

Received August 14, 2019, accepted August 29, 2019, date of publication September 11, 2019, date of current version September 24, 2019.

Digital Object Identifier 10.1109/ACCESS.2019.2940418

FCA-Net: Adversarial Learning for Skin Lesion Segmentation Based on Multi-Scale Features and Factorized Channel Attention

VIVEK KUMAR SINGH¹, MOHAMED ABDEL-NASSER^{1,4}, HATEM A. RASHWAN¹, FARHAN AKRAM^{3,5}, NIDHI PANDEY¹, ALAIN LALANDE², BENOIT PRESLES², SANTIAGO ROMANI¹, AND DOMENEC PUIG¹

¹Department of Computer Engineering and Mathematics, Rovira i Virgili University, 43007 Tarragona, Spain

²ImViA EA 7535, University of Burgundy, 21078 Dijon, France

³Imaging Informatics Division, Bioinformatics Institute, A*STAR, Singapore 138671

⁴Department of Electrical Engineering, Aswan University, Aswan 81542, Egypt

⁵Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, Abu Dhabi 127788, United Arab Emirates

Corresponding author: Vivek Kumar Singh (vivekkumar.singh@urv.cat)

This work was supported by the Spanish Government through Project under Grant DPI2016-77415-R.

ABSTRACT Skin lesion segmentation in dermoscopic images is still a challenge due to the low contrast and fuzzy boundaries of lesions. Moreover, lesions have high similarity with the healthy regions in terms of appearance. In this paper, we propose an accurate skin lesion segmentation model based on a modified conditional generative adversarial network (cGAN). We introduce a new block in the encoder of cGAN called factorized channel attention (FCA), which exploits both channel attention mechanism and residual 1-D kernel factorized convolution. The channel attention mechanism increases the discriminability between the lesion and non-lesion features by taking feature channel interdependencies into account. The 1-D factorized kernel block provides extra convolutions layers with a minimum number of parameters to reduce the computations of the higher-order convolutions. Besides, we use a multi-scale input strategy to encourage the development of filters which are scale-variant (i.e., constructing a scale-invariant representation). The proposed model is assessed on three skin challenge datasets: ISBI2016, ISBI2017, and ISIC2018. It yields competitive results when compared to several state-of-the-art methods in terms of Dice coefficient and intersection over union (IoU) score. The codes of the proposed model are publicly available at <https://github.com/vivek231/Skin-Project>.

INDEX TERMS Skin lesion, conditional generative adversarial network, channel attention, factorized kernel, residual convolution.

I. INTRODUCTION

According to the world health organization (WHO), around 100,000 melanoma skin cancer cases appear every year [1]. For early diagnosis, different diagnostic algorithms [2], such as the ABCD Dermoscopy Rule, and 7-Point Check List, have been utilized. For example, the ABCD rule of dermoscopy helps dermatologist to discriminate between benign and malignant tumors by analyzing the following features in skin images: asymmetry (A), border irregularity (B), color (C), and dermoscopic structures (D).

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

Nowadays, computer-aided diagnosis (CAD) systems are widely used for the early-stage diagnosis of skin diseases using dermoscopic images. These CAD systems are used to train inexperienced dermatologists and to devise automated diagnostic procedures. One of the important tasks of these CAD systems is to accurately segment the lesions from the dermoscopic images that help follow-up lesions. Moreover, CAD systems are also designed to extract basic features (e.g., ABCD features) that can be used for in-depth pattern, shape or region of interest analysis.

Figure 1 presents four examples of skin images containing lesions. As shown in Figure 1, there are many challenges for skin lesion segmentation methods to properly segregate the observed lesions, such as the presence of hair, illumination

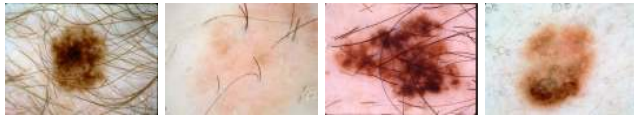


FIGURE 1. Examples of skin lesions with presence of hair, illumination changes, noise, color variations and fuzzy boundaries.

changes, noise, color variations, and fuzzy boundaries [3]. These challenges degrade the performance of the automatic segmentation methods.

Several skin lesion segmentation methods have been proposed in the literature [4]–[6], and [7], which are based on traditional computer vision, machine learning and/or deep learning techniques. Regarding traditional computer vision techniques, adaptive thresholding, region growing, and contour-based methods have been used to segment skin lesions [8]. However, these methods yield poor results on low contrast skin images.

Recently, different deep learning techniques have been used to segment biomedical images [9]–[12]. Several approaches [7], [13] have been proposed to automatically segment skin lesions. These methods yield more accurate results than traditional ones. However, most methods apply several pre-processing (e.g., hair removal, color space transformation, and data augmentation) or post-processing techniques (e.g. morphological operations) to improve their results. Among of used deep models, conditional generative adversarial network (cGAN), a cutting-edge idea in image-to-image translation, has been used [14] to segment skin lesions. It gave a Dice coefficient of 86.7%.

