Gray-HSV’, ‘XYZ-YIQ’, and ‘XYZ-Lab’ color spaces, and using the attention modules with ‘RGB-YCbCr’ color spaces achieves the best results in TPR, DIC, and JAC evaluation metrics. For PH2 dataset, using attention modules improves the values of ACC, TPR, DIC, and JAC with all color spaces as shown in Table 13. Using the attention module with ‘LUV-YIQ’ color spaces achieves the best results for ACC, DIC, and JAC evaluation metrics. Table 14 depicts the results of using proposed attention modules with the ISIC 2017 dataset to train and test the TICU-Net with the outlined color spaces. Using a combination of ‘RGB-YCbCr-YIQ’, all evaluation metrics are increased, and using ‘RGB-YIQ-Gray’ color spaces achieved the best values for ACC, TNR, DIC, and JAC. Using ‘Gray-YCbCr-YIQ’, color spaces yield the best TPR, and ‘XYZ-HSV-Gray’ yield the best TNR. Table 15 views the results of using proposed attention modules with the PH2 dataset for testing and ISIC 2017 for training. Using a combination of ‘RGB-XYZ-Lab’, ‘LUV-YCbCr-RGB’, and ‘LUV-YCbCr-YIQ’ increases the values of all evaluation metrics, and using ‘Lab-LUV-Gray’ color

TABLE 21. Effect of using image normalization on TICU-Net performance using PH2 dataset.

Color spaces	Normalizing image	ACC(%)	SEN(%)	SPE(%)	DIC(%)	JAC(%)
Lab-LUV-Gray	with	92.89	92.31	93.16	89.32	80.71
	without	93.34	91.76	94.09	89.88	81.62
RGB-XYZ-Gray	with	89.29	97.94	85.17	85.49	74.66
	without	92.62	92.05	92.89	88.94	80.07
RGB-XYZ-Lab	with	92.24	93.83	91.49	88.63	79.58
	without	92.33	94.70	91.21	88.84	79.92
Gray-YIQ-XYZ	with	89.13	97.43	85.19	85.25	74.29
	without	92.82	90.70	93.83	89.06	80.27
LUV-YCbCr-RGB	with	91.66	88.63	93.10	87.27	77.41
	without	92.81	93.17	92.64	89.31	80.68
LUV-YCbCr-Gray	with	90.98	96.99	88.12	87.39	77.61
	without	93.06	93.90	92.66	89.70	81.33
LUV-YCbCr-YIQ	with	93.16	89.70	94.80	89.42	80.87
	without	92.70	92.15	92.96	89.06	80.27

spaces achieved the best values for ACC, TNR, DIC, and JAC, and Using ‘RGB-XYZ-Lab’ color spaces yield the best TPR.

C. EFFECT OF NORMALIZING INPUT IMAGES ON THE PROPOSED CU-NET MODELS

This experiment is conducted using the same color spaces that are used in subsection IV-B on the proposed CU-Net models except by applying image normalization to the input image. The image normalization technique utilized in this experiment maps the intensity values of each channel of the input image to new values to increase the contrast of the normalized image. The intensity of the normalized image have 1% values saturated at low and high intensities. Using ISIC 2017 dataset, Table 16 shows the comparison between the SICU-Net segmentation results with and without image normalization. The table shows that all evaluation metrics are improved and the best ACC, TPR, DIC and JAC are yielded using XYZ color space, while the best TNR is yielded by using HSV color space. Table 18 indicates the comparison between DICU-Net segmentation results with and without image normalization. The results reveal that all evaluation metrics are increased and the best TNR is achieved with ‘RGB-XYZ’ color space, and the best ACC, TPR, DIC and JAC values are achieved with ‘Gray-HSV’ color spaces. In Table 20, using image normalization with ‘RGB-YCbCr-YIQ’, ‘Gray-YCbCr-YIQ’, and ‘XYZ-HSV-Gray’ color spaces improved the values of ACC, DIC, and JAC, and the best ACC is achieved by ‘RGB-YCbCr-YIQ’, and the best TPR is yielded by using ‘RGB-YIQ-Gray’, while using ‘Gray-YCbCr-YIQ’ achieved the best TNR. and using ‘XYZ-HSV-Gray’ achieved the best DIC and JAC coefficient.

For PH2 dataset, Table 17 shows the comparison between the SICU-Net segmentation results with and without image normalization. The results show that all evaluation metrics are improved and the best ACC, DIC and JAC are achieved with YCbCr color space, the best TNR is yielded from YIQ color space, and the best TPR is yielded from RGB color space.

TABLE 22. Evaluation of segmentation performance TICU-Net according to each loss function for ISIC 2017.

Color space	Loss function	ACC%	TPR%	TNR%	DIC%	JAC%
XYZ-HSV-Gray	Dice loss	92.51	76.03	97.85	83.25	71.30
	Sens. Spec. loss	91.63	87.67	92.91	83.68	71.94
	Cross entropy loss	91.68	69.72	98.79	80.39	67.21
	Proposed composite loss	93.13	83.64	96.21	85.63	74.88

TABLE 23. ACC (Accuracy), TPR (True positive rate), and TNR (True negative rate) evaluation of our proposed models and state-of-the-art segmentation architectures on ISIC 2017 training and testing set.