Enhancing the contextual information extracted by some layers of the cGAN model could increase the segmentation accuracy. In order to enhance the contextual information, a dual attention block was proposed that integrates the spatial and channel long-range dependencies for general scene segmentation [15]. However, the dual attention block significantly increases the number of training parameters, especially in the spatial attention branch. Therefore, we substituted that branch with a residual connection joined with four layers of 1-D factorized kernel convolution. The factorization method proposed in [16] further helps in reducing the trainable parameters of the equivalent two layers of 2D kernel convolution. Consequently, for our skin lesion segmentation method, we introduce a novel layer, called Factorized Channel Attention (FCA), which integrates channel attention and residual 1-D factorized kernel convolution. On one hand, we assume that a high cross-channel correlation in activation maps of the encoder layers indicates the presence of relevant cues for distinguishing skin lesion pixels from normal skin pixels. On the other hand, we also assume that the residual convolutions can learn some spatial dependencies in neighboring positions of the feature maps, which lead to a more compact pixel labeling of the lesion regions, but with a minimal set of trained parameters. Consequently, by integrating both methods in [15] and [16] we are able to formulate an efficient and accurate segmentation network.

The main contributions of this paper are outlined below:

- 1) We propose a fully automated skin lesion segmentation model based on cGAN, which can learn more effective features for small-size skin lesions without using pre-processing (e.g., color space transformation) or data augmentation techniques.
- 2) To model channel and spatial inter-dependencies inside the feature maps of the encoder layers, we introduce the FCA block, which integrates channel attention and residual 1-D factorized convolutions to boost feature discriminability between the lesion and non-lesion pixels.
- 3) We also use a multi-scale input strategy, in which the input images are resized into three different scales of the original size. Thus, the FCA-Net can explicitly deal with variation in resolution, object size and image scale, by encouraging the development of filters which are scale-variant, while constructing a scale-invariant representation.

Section II of this paper presents the related works on skin lesion segmentation. Section III explains the proposed model. The results and discussion of the proposed model are provided in Section IV. Lastly, Section V summarizes our study and provides some insights into future work.

II. RELATED WORK

In the literature, various skin lesion segmentation methods have been proposed. Table 1 summarizes some of these methods that have been recently published. These methods include traditional computer vision techniques, convolutional neural network (CNN) and generative adversarial network (GAN) based methods.

A. TRADITIONAL COMPUTER VISION METHODS

These methods usually exploit pixel values, color, texture and shape statistics, i.e., hand-crafted features used for the segmentation process. For instance, Rahman *et al.* [8] proposed an automatic lesion segmentation using adaptive thresholding and region growing methods to segment skin lesions. These areas were then fed them into an extreme learning machine (ELM) to classify skin lesions. The main drawback of thresholding based methods is that they can achieve good results only if there is a high contrast between the lesion area and the surrounding skin region, which is not always the case. Also, Wong *et al.* [29] suggested an iterative stochastic region-merging approach, which was employed to segment skin lesions from macroscopic images. In this method, stochastic region merging was initialized on a pixel level, and then on a regional level until convergence. An active contour method (snakes) based on gradient vector flow (GVF) was proposed in [30] for lesion contour extraction. An extension of GVF based on a mean shift method was proposed in [31]. In [32], two contour-based methods were applied to skin images: adaptive snake and active contour. However, contour-based methods usually fail in the presence of hair or

TABLE 1. Summary of skin lesion segmentation methods. Dashes (–) indicate that the information is not reported in the referred references.

Study	Dataset	Architecture	Data augmentation/Preprocessing	Input Size	Loss function	Post-processing
Yu <i>et al.</i> [17]	ISBI2016	FCRN	Rotate, flipping, translation, and random noise	Resize 250×250	-	Thresholding of 0.5
Bissoto <i>et al.</i> [18]	ISIC2018	U-Net	Rotate, illumination, scale, and flipping	Resize 256×256	Binary cross entropy+IoU	Thresholding, hole filling
Bisla <i>et al.</i> [19]	ISBI2017, ISIC2018	U-Net	Random masking	Resize 380×380	Binary cross entropy	Hole filling
Almasani <i>et al.</i> [5]	ISBI2017	FrCN	Rotate and colorspace HSV	Resize 192×256	Cross entropy	-
Mirkharaji <i>et al.</i> [20]	ISBI2016	U-Net	Rotation and flipping	Resize 336×336	Dice	-
Xiaomeng [21]	ISBI2017	U-Net	Rotation, flipping, and scaling	-	Cross entropy+MSE	Thresholding of 0.5, hole filling
Rahman <i>et al.</i> [8]	ISBI2016	Adaptive thresholding and region growing	-	-	-	-
Yuan <i>et al.</i> [17]	ISBI2016, ISBI2017	CDNN	Rotate, flipping, shifting, and scaling	Resize 192×192	IoU	Thresholding of 0.5, hole filling
Berseth <i>et al.</i> [7]	ISBI2017	U-Net	Rotation, flipping, and zooming	Resize 192×192	-	CRF
Venkatesh <i>et al.</i> [22]	ISBI2017	U-Net	Rotation, flipping, translation, and scaling	Resize 256×256	Binary cross entropy+IoU	-
Galdran <i>et al.</i> [23]	ISBI2017	U-Net	Rotation, flipping, illumination, scaling, translation, and change RGB to Gray scale	-	IoU	-
Sulaiman <i>et al.</i> [24]	ISBI2017	U-Net	Rotation, flipping, color shifting, translation, and scaling	Resize 512×512	Dice	-
Sarker <i>et al.</i> [25]	ISBI2017	U-Net	-	Resize 384×384	NLL + EPE	-
Jahanifar <i>et al.</i> [26]	ISBI2016, ISBI2017	Saliency detection	-	Resize 300×400	-	Thresholding
Lei Bi <i>et al.</i> [27]	ISBI2016, ISBI2017	FCN	Random cropping and flipping	Resize 1000×1000	Cross entropy	Thresholding, hole filling
Xue <i>et al.</i> [14]	ISBI2017	GAN	Random cropping, flipping, and illumination change	Resize 128×128	L1-norm	-
Izadi <i>et al.</i> [28]	DermoFit	GAN	Random cropping, flipping, and elastic deformation	-	Binary cross entropy	-