Method	Benign cases			Melanoma cases			SK cases			Overall		
	ACC%	TPR%	TNR%	ACC%	TPR%	TNR%	ACC%	TPR%	TNR%	ACC%	TPR%	TNR%
FCN [66]	94.45	85.25	96.95	88.25	70.67	96.18	90.96	75.10	95.91	92.72	79.98	96.66
U-Net [29]	92.89	76.76	97.26	84.98	58.71	96.81	84.83	43.81	97.64	90.14	67.15	97.24
SegNet [67]	93.93	85.19	96.30	87.90	73.78	94.26	87.29	70.58	92.50	91.76	80.05	95.37
FrCN [18]	95.62	88.95	97.44	90.78	78.94	96.04	91.29	82.37	94.08	94.03	85.40	96.69
DeeplabV3+ [68]	95.67	88.54	97.21	89.65	77.71	96.37	90.06	82.37	94.08	94.03	85.40	97.25
SICU-Net [XYZ]	93.83	78.76	98.16	89.22	75.61	95.68	88.04	74.59	98.55	93.66	84.34	94.65
DICU-Net [Gray-HSV]	94.19	90.69	95.19	89.69	77.20	95.61	88.72	75.46	93.06	92.12	80.82	95.78
TICU-Net [XYZ-Gray-HSV]	94.96	83.77	98.18	88.71	70.02	97.58	88.73	69.09	95.15	93.13	83.64	96.21

TABLE 24. DIC (Dice coefficient), and JAC (Jaccard coefficient) evaluation of our proposed models and state-of-the-art segmentation architectures on ISIC 2017 training and testing set.

Method	Benign cases		Melanoma cases		SK cases		Overall	
	DIC%	JAC%	DIC%	JAC%	DIC%	JAC%	DIC%	JAC%
FCN [66]	86.77	76.63	78.89	65.14	79.81	66.40	83.83	72.17
U-Net [29]	82.16	69.72	70.82	54.83	57.88	40.73	76.27	61.64
SegNet [67]	85.69	74.97	79.11	65.45	72.54	56.91	82.09	69.63
FrCN [18]	89.63	81.28	84.02	72.44	81.83	69.25	87.08	77.11
DeeplabV3+ [68]	88.29	81.09	80.86	71.30	77.05	67.55	85.16	77.15
SICU-Net [XYZ]	85.07	74.02	81.87	69.30	75.07	60.09	85.15	74.13
DICU-Net [Gray-HSV]	87.45	77.70	82.82	70.67	76.74	62.25	83.38	71.50
TICU-Net [XYZ-Gray-HSV]	88.13	78.78	79.97	66.62	75.13	60.17	85.63	74.88

TABLE 25. ACC (Accuracy), TPR (True positive rate), and TNR (True negative rate) evaluation of our proposed models and state-of-the-art segmentation architectures on ISIC 2017 training set and PH2 testing set.

Method	Benign cases			Melanoma cases			Overall		
	ACC%	TPR%	TNR%	ACC%	TPR%	TNR%	ACC%	TPR%	TNR%
FCN [66]	94.49	95.35	94.09	92.82	90.30	94.02	92.82	90.30	94.02
U-Net [29]	94.66	86.68	97.63	84.36	70.58	98.47	92.55	81.63	97.76
SegNet [67]	95.19	91.57	96.57	86.04	75.50	96.83	93.36	86.53	96.61
FrCN [18]	95.20	94.48	95.46	94.64	91.57	96.55	95.08	93.72	95.65
SICU-Net [YCbCr]	95.32	97.63	94.57	86.00	87.51	83.40	91.75	92.78	91.25
DICU-Net [LUV-YCbCr]	92.96	99.68	90.78	86.80	86.98	86.47	93.23	93.47	93.12
TICU-Net [LUV-YCbCr-YIQ]	95.08	98.64	93.92	81.54	74.60	93.47	93.16	89.70	94.80

Table 19 displays the comparison between the DICU-Net segmentation results with and without image normalization. The table shows that the best evaluation metrics are achieved with ‘LUV-YCbCr’ color spaces. Table 21 shows the effect of using image normalization on the segmentation results of TICU-Net. Using image normalization improves the results of TPR for most of all the outlined color space combinations. The best TNR, DIC and JAC are achieved with LUV-YCbCr-YIQ color spaces, the best TPR is yielded from RGB-XYZ-Gray combination color spaces, and the best ACC is achieved from Lab-LUV-Gray color spaces combination. Figure 11 illustrates examples of the resulted predicted images obtained from SICU-Net (Third row), DICU-Net (fourth row), and TICU-Net (fifth row) models for both ISIC 2017 and PH2

datasets. The images in the figure are chosen randomly from each class of the two dataset.

D. STUDDING SEGMENTATION PERFORMANCE ACCORDING TO EACH LOSS FUNCTION COMPONENT

This experiment investigates the effect of the proposed composite loss function which contains Dice loss, sensitivity and specificity loss, and Cross Entropy loss function. Table 22 displays the effect of applying each loss function separately on the ‘XYZ-HSV-Gray’ color space combination for TICU-Net model. The results reveal that the proposed composite loss function outperforms each individual loss function

TABLE 26. DIC (Dice coefficient), and JAC (Jaccard coefficient) evaluation of our proposed models and state-of-the-art segmentation architectures on ISIC 2017 training set and PH2 testing set.

Method	Benign cases		Melanoma cases		Overall	
	DIC%	JAC%	DIC%	JAC%	DIC%	JAC%
FCN [66]	90.96	82.59	89.03	80.22	89.03	80.22
U-Net [29]	89.88	80.63	82.04	69.55	87.61	77.95
SegNet [67]	91.32	84.03	84.55	73.23	89.36	80.77
FrCN [18]	91.38	84.13	92.92	86.77	91.77	84.79
SICU-Net [YCbCr]	91.07	83.61	88.77	79.81	87.87	78.37
DICU-Net [LUV-YCbCr]	87.38	77.59	89.29	80.64	89.90	81.65
TICU-Net [LUV-YCbCr-YIQ]	90.74	83.05	83.64	71.87	89.42	80.87

TABLE 27. Performance evaluation of different segmentation algorithms on ISIC 2017 dataset.