air bubbles and if the transition between the lesion and the surrounding skin is smooth.

B. CNN-BASED METHODS

Nowadays, deep convolutional neural networks are widely used to analyze natural and medical images for tasks of detection, segmentation, and classification. Several deep learning-based image segmentation methods have been proposed in the last five years. One of the prominent models is the fully convolutional network (FCN) [33] that includes encoder and decoder layers. In [34], the U-Net model was proposed for biomedical image segmentation. It adapted the FCN model by using a skip connection from each encoder layer to the corresponding decoder layer to keep the features extracted from the first layers. Furthermore, the SegNet model was proposed in [35] to improve the accuracy of image segmentation by using a max pooling in the decoder layers that extend from the corresponding encoder layer to achieve a non-linear upsampling of their input feature maps.

Recently, researchers have used state-of-the-art image segmentation based deep learning models to obtain more accurate skin lesion segmentation. In [17], a CNN-based fully convolutional residual network (FCRN) and multiscale contextual information were proposed to segment skin lesions. However, this method is not able to properly segment low contrast dermoscopic images and that includes hairs and irregular skin lesion shapes. Also, this method cannot fully utilize the discrimination capability of the deep CNN with limited training data, according to the authors of the paper. Furthermore, Bissoto *et al.* [18] used the U-Net network to segment skin lesions. They assessed their model on the ISIC2018 skin lesion dataset and achieved an IoU score of 72.8%. The main limitation of this method is that it requires several pre-processing steps for removing noise present in dermoscopic images.

To accurately segment the lesion boundaries, Vesal *et al.* [24] proposed the SkinNet model, which is based on the U-Net architecture. The authors replaced standard

convolution layers at every level of both the encoder and decoder with densely connected convolution layers. The SkinNet model was evaluated on ISBI2017 dataset and achieved a Dice coefficient of 85.10% and an IoU score of 76.7%. To extract rich features from a dermoscopic image, Sarkar *et al.* [25] utilized a residual network weighting and a spatial pyramid pooling network. They also proposed the use of a loss function called End Point Error (EPE) to preserve the lesion boundaries. Furthermore, Jahanifar *et al.* [26] integrated a multilevel segmentation algorithm, regional contrast, background descriptors, and a random forest regressor to create saliency scores for each region in the image. This method gives poor segmentation results with low contrast dermoscopic images.

To add image appearance information as well as contextual information, Mirikharaji *et al.* [20] employed the U-Net based method to predict the pixel-wise probability of a skin lesion segmentation. It achieved a Dice coefficient of 90.11% and an IoU score of 83.30% on ISBI2016 dataset. Furthermore, Li *et al.* [21] proposed a transformation consisting of a self-ensemble model, which enhances the regularization effects by utilizing the unlabeled data. It achieved a Dice coefficient of 87.40% and IoU scores 79.87% on ISBI2017 dataset. In turn, Venkatesh *et al.* [22] used the U-Net model that is based on multi-scale input with a shortcut connection at each block of the U-Net. The suggested method has evaluated on ISBI2017 dataset, obtaining a Dice coefficient and IoU scores of 85.60% and 76.40%, respectively. Galdran *et al.* [23] also exploited the U-Net architecture and used color constancy methods to normalize the color throughout the dataset images while retaining the estimated illumination information, enabling them to randomly change the color and illumination of normalized images during the training process. They achieved a Dice coefficient of 82.40% on ISBI2017 dataset.

Bi *et al.* [27] proposed a FCN-based class-specific training to extract visible features of different kinds of skin lesions and a probability-based step-wise combination of

the derived class-specific segmentation maps to guarantee visible persistence of the segmented regions. Also, Yuan [6] introduced a FCN-based model, which contains 29 layers to segment skin lesions from dermoscopic images. They have employed an upsampling and deconvolutional layers to compute multi-resolution loss while carrying over the global perspective from pooling layers. However, their model requires several pre- and post-processing operations, such as color space transformations and threshold selection. Moreover, Al-Masni *et al.* [5] proposed a full resolution convolutional networks (FrCN), that learns the full resolution features of each pixel of the input data without the requirement of pre or post-processing methods.