Method	Data selection	Year	ACC%	SEN%	SPE%	DIC%	JAC%
U-Net [29]	ISIC 2017 dataset contains 2000 training images,150 validation images and 600 testing images, respectively.	2015	91.64	81.72	96.80	81.59	72.34
SSP [69]		2018	92.14	83.18	97.15	83.04	75.21
SegAN [70]		2018	94.10	-	-	86.70	78.50
DDN [71]		2018	92.35	83.64	95.97	84.29	75.27
AttU-Net [72]		2019	91.45	79.98	97.76	80.82	71.73
DSM [73]		2019	92.86	83.72	96.58	84.15	75.72
LCASA-Net [74]		2019	94.70	-	-	87.90	80.90
ERU [75]		2020	91.98	82.97	96.62	84.13	75.18
SESV [76]		2020	92.23	83.26	96.68	83.92	75.31
ME-DCNN [8]		2020	93.11	83.25	96.84	84.27	76.03
DAGAN [77]		2020	93.04	83.63	97.16	84.25	75.94
FAT-Net [78]		2021	93.26	83.92	97.25	85.00	76.53
U-Nets [52]		The available 2000 training image set is divided into 1800 images for training and the remaining 200 images for test of ISBI 2017	2018	-	-	-	77.00
SLS-Deep [79]	2000 challenge training images, and 600 challenge testing images of ISBI 2017	2018	93.60	81.60	98.30	87.80	78.20
SLS-Net [26]		2018	97.61	87.81	99.92	90.63	81.98
DCL-PSI [80]		2019	94.10	86.20	96.70	85.70	77.80
CDNN [60]		2019	93.40	82.50	97.50	84.90	76.50
FrCN [18]	The official 2000 dermoscopy training images for training, and 600 dermoscopy images test dataset of the ISBI 2017	2018	94.03	85.40	96.69	87.08	77.11
YOLOv3 [21]		2019	93.39	90.82	92.68	84.26	74.81
Fuzzy Classification [62]	The training data set and test data set were used with 2000 and 600 images respectively in the 2017 ISBI Challenge	2019	88.40	86.90	92.30	76.00	66.50
R2U-net [58]	The official 2000 deormoscopy images for training of ISIC 2017 are divided into 1250 samples for training,150 samples for validation data, and 600 samples as test	2019	94.24	94.14	94.25	-	94.21
MCGU-Net [59]		2020	95.70	85.02	98.55	-	95.70
FRCU-Net [53]		2021	97.27	91.50	98.61	-	97.27
SICU-Net [XYZ]	The official 2000 challenge training images, and 600 challenge testing images of ISIC 2017		92.63	86.37	94.65	85.15	74.13
DICU-Net [Gray-HSV]		2022	92.49	85.38	94.79	84.77	73.56
TICU-Net [XYZ-Gray-HSV]			93.13	83.64	96.21	85.63	74.88

which leads to improve the accuracy by 0.62%, Dice by 1.95% and Jaccard coefficient by 2.94 %.

E. COMPARISON WITH STATE-OF-THE-ART METHODS

This experiment introduces a comparison between our proposed models and other state-of-the-art methods that utilize

the same dataset. The compared methods were obtained from [18], [23], [59], [78], [83].

Tables 23 and 24 present the performance of our proposed models with other recently developed fully convolutional networks such as: FCN [84], U-Net [30], SegNet [79], FrCN [18], DeeplabV3+ [67] which used ISIC 2017 at the level of dataset categories. Tables show that the proposed

TABLE 28. Performance evaluation of different segmentation algorithms on ISIC 2018 dataset.

Method	Data selection	Year	ACC%	SEN%	SPE%	DIC%	JAC%
U-Net [29]		2015	94.04	88.00	96.97	85.45	77.33
AttU-Net [72]		2019	93.76	86.00	98.26	85.66	77.64
ResUNet++ [81]	The ISIC 2018 (2594) training image dataset is divided into different sets of training (2044) and testing (550).	2019	93.82	87.35	97.21	85.36	77.21
FTL [33]		2018	94.12	87.54	96.32	86.93	78.25
ERU [75]	The ISIC 2018 (2594) training image dataset is divided into two parts with ratio 4:1 as training sets and validation sets	2020	94.35	90.32	96.92	88.12	80.56
DAGAN [77]		2020	93.24	90.72	95.88	88.07	81.13
CKD-Net [82]	The ISIC 2018 (2594) training image dataset is divided into 2296 images for training and 300 images for testing randomly	2021	94.92	90.55	97.01	87.79	80.41
FAT-Net [78]		2021	95.78	91.00	96.99	89.03	82.02
CPFNet [83]	Split the data for into five fold cross-validation	2020	96.3	-	-	89.9	82.9
AFLN-DGCL [23]		2021	96.30	-	-	90.00	83.50
Attn_U-Net+GN [9]	The ISIC 2018 (2594) training image dataset is divided into different sets of training (2044) and testing (550).	2020	95.00	94.00	95.00	91.00	83.00
BCDU-Net [61]		2019	93.70	78.50	98.20	-	93.70
FRCU-Net [53]	The official 2594 images for training of ISIC 2018 are divided into 1815 images for training, 259 for validation and 520 for testing.	2021	96.30	90.40	97.90	-	96.30
MCGU-Net [59]		2020	95.50	84.80	98.60	-	95.50
SICU-Net [XYZ]	2594 dermoscopic images were divided into (80%) 2076 training images and (20%) 518 images for testing.		94.69	89.39	96.95	90.94	83.39
DICU-Net [Gray-HSV]		2022	94.14	93.02	94.61	90.44	82.55
TICU-Net [XYZ-Gray-HSV]			94.58	91.57	95.85	90.96	83.42

SICU-Net performed the best in TNR with 98.55 % score for SK cases, and the proposed DICU-Net performed the best in TPR with 90.69 % score for Benign class. While, the proposed TICU-Net achieved the best in TNR for benign cases and the best for TNR for melanoma cases with values 98.18 and 97.58, respectively.

Tables 25 and 26 present the performance of our proposed models with other fully convolutional networks: FCN, U-Net, SegNet, and FrCN methods on the PH2 dataset that is used for testing only at the level of dataset categories. Tables show that the proposed SICU-Net performed the best in ACC with 95.32 % score, and the proposed DICU-Net performed the best in TPR with 99.68 % score, all for Benign class.

The results of some of the state-of-art methods such [54], [59], [60] utilize different partitioning for the benchmark databases provided by the creator as shown in the second column table 27 and 28. The differences in data partitioning make somehow unfair comparison between these methods and our proposed method. Although our proposed method could not achieve the best performance compared with other recent works, the idea of using multi-input paths with different color spaces can be added to any of the existing models to improve the robustness of these models due to color contrast variations.

V. DISCUSSION

The proposed color U-Net models (CU-Net) using single, dual, and triple color inputs are designed to overcome the problem of color variations in skin lesion images. Most of the existing deep network architectures utilize only a single input path, while this work explores the effectiveness of combining multiple inputs with different color spaces.

The proposed CU-Net variants are fed with different color spaces of the input image to exploit distinct features that appear in some color channels in specific color spaces. The proposed models help to significantly improve the performance of U-Net semantic segmentation deep model. The interconnections between encoder and decoder paths using attention network enrich the features and hence improve the classification results. Since each channel of the extracted feature maps is working as a feature detector, we take the advantage of the inter-relationship of the channel features to focus on the meaning of an input image using the channel attention module. However, one drawback of the proposed method is that the complexity of the network increases as we add more inputs and their corresponding encoder paths. Although our proposed method could not achieve the best performance compared with other recent works, the idea of using multi-input paths with different color spaces can be added to any of the state-of-the-art models to improve the robustness of these models against color contrast variations.

The experiments conducted using three different benchmark datasets reveal some interested remarks, which can be summarized as follows:

- No specific color space is the best among others. However, the combination of different color spaces improves the results of various performance metrics.
- Using channel attention modules in the proposed CU-Net models in the interconnection between the encoder and decoder paths improves the results of some evaluation metrics.
- Using image normalization improves the results of TPR metric for most of all color-space combinations which

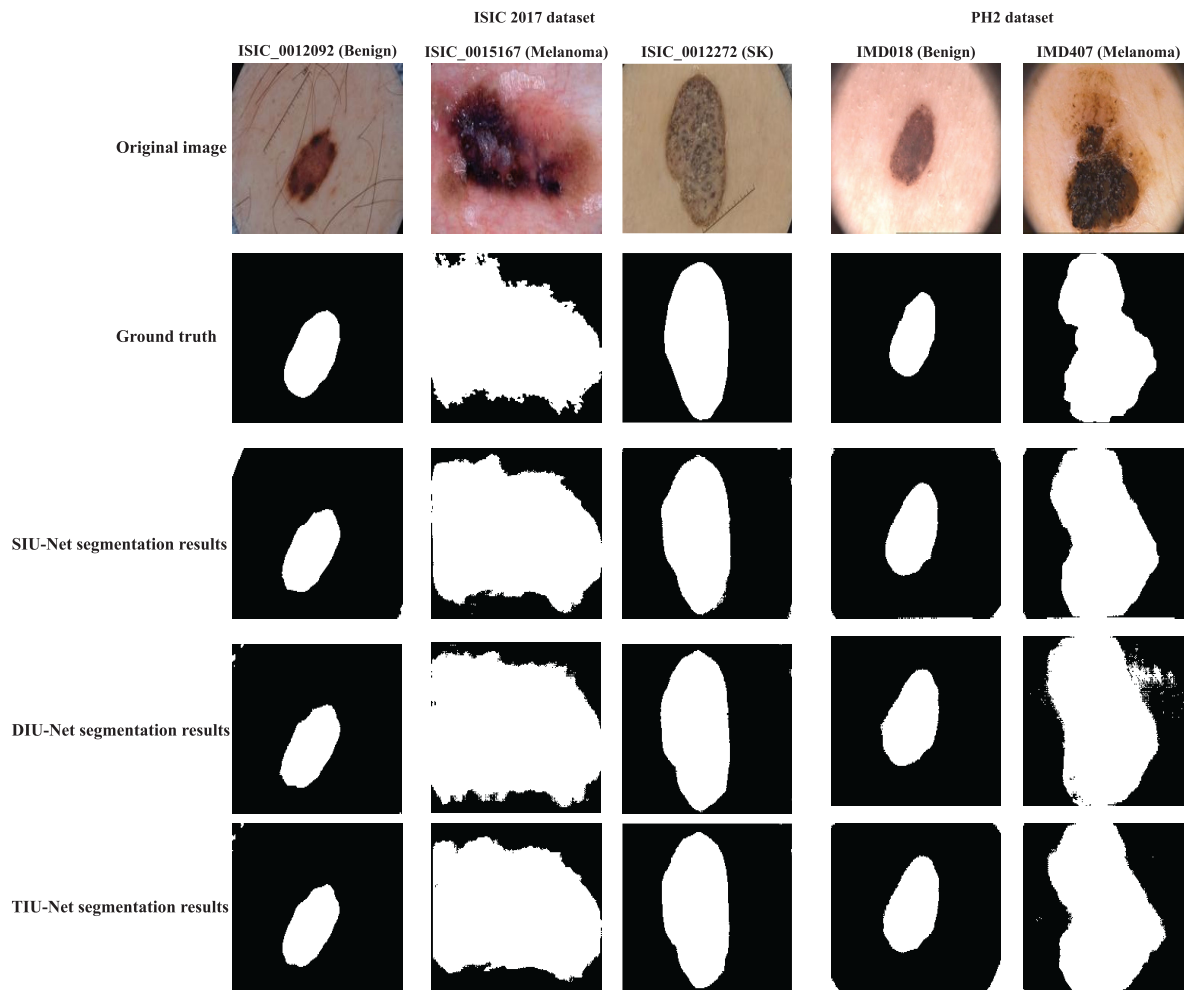


FIGURE 11. SIU-Net, DICU-Net, and U-Net segmentation results for chosen images from each dataset class (one image from each dataset class).

is considered the most important metric between all metrics.