C. GAN-BASED METHODS

Recently, several GAN-based segmentation models have been applied to medical images. For instance, cGAN has been used in [11] and [36] to segment breast cancer subtypes, in which a Dice loss was used in the generator network to refine pixel-wise segmentation results. To segment skin lesions, Xue *et al.* [14] proposed an adversarial-based model with residual blocks and skip connections. The model was evaluated on ISBI2017 dataset, achieving a Dice coefficient of 86.70% and an IoU score of 78.50%. Moreover, Bisla *et al.* [19] introduced the Deep Convolutional Generative Adversarial Network (DCGAN) and ResNet-50 models to jointly segment the skin lesion and classify the lesions into benign and malignant. They exploited pre-processing steps to suppress the artifacts from the skin images. With ISBI2017 and ISIC2018 test datasets, they obtained IoU scores of 77.00% and 70.20%, respectively.

Most of the methods stated in Table 1 have utilized data augmentation/preprocessing techniques in the training phase while others applied postprocessing techniques (e.g., morphological operations) on the resulting masks. In turn, during the training of the proposed FCA-Net, we utilize the original images of ISBI2016, ISBI2017, and ISIC2018 datasets without applying any data augmentation technique. For a fair comparison with the state-of-the-art methods, we separately trained and tested the proposed model on the training and testing sets of the aforementioned datasets. The proposed cGAN model that includes FCA blocks and a multi-scale stage highlights the most important features (of a highly receptive field) and disregards the artifacts from images.

III. METHODOLOGY

A. THE FACTORIZED CHANNEL ATTENTION BLOCK

Figure 2 presents the design of the proposed FCA block, which applies a weighted aggregation between the output features of two mechanisms: channel attention (see upper branch) [15] and residual 1-D factorized convolutions (see lower branch) [16]. The proposed FCA block increases the representational power of features computed by encoder layers in generator and discriminator networks.

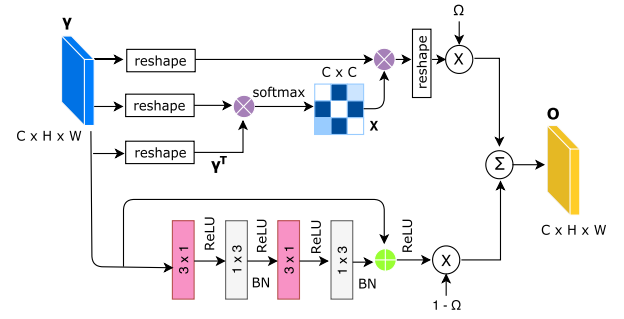


FIGURE 2. Proposed FCA block with integration of channel attention and residual 1-D factorized convolution.

1) CHANNEL ATTENTION

This mechanism is intended to boost feature channels that have similar values in the same image positions. Assume that $\gamma \in \mathbb{R}^{C \times H \times W}$ is the activation map (i.e. set of features) obtained by the original encoder layer. To calculate the channel attention map $X \in \mathbb{R}^{C \times C}$, γ is firstly reshaped to $\mathbb{R}^{C \times N}$ ($N = H \times W$), then multiplied by its transpose, and finally normalized with the softmax function:

$$x_{ji} = \frac{\exp(\gamma_i \cdot \gamma_j)}{\sum_{i=1}^C \exp(\gamma_i \cdot \gamma_j)} \quad (1)$$

where γ_i and γ_j are vectors of length N , containing the values of all map positions in channels i and j , respectively, and $\gamma_i \cdot \gamma_j$ represents their dot product. Hence, x_{ji} represents a normalized correlation degree between those two channels. The output of the channel attention branch, $O_1 \in \mathbb{R}^{C \times H \times W}$, can be expressed for each channel j as follows:

$$O_{1j} = \eta \sum_{i=1}^C (x_{ji} \gamma_i) + \gamma_j \quad (2)$$

where $\sum_{i=1}^C (x_{ji} \gamma_i)$ includes the feature values of all channels modulated by the correlation degree between each channel with respect to the j^{th} channel. Moreover, this summation is weighted by η , which is a learned weighting parameter and then added to the original activation map. In this way, channels that present more similarities increase their relevance in the output. This mechanism improves the segmentation accuracy because relevant patterns corresponding to skin lesion areas create high activation values in several feature channels, while other irrelevant patterns, like healthy skin areas or hairs, may have associated very few feature channels as they are representatives.

2) RESIDUAL 1-D FACTORIZED CONVOLUTIONS

This mechanism is intended to boost feature values that are similar in different image positions. The core working is the same as a typical residual convolution, i.e., to learn the difference between input and output activation maps, but using factorized 1-D kernels instead of regular 2D kernels. We hypothesize that the output of this branch will tend to detect image areas that present similar skin patterns in neighboring image positions. This output does not entirely take

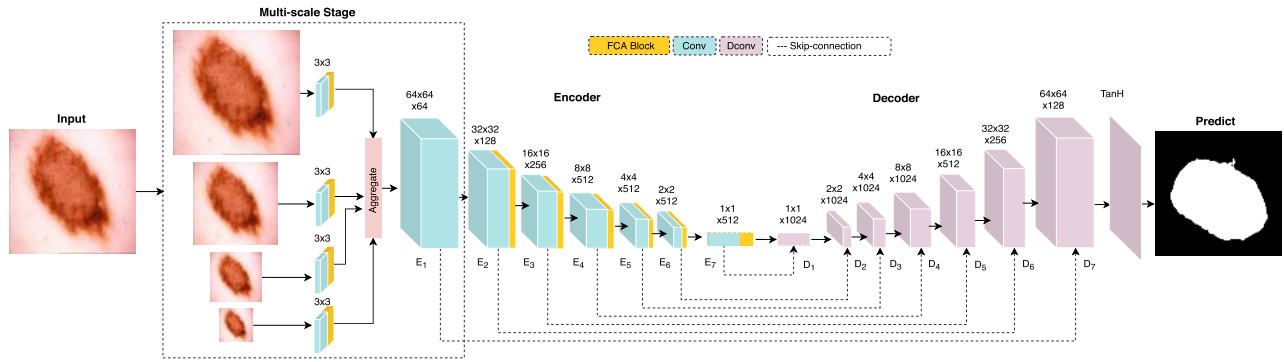


FIGURE 3. The architecture of the generator network.

over the role of the full spatial attention mechanism, where similar patterns in disconnected image areas can be enhanced. Nevertheless, in the skin lesion context, the residual convolution is enough if we assume that most of the lesions are contained in a contiguous image region.

Factorized kernels can effectively preserve the spatial information of regular 2D kernels and maintain the accuracy with significantly less computation.

Assume that $W \in \mathbb{R}^{C \times d^h \times d^v \times F}$ are the weights of a typical 2D convolutional layer, where C is the number of input planes (i.e. channels), F is the number of output planes (i.e. feature maps) and $d^h \times d^v$ is the kernel size of each feature map (typically $d^h \equiv d^v \equiv d$). $f^i \in \mathbb{R}^{d^h \times d^v}$ is the i^{th} kernel in the layer. As proposed in [16], f^i can be expressed as a linear combination of 1-D filters:

$$f^i = \sum_{k=1}^K \sigma_k^i \bar{v}_k^i (\bar{h}_k^i)^T \quad (3)$$

where σ_k^i is a scalar weight, K is the rank of f^i , \bar{v}_k^i and $(\bar{h}_k^i)^T$ are vectors of length d . The i^{th} output of the decomposed layer O_{2i} can be expressed as a function of its input γ , as follows:

$$O_{2i} = \varphi \left(b_i^h + \sum_{l=1}^L \bar{h}_{il}^T * \left[\varphi \left(b_i^v + \sum_{c=1}^C \bar{v}_{lc} * \gamma_c \right) \right] \right) + \gamma_i \quad (4)$$

where $\varphi(\cdot)$ represents the non-linearity of the 1-D decomposed filters (where we used ReLU), b_i^h and b_i^v are the horizontal and vertical biases of each filter. The residual strategy of this branch combines the original features provided by the previous layer with a new set of feature maps with 1-D convolutional filters. 1-D convolutional filters have intrinsically less computational cost and less number of parameters than their 2D equivalent filters. Additionally, the 1-D combinations improve the compactness of the generator layers by minimizing redundancies in the features coming from the previous 2D convolution layers and theoretically improve the learning capacity. Besides, the residual 1-D kernel factorization is faster in terms of the computation time than the normal non-bottleneck [37] and has fewer parameters than the bottleneck design keeping a high learning capacity and accuracy.

To determine the final output of the FCA block, we aggregate the channel attention and the residual 1-D factorized convolutions outputs as follows:

$$O = (1 - \Omega) \times O_1 + \Omega \times O_2 \quad (5)$$

Here, Ω is the weighting factor. We checked the system performance for Ω values from 1.0 to 0.0, in steps of 0.1. We have found that $\Omega = 0.3$ provides the best results.

B. NETWORK ARCHITECTURE

The proposed model comprises a generator and a discriminator network. The generator network includes an encoder section and a decoder section. As shown in Figure 3, both encoder and decoder sections include seven sequential layers (E_n refers to an encoder layer and D_n refers to a decoder layer).