- The proposed composite loss function outperforms each individual loss function.
- The proposed TICU-Net achieves the best values for Dice and jaccard coefficients compared with SICU-Net and DICU-Net.

VI. CONCLUSION AND FUTURE WORK

Color contrast variations of dermoscopy images posed a significant obstacle for an accurate diagnosis of the infected regions from the healthy ones. To address this problem, this paper presented three new convolution neural network models for skin lesion segmentation. Unlike other existing deep learning models, the proposed CU-Net models solve the low contrast and color discrepancy skin lesion segmentation problems. The three proposed models including single, dual, and triple input color U-Net (SICU-Net, DICU-Net, TICU-Net) are fed by multiple input images with different color spaces. The combination of multiple color spaces not

only improves the segmentation results but also increases the robustness of the model to skin lesion color variations. The optimal selection of color space is significantly affect the performance of the proposed color U-Net models. Deep convolutional network with multiple input paths enriches the extracted features and improve the performance of its counter single input network. Moreover, channel-wise attention modules are used to focus on the interested features extracted from each input path. We also utilised the pre-processing by using (1) color space transformation to choose the optimum color space, and (2) applying a simple image normalization to enhance image contrast. The performance of triple input network is better than both single and dual inputs, but on the condition that appropriate color spaces are selected. The three well-known ISIC 2018, ISIC 2017 and PH2 datasets are used to prove the robustness of the proposed models. Experimental results reveal that the three proposed CU-Net models significantly outperform the performance of the traditional U-Net model. The experimental results conducted on ISIC 2018 are comparable with other state-of-the-art

methods. Further improvements can be made by applying other attention network architectures to emphasize more features in the image.

ACKNOWLEDGMENT

Saleh Aly would like to thank the Deanship of Scientific Research at Majmaah University for supporting this work under Project No. R-2022-27.