1) THE ENCODER

We use a multi-scale input strategy [38], where the input images are resized into three different scales with ratios of 1/8, 1/4 and 1/2 of the original size. In this way, the FCA-Net explicitly deals with variation in resolution, object size and image scale, by encouraging the development of filters which are scale-variant, while constructing a scale-invariant representation. Scale-variant filters help segment some small skin lesion pixels. After each scale, we add a convolution layer with 3×3 kernels along with the proposed FCA block to extract more rich features from a skin lesion. The sizes of the features that are fed into the aggregation module from up to down are $128 \times 128 \times 64$, $64 \times 64 \times 64$, $32 \times 32 \times 64$ and $16 \times 16 \times 64$, respectively. Before aggregating the features of all scales, lower-scale features are upsampled to the size of the feature vector extracted from the original image ($128 \times 128 \times 64$), and then we input them into the encoder layers. We added the proposed FCA block in all layers of the encoder part, and we did not add it to the decoder layers. We used batch normalization with *LeakyReLU* (slope 0.2) after the first six layers of encoder and the *ReLU* activation function after E_7 . The size of all convolutional kernels is 4×4 with a stride of 2. The encoder can learn low-level features

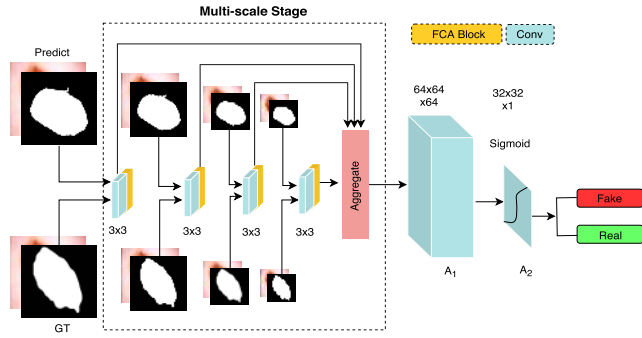


FIGURE 4. The architecture of the discriminator network.

of the skin images, such as spatial information (e.g., edge, intensity, texture) throughout the training process.

2) THE DECODER

To avoid overfitting, we used batch normalization and dropout (rate = 0.5) in D_1 , D_2 and D_3 . We added the *ReLU* activation function after each layer of the decoder. The size of the deconvolutional kernels is 4×4 with a stride of 2. We also added skip connections between each convolutional layer to its corresponding deconvolutional layer. To outline the segmented skin lesion into a binary mask, we added *Tanh* after D_7 . We used a threshold of 0.5 to convert the output of *Tanh* activation function to binary masks.

3) THE DISCRIMINATOR NETWORK

In Figure 4, similar to the generator network, we apply a multi-scale stage with the FCA block to enhance the scale independence of the discriminator between the ground truth and the generated masks. The masks are also resized into three scales with ratios 1/8, 1/4 and 1/2 of the original image size. The extracted features are up-sampled, concatenated and inputted into two successive convolutional layers A_1 and A_2 . We add *LeakyReLU* activation function in A_1 and *sigmoid* function in A_2 .

C. LOSS FUNCTION

To optimize the proposed segmentation model, we employ a loss function composed of three terms: Adversarial loss as Binary Cross Entropy (BCE), ℓ_{L1} loss and End Point Error (EPE) proposed in [25]. Assume x is the skin image containing a lesion, y is the ground truth mask, $G(x, z)$ and $D(x, G(x, z))$ are the outputs of the generator and the discriminator, respectively, the loss function of the generator network G is as follows:

$$\begin{aligned} \ell_{Gen}(G, D) = & \mathbb{E}_{x,y,z}(-\log(D(x, G(x, z)))) \\ & + \lambda \mathbb{E}_{x,y,z}(\ell_{L1}(y, G(x, z))) \\ & + \beta \mathbb{E}_{x,y,z}(\ell_{EPE}(y, G(x, z))) \end{aligned} \quad (6)$$

where z is a random variable, and β and λ are empirical weighting factors. The ℓ_{L1} loss forces the model to suppress the outliers and artifacts and speed up the optimization process.

The EPE loss [39] compares the magnitude and orientation of the edges of the predicted mask with its ground truth for preserving the boundaries of the segmented regions. The EPE can be defined as:

$$L_{epe} = \sqrt{(G(x, z)_x - y_x)^2 + (G(x, z)_y - y_y)^2} \quad (7)$$

where $(G(x, z)_x, G(x, z)_y)$ and (y_x, y_y) are the first derivatives in x and y directions of $G(x, z)$ and y , respectively.

In the discriminator network, we only used the BCE loss, which is defined as:

$$\begin{aligned} \ell_{Dis}(G, D) = & \mathbb{E}_{x,y,z}(-\log(D(x, y))) \\ & + \mathbb{E}_{x,y,z}(-\log(1 - D(x, G(x, z)))) \end{aligned} \quad (8)$$

The optimizer will fit D to maximize the loss values for ground truth masks (by minimizing $-\log(D(x, y))$) and minimize the loss values for generated masks (by minimizing $-\log(1 - D(x, G(x, z)))$). The generator and discriminator networks are optimized concurrently, one optimization step for both networks at each iteration, where G tries to generate an accurate lesion segmentation mask and D learns how to discriminate between the synthetic and the real segmentation masks.