REFERENCES

- [1] M. Shorfuzzaman, "An explainable stacked ensemble of deep learning models for improved melanoma skin cancer detection," *Multimedia Syst.*, pp. 1–15, Apr. 2021, doi: 10.1007/s00530-021-00787-5.
- [2] A. Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: A survey of state-of-the-art," *Artif. Intell. Rev.*, vol. 54, pp. 811–841, Jun. 2020.
- [3] R. Ashraf, S. Afzal, A. U. Rehman, S. Gul, J. Baber, M. Bakhtyar, I. Mehmood, O.-Y. Song, and M. Maqsood, "Region-of-interest based transfer learning assisted framework for skin cancer detection," *IEEE Access*, vol. 8, pp. 147858–147871, 2020.
- [4] A. A. Adegun and S. Viriri, "Deep learning-based system for automatic melanoma detection," *IEEE Access*, vol. 8, pp. 7160–7172, 2020.
- [5] L. Ichim and D. Popescu, "Melanoma detection using an objective system based on multiple connected neural networks," *IEEE Access*, vol. 8, pp. 179189–179202, 2020.
- [6] T. J. Brinker, A. Hekler, A. H. Enk, C. Berking, S. Haferkamp, A. Hauschild, M. Weichenthal, J. Klode, D. Schadendorf, T. Holland-Letz, C. von Kalle, S. Fröhling, B. Schilling, and J. S. Utikal, "Deep neural networks are superior to dermatologists in melanoma image classification," *Eur. J. Cancer*, vol. 119, pp. 11–17, Sep. 2019.
- [7] N. Razmjoo, B. Somayeh Mousavi, F. Soleymani, and M. H. Khotbesara, "A computer-aided diagnosis system for malignant melanomas," *Neural Comput. Appl.*, vol. 23, nos. 7–8, pp. 2059–2071, Dec. 2013.
- [8] F. Xie, J. Yang, J. Liu, Z. Jiang, Y. Zheng, and Y. Wang, "Skin lesion segmentation using high-resolution convolutional neural network," *Comput. Methods Programs Biomed.*, vol. 186, Apr. 2020, Art. no. 105241.
- [9] R. Arora, B. Raman, K. Nayyar, and R. Awasthi, "Automated skin lesion segmentation using attention-based deep convolutional neural network," *Biomed. Signal Process. Control*, vol. 65, Mar. 2021, Art. no. 102358.
- [10] M. K. Hasan, L. Dahal, P. N. Samarakoon, F. I. Tushar, and R. Martí, "DSNet: Automatic dermoscopic skin lesion segmentation," *Comput. Biol. Med.*, vol. 120, May 2020, Art. no. 103738.
- [11] M. Pezhman Pour and H. Seker, "Transform domain representation-driven convolutional neural networks for skin lesion segmentation," *Expert Syst. Appl.*, vol. 144, Apr. 2020, Art. no. 113129.
- [12] G. G. De Angelo, A. G. C. Pacheco, and R. A. Krohling, "Skin lesion segmentation using deep learning for images acquired from smartphones," in *Proc. Int. Joint Conf. Neural Netw. (IJCNN)*, Jul. 2019, pp. 1–8.
- [13] H. Nisar, Y. K. Ch'ng, T. Y. Chew, V. V. Yap, K. H. Yeap, and J. J. Tang, "A color space study for skin lesion segmentation," in *Proc. IEEE Int. Conf. Circuits Syst. (ICCS)*, Sep. 2013, pp. 172–176.
- [14] A. A. Abbas, X. Guo, W. H. Tan, and H. A. Jalab, "Combined spline and B-spline for an improved automatic skin lesion segmentation in dermoscopic images using optimal color channel," *J. Med. Syst.*, vol. 38, no. 8, p. 80, 2014.
- [15] P. Ganesan, B. S. Sathish, and L. M. Leo Joseph, "HSL color space based skin lesion segmentation using fuzzy-based techniques," in *Advances in Electrical Control and Signal Systems* (Lecture Notes in Electrical Engineering), vol. 665, Singapore: Springer, 2020, pp. 903–911.
- [16] M. Hajabdollahi, R. Esfandiarpour, P. Khadivi, S. M. R. Soroushmehr, N. Karimi, and S. Samavi, "Simplification of neural networks for skin lesion image segmentation using color channel pruning," *Computerized Med. Imag. Graph.*, vol. 82, Jun. 2020, Art. no. 101729.
- [17] Y. Yuan, M. Chao, and Y.-C. Lo, "Automatic skin lesion segmentation using deep fully convolutional networks with jaccard distance," *IEEE Trans. Med. Imag.*, vol. 36, no. 9, pp. 1876–1886, Sep. 2017.
- [18] 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.
- [19] G. Schaefer, M. I. Rajab, M. Emre Celebi, and H. Iyatomi, "Colour and contrast enhancement for improved skin lesion segmentation," *Computerized Med. Imag. Graph.*, vol. 35, no. 2, pp. 99–104, Mar. 2011.
- [20] M. H. Jafari, N. Karimi, E. Nasr-Esfahani, S. Samavi, S. M. R. Soroushmehr, K. Ward, and K. Najarian, "Skin lesion segmentation in clinical images using deep learning," in *Proc. 23rd Int. Conf. Pattern Recognit. (ICPR)*, Dec. 2016, pp. 337–342.
- [21] H. M. Ünver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of YOLO and GrabCut algorithm," *Diagnostics*, vol. 9, no. 3, p. 72, Jul. 2019.
- [22] P. Shan, Y. Wang, C. Fu, W. Song, and J. Chen, "Automatic skin lesion segmentation based on FC-DPN," *Comput. Biol. Med.*, vol. 123, Aug. 2020, Art. no. 103762.
- [23] P. Tang, X. Yan, Q. Liang, and D. Zhang, "AFLN-DGCL: Adaptive feature learning network with difficulty-guided curriculum learning for skin lesion segmentation," *Appl. Soft Comput.*, vol. 110, Oct. 2021, Art. no. 107656.
- [24] L. Zhang, G. Yang, and X. Ye, "Automatic skin lesion segmentation by coupling deep fully convolutional networks and shallow network with textons," *J. Med. Imag.*, vol. 6, no. 2, 2019, Art. no. 024001.
- [25] X. Wang, X. Jiang, H. Ding, Y. Zhao, and J. Liu, "Knowledge-aware deep framework for collaborative skin lesion segmentation and melanoma recognition," *Pattern Recognit.*, vol. 120, Dec. 2021, Art. no. 108075.
- [26] L. Bi, J. Kim, E. Ahn, A. Kumar, D. Feng, and M. Fulham, "Step-wise integration of deep class-specific learning for dermoscopic image segmentation," *Pattern Recognit.*, vol. 85, pp. 78–89, Jan. 2019.
- [27] M. M. K. Sarker, H. A. Rashwan, F. Akram, V. K. Singh, S. F. Banu, F. U. H. Chowdhury, K. A. Choudhury, S. Chambon, P. Radeva, D. Puig, and M. Abdel-Nasser, "SLSNet: Skin lesion segmentation using a lightweight generative adversarial network," *Expert Syst. Appl.*, vol. 183, Nov. 2021, Art. no. 115433.
- [28] W. Li, A. N. Joseph Raj, T. Tjahjadi, and Z. Zhuang, "Digital hair removal by deep learning for skin lesion segmentation," *Pattern Recognit.*, vol. 117, Sep. 2021, Art. no. 107994.
- [29] X. Tong, J. Wei, B. Sun, S. Su, Z. Zuo, and P. Wu, "ASCU-net: Attention gate, spatial and channel attention U-net for skin lesion segmentation," *Diagnostics*, vol. 11, no. 3, p. 501, Mar. 2021.
- [30] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 234–241.
- [31] H. Iyatomi, H. Oka, M. E. Celebi, M. Hashimoto, M. Hagiwara, M. Tanaka, and K. Ogawa, "An improved internet-based melanoma screening system with dermatologist-like tumor area extraction algorithm," *Computerized Med. Imag. Graph.*, vol. 32, no. 7, pp. 566–579, Oct. 2008.
- [32] 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.
- [33] G. Betta, G. Di Leo, G. Fabbrocini, A. Paolillo, and M. Scalvenzi, "Automated application of the '7-point checklist' diagnosis method for skin lesions: Estimation of chromatic and shape parameters," in *Proc. IEEE Instrum. Meas. Technol. Conf.*, vol. 3, May 2005, pp. 1818–1822.
- [34] N. Abraham and N. M. Khan, "A novel focal Tversky loss function with improved attention U-net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2019, pp. 683–687.
- [35] K. Taouil and N. Romdhane, "Automatic segmentation and classification of skin lesion images," in *Proc. 2nd Int. Conf. Distrib. Frameworks Multimedia Appl.*, May 2006, pp. 211–222.
- [36] M. E. Celebi, Y. A. Aslandogan, and P. R. Bergstresser, "Mining biomedical images with density-based clustering," in *Proc. Int. Conf. Inf. Technol.: Coding Comput. (ITCC)*, vol. 1, Apr. 2005, pp. 163–168.
- [37] M. E. Celebi, Y. A. Aslandogan, W. V. Stoecker, H. Iyatomi, H. Oka, and X. Chen, "Unsupervised border detection in dermoscopy images," *Skin Res. Technol.*, vol. 13, no. 4, pp. 454–462, 2007.
- [38] 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.