IV. EXPERIMENTAL RESULTS AND DISCUSSION

Datasets: To assess the efficacy of the proposed model, we use three skin lesion challenge datasets, which are publicly available: ISBI2016¹ [40], ISBI2017² [13] and ISIC2018³ [41]. The images of the datasets were acquired using different devices at several medical centers worldwide. In the ISBI2016 dataset, the training set has 900 annotated images while the testing set has 379 annotated images. The size of the images varies from 542×718 to 2848×4288 pixels. The ISBI2017 dataset has three sets: the training set (2000 images), validation set (150 images) and testing set (600 images). In the ISIC2018 dataset, the training set has 2594 images, the validation set has 100 images and the testing set contains 1000 images. Ground truth of mask images (provided for training and used internally for scoring validation and test phases) were generated using several techniques, but all data were reviewed and curated by practicing dermatologists with expertise in dermoscopy. Our model has been trained on randomly chosen 2546 skin lesion images from the ISIC 2018 dataset and validated on the rest 48 images. The model is then evaluated on the validation and testing sets. This process has only been used to tune the hyperparameters of the model. Note that the trained model is evaluated on the testing sets of ISBI2016 and ISBI2017 datasets. Our model is further assessed on the ISIC2018 validation and test sets.

Evaluation Metrics: Assume A is the ground truth and B is the segmented mask (using our model). The true positive (TP) rate is defined as $TP = A \cap B$, which is the region of the segmented part common in both A and B .

¹<https://challenge.kitware.com/#challenge/560d7856cad3a57cfde481ba>

²<https://challenge.kitware.com/#challenge/583f126bcad3a51cc66c8d9a>

³<https://challenge.kitware.com/#challenge/5aab46f156357d5e82b00fe5>

TABLE 2. Analyzing different configurations of the proposed method with ISBI2016 and ISBI2017 datasets.

Methods	ISBI2016					ISBI2017				
	ACC	Dice	IoU	SEN	SPE	ACC	Dice	IoU	SEN	SPE
BL	92.66	88.23	81.59	86.78	94.42	92.08	83.52	69.36	76.47	94.11
BL+CA	93.67	88.97	82.62	87.78	94.91	93.86	83.99	73.50	79.37	94.64
BL+FK	94.95	90.33	84.16	86.36	96.61	94.29	84.52	75.10	83.46	96.79
FCA-Net w/o MS	95.69	91.75	86.01	91.47	97.88	95.11	86.54	76.88	87.36	96.92
FCA-Net	96.97	93.94	87.58	92.42	98.62	96.29	88.28	78.94	88.09	97.36

The false positive (FP) rate is defined as $\bar{A} \cap B$, which is the segmented area not belonging to A. The false-negative (FN) rate is defined as $A \cap \bar{B}$, which is the true area missed by the segmentation model. To assess the performance of the segmentation models, we use the accuracy (ACC), Dice Coefficient (Dice), Jaccard index (IoU), sensitivity (SEN), and specificity (SPE) [42].

Parameter Settings: Each input image is resampled to 128×128 pixels and apply a normalization for rescaling the pixel values between [0, 1] before feeding it into our network. The hyper-parameters of the model were empirically tuned. We used Adam optimizer with $\beta_1 = 0.5$ and $\beta_2 = 0.999$, a learning rate of 0.0002 and a batch size of 2. The weighting factors of the ℓ_{L1} and EPE losses (λ and β) were set to 100 and 50, respectively. We have trained our model till 300 epochs and during training, the best results were obtained with 240 epochs.

Implementation Details: The experiments were conducted out on an NVIDIA GeForce GTX 1070 with 8 GB of video RAM. The operating system was Ubuntu 16.04 using a 3.4 GHz Intel Core-i7 with 16 GB of RAM. The main required packages involve Python 3.6, CUDA 9.1, cuDNN 7.0 and PyTorch 0.4.1. The codes of the proposed model are publicly available at <https://github.com/vivek231/Skin-Project>.

A. ABLATION STUDY

To demonstrate the effect of each part of the proposed block, an ablation study has been done. We firstly trained a baseline (BL) model without adding the channel attention or factorized convolution blocks. Then, we added the channel attention block to the encoder layers of the generator network (called the BL+CA model). Furthermore, the factorized kernel was also separately added to the encoding layers (called the BL+FK model). Finally, the proposed model (FCA-Net) is constructed by adding both CA and FK blocks with and without multi-scale. Note that all models used in this ablation study have been trained on the ISIC2018 dataset and tested on the ISBI2016 and ISBI2017 that appear in the training set of the ISIC2018 dataset.

Table 2 presents the results of the BL, BL+CA and BL+FK models. With the dataset of ISBI2016 dataset, the BL model yields a Dice and IoU scores of 88.23% and 81.59% respectively, while the BL+CA model provides a small improvement of 0.5% and 1.0% for Dice and IoU scores, respectively. This improvement is achieved because the CA

block explicitly models inter-dependencies among channels. Besides, the BL+FK model gives a Dice of 90.33% and IoU scores of 84.16% yielding better improvement than BL and BL+CA models. In turn, the proposed FCA-Net achieves an improvement of around 4.0% and 3.0% of Dice and IoU scores, respectively, better than the BL+FK model. Similarly, on ISBI2017 dataset, FCA-Net improved the results comparing with other checked strategies and achieved a Dice of 88.28% and IoU scores of 78.94%.