- [39] A. S. Ashour, Y. Guo, E. Kucukkulahli, P. Erdogmus, and K. Polat, "A hybrid dermoscopy images segmentation approach based on neutro-sophic clustering and histogram estimation," *Appl. Soft Comput.*, vol. 69, pp. 426–434, Aug. 2018.
- [40] H. Park, T. Schoepflin, and Y. Kim, "Active contour model with gradient directional information: Directional snake," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 11, no. 2, pp. 252–256, Feb. 2001.
- [41] X. Yuan, N. Situ, and G. Zouridakis, "A narrow band graph partitioning method for skin lesion segmentation," *Pattern Recognit.*, vol. 42, pp. 1017–1028, Jun. 2009.
- [42] Q. Abbas, M. E. Celebi, I. F. Garcia, and M. Rashid, "Lesion border detection in dermoscopy images using dynamic programming," *Skin Res. Technol.*, vol. 17, no. 1, pp. 91–100, Feb. 2011.
- [43] Ş. Öztürk and U. Özkaya, "Skin lesion segmentation with improved convolutional neural network," *J. Digit. Imag.*, vol. 33, no. 4, pp. 958–970, Aug. 2020.
- [44] M. A. Khan, T. Akram, Y.-D. Zhang, and M. Sharif, "Attributes based skin lesion detection and recognition: A mask RCNN and transfer learning-based deep learning framework," *Pattern Recognit. Lett.*, vol. 143, pp. 58–66, Mar. 2021. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0167865520304463>
- [45] S. Kadry, D. Taniar, R. Damasevicius, V. Rajinikanth, and I. A. Lawal, "Extraction of abnormal skin lesion from dermoscopy image using VGG-SegNet," in *Proc. 7th Int. Conf. Bio Signals, Images, Instrum. (ICBSII)*, Mar. 2021, pp. 1–5. [Online]. Available: <https://ieeexplore.ieee.org/document/9445180/>
- [46] S. R. Hashemi, S. S. M. Salehi, D. Erdogmus, S. P. Prabhu, S. K. Warfield, and A. Gholipour, "Asymmetric loss functions and deep densely-connected networks for highly-imbalanced medical image segmentation: Application to multiple sclerosis lesion detection," *IEEE Access*, vol. 7, pp. 1721–1735, 2018.
- [47] D. Połap, A. Winnicka, K. Serwata, K. Kęsik, and M. Woźniak, "An intelligent system for monitoring skin diseases," *Sensors*, vol. 18, no. 8, p. 2552, Aug. 2018. [Online]. Available: <http://www.mdpi.com/1424-8220/18/8/2552>
- [48] M. A. Khan, M. Sharif, T. Akram, R. Damaševičius, and R. Maskeliūnas, "Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization," *Diagnostics*, vol. 11, no. 5, p. 811, Apr. 2021. [Online]. Available: <https://www.mdpi.com/2075-4418/11/5/811>
- [49] S. Qamar, P. Ahmad, and L. Shen, "Dense encoder-decoder-based architecture for skin lesion segmentation," *Cogn. Comput.*, vol. 13, pp. 583–594, Feb. 2021.
- [50] H. N. Pham, C. Y. Koay, T. Chakraborty, S. Gupta, B. L. Tan, H. Wu, A. Vardhan, Q. H. Nguyen, N. R. Palaparthi, B. P. Nguyen, and M. C. H. Chua, "Lesion segmentation and automated melanoma detection using deep convolutional neural networks and XGBoost," in *Proc. Int. Conf. Syst. Sci. Eng. (ICSSE)*, Jul. 2019, pp. 142–147.
- [51] P. Tang, Q. Liang, X. Yan, S. Xiang, W. Sun, D. Zhang, and G. Coppola, "Efficient skin lesion segmentation using separable-unet with stochastic weight averaging," *Comput. Methods Programs Biomed.*, vol. 178, pp. 289–301, Sep. 2019.
- [52] M. Dash, N. D. Londhe, S. Ghosh, A. Semwal, and R. S. Sonawane, "PsLSNet: Automated psoriasis skin lesion segmentation using modified U-Net-based fully convolutional network," *Biomed. Signal Process. Control*, vol. 52, pp. 226–237, Jul. 2019.
- [53] B. S. Lin, K. Michael, S. Kalra, and H. R. Tizhoosh, "Skin lesion segmentation: U-Nets versus clustering," in *Proc. IEEE Symp. Ser. Comput. Intell. (SSCI)*, Nov. 2017, pp. 1–7.
- [54] R. Azad, I. 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. Workshops (ICCVW)*, Oct. 2021, pp. 3274–3283.
- [55] G. Huang, Z. Liu, L. Van Der Maaten, and K. Q. Weinberger, "Densely connected convolutional networks," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 2261–2269. [Online]. Available: <https://ieeexplore.ieee.org/document/8099726/>
- [56] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jun. 2016, pp. 770–778.
- [57] C. Szegedy, S. Ioffe, V. Vanhoucke, and A. A. Alemi, "Inception-V4, inception-ResNet and the impact of residual connections on learning," in *Proc. 31st AAAI Conf. Artif. Intell.*, 2017, pp. 4278–4284.
- [58] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," in *Proc. 3rd Int. Conf. Learn. Represent. (ICLR)*, 2015, pp. 1–14.
- [59] M. Z. Alom, C. Yakopcic, M. Hasan, T. M. Taha, and V. K. Asari, "Recurrent residual U-Net for medical image segmentation," *J. Med. Imag.*, vol. 6, no. 1, 2019, Art. no. 014006.
- [60] 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*.
- [61] Y. Yuan and Y.-C. Lo, "Improving dermoscopic image segmentation with enhanced convolutional-deconvolutional networks," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 519–526, Mar. 2019.
- [62] S. E. Reza Azad, M. Asadi-Aghbolaghi, and M. Fathy, "Bi-directional ConvLSTM U-Net with densely connected convolutions," in *Proc. IEEE/CVF Int. Conf. Comput. Vis.*, Oct. 2019, pp. 406–415.
- [63] 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.
- [64] L. Chen, H. Zhang, J. Xiao, L. Nie, J. Shao, W. Liu, and T.-S. Chua, "SCA-CNN: Spatial and channel-wise attention in convolutional networks for image captioning," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, Jul. 2017, pp. 6298–6306.
- [65] F. Milletari, N. Navab, and S.-A. Ahmadi, "V-Net: Fully convolutional neural networks for volumetric medical image segmentation," in *Proc. 14th Int. Conf. 3D Vis. (3DV)*, Oct. 2016, pp. 1–11.
- [66] T. Brosch, Y. Yoo, L. Y. Tang, D. K. Li, A. Traboulsée, and R. Tam, "Deep convolutional encoder networks for multiple sclerosis lesion segmentation," in *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*, vol. 9351. Cham, Switzerland: Springer, 2015, pp. 3–11, doi: [10.1007/978-3-319-24574-4_1](https://doi.org/10.1007/978-3-319-24574-4_1).
- [67] L.-C. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "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, Apr. 2017.
- [68] Z. Mirikharaji and G. Hamarneh, "Star shape prior in fully convolutional networks for skin lesion segmentation," *Lecture Notes in Computer Science (Including Subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* (Lecture Notes in Computer Science), vol. 11073. 2018, pp. 737–745.
- [69] Y. Xue, T. Xu, and X. Huang, "Adversarial learning with multi-scale loss for skin lesion segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, 2018, pp. 859–863, doi: [10.1109/ISBI.2018.8363707](https://doi.org/10.1109/ISBI.2018.8363707).
- [70] Y. Li and L. Shen, "Skin lesion analysis towards melanoma detection using deep learning network," *Sensors*, vol. 18, no. 2, p. 556, 2018.
- [71] 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, doi: [10.1016/j.media.2019.01.012](https://doi.org/10.1016/j.media.2019.01.012).
- [72] G. Zhang, X. Shen, S. Chen, Y. E. Luo, J. I. E. Yu, and J. Lu, "DSM: A deep supervised multi-scale network learning for skin cancer segmentation," *IEEE Access*, vol. 7, pp. 140936–140945, 2019.
- [73] D. Ma, H. Wu, J. Sun, C. Yu, and L. Liu, "A light-weight context-aware self-attention model for skin lesion segmentation," in *Proc. Pacific Rim Int. Conf. Artif. Intell.*, 2019, pp. 501–505.
- [74] D. K. Nguyen, T.-T. Tran, C. P. Nguyen, and V.-T. Pham, "Skin lesion segmentation based on integrating efficientnet and residual block into U-Net neural network," in *Proc. 5th Int. Conf. Green Technol. Sustain. Develop. (GTSD)*, Nov. 2020, pp. 366–371.
- [75] Y. Xie, G. S. Member, J. Zhang, and H. Lu, "SESV: Accurate medical image segmentation by predicting and correcting errors," *IEEE Trans. Med. Imag.*, vol. 40, no. 1, pp. 286–296, Jan. 2021.
- [76] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, Y. Xu, J. Qin, S. Chen, T. Wang, and S. Wang, "Skin lesion segmentation via generative adversarial networks with dual discriminators," *Med. Image Anal.*, vol. 64, Aug. 2020, Art. no. 101716.
- [77] 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, doi: [10.1016/j.media.2021.102327](https://doi.org/10.1016/j.media.2021.102327).

- [78] M. K. Sarker, H. A. Rashwan, and F. Akram, "SLSDeep: Skin lesion segmentation based on dilated residual and pyramid pooling networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.*, 2018, pp. 1–9.
- [79] 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, Jan. 2017.
- [80] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. D. Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An advanced architecture for medical image segmentation," in *Proc. IEEE Int. Symp. Multimedia (ISM)*, Dec. 2019, pp. 225–230.
- [81] Q. Jin, H. Cui, C. Sun, Z. Meng, and R. Su, "Cascade knowledge diffusion network for skin lesion diagnosis and segmentation," *Appl. Soft Comput.*, vol. 99, Feb. 2021, Art. no. 106881, doi: [10.1016/j.asoc.2020.106881](https://doi.org/10.1016/j.asoc.2020.106881).
- [82] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, Y. Ma, D. Xiang, W. Zhu, and X. Chen, "CPFNet: Context pyramid fusion network for medical image segmentation," *IEEE Trans. Med. Imag.*, vol. 39, no. 10, pp. 3008–3018, Oct. 2020.
- [83] M. Goyal, A. Oakley, P. Bansal, D. Dancey, and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2020.
- [84] 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.

RANIA RAMADAN received the B.Sc. and M.Sc. degrees in computer science from Sohag University, Sohag, Egypt, in 2000 and 2009, respectively. She is currently pursuing the Ph.D. degree in medical image segmentation using deep learning. Her research interests include image segmentation, medical image analysis, machine learning, deep learning, image processing, computer vision, and neural networks.



SALEH ALY (Senior Member, IEEE) received the B.Sc. and M.Sc. degrees in electrical and computer engineering from Assiut University, Assiut, Egypt, in 1997 and 2004, respectively, and the Ph.D. degree from the Department of Intelligent Systems, Graduate School of Information Science and Electrical Engineering, Kyushu University, Japan, in 2010. He is currently a Professor in computer engineering with the Faculty of Engineering, Aswan University, Egypt. In 2012 and 2014, he was a Visiting Researcher with the Department of Media Science, Nagoya University, and the Department of Computer Science, Tsukuba University, Japan, respectively. His research interests include deep learning, neural networks, pattern recognition, medical image processing, machine learning, and computer vision.

• • •