To demonstrate the effectiveness of the multi-scale stage, in Table 2 we provide the results of FCA-Net with and without employing the multi-scale stage (w/o MS). As shown, the multi-scale stage improves the segmentation performance of the proposed FCA-Net model with an increment of approximately 2% in Dice and IoU scores.

In Fig. 5, we present a couple of difficult samples, jointly with visualizations of the activation maps created by the second (Conv2) and third (Conv3) encoder layers, as well as the output of the last decoder layer (Deconv7) compared with the ground truth (more examples in Fig. 8). The layer outputs have been obtained with four variants of the FCA block (see figure captions).

The BL variant (basic cGAN) can distinguish lesion from non-lesion areas in low-level layers of the network, as shown with pink and green zones in Conv2 activation maps, but artifacts like hair and texture/color variability are interfering with the detection of the lesion area. Besides, for image 2 there is a false lesion detection at the left of the true lesion. Although next layers can get rid of such inaccuracies up to a certain degree, the final outputs show a high amount of false negatives (red pixels in output 1) or a fair amount of false positives (green pixels in output 2). Also, note that the output for image 1 comprises small holes (red spots inside the yellow area).

The CA variant obtains a more consistent identification of the target classes. In Conv2, almost all lesion region renders one single color (dark red), although their pixels show varying shading due to different activation degrees (especially for image 1). In Conv3, the lesion areas are more compact (shaded in dark blue), but for image 1 there is a visible break. Nevertheless, this groove will be filled in by further layers. Noticeably, the output for image 1 has reduced the number of FN with respect to the baseline output, although there are still too many red pixels. The output for image 2 has not trimmed the false (green) elongation at the right of the lesion, but it has trimmed the extra green pixels at its left boundary, so the performance metrics have been improved. Despite that these

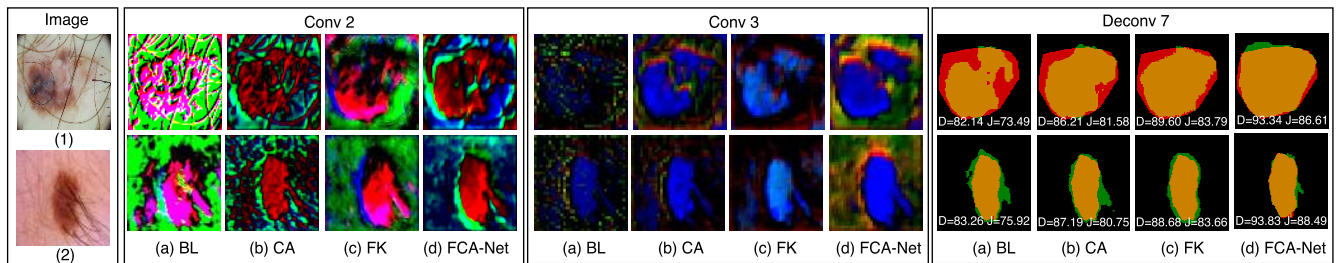


FIGURE 5. Visualization of two sample images and the corresponding activation maps generated by the second and third convolutional layers of the generator network in four variants w.r.t. the use of the FCA block: BL is our baseline cGAN; CA includes the channel attention branch; FK includes the residual 1-D factorized convolutions; FCA-Net is our fully-fledged network. The figure also shows the output (Deconv7) of the variants, graphically compared with the ground truth segmentation, color-coded as yellow: TP, green: FP, red: FN and black: TN, as well as the Dice and IoU indexes for each experiment.

TABLE 3. Analyzing the effect of the FCA block on different segmentation models on ISBI2016 dataset.

Methods	Without FCA Block					With FCA Block				
	ACC	Dice	IoU	SEN	SPE	ACC	Dice	IoU	SEN	SPE
FCN8	90.10	80.60	72.52	81.11	97.17	91.02	83.37	74.42	85.15	98.39
U-Net	89.23	79.51	71.66	80.26	96.87	91.22	82.10	73.92	84.92	97.41
SegNet	92.36	83.74	75.92	84.67	97.02	93.41	86.61	78.45	87.51	97.87
Link-Net	92.97	84.76	77.01	85.97	97.26	93.74	87.53	79.46	88.02	97.98
RefineNet	93.03	83.97	75.58	86.08	95.71	93.19	85.82	77.71	89.41	96.76
FCA-Net	92.66	88.23	81.59	86.78	94.42	96.97	93.94	87.58	92.42	98.62

TABLE 4. Analyzing the effect of the FCA block on different segmentation models on ISBI2017 dataset.

Methods	Without FCA Block					With FCA Block				
	ACC	Dice	IoU	SEN	SPE	ACC	Dice	IoU	SEN	SPE
FCN8	86.38	64.55	57.38	61.71	88.26	89.93	68.85	59.98	64.90	92.82
U-Net	85.84	63.03	55.39	60.87	88.24	90.04	67.89	59.39	63.20	95.03
SegNet	88.76	72.45	64.30	71.52	92.44	91.66	75.24	66.06	74.68	96.02
LinkNet	89.62	75.97	69.37	73.81	94.35	92.57	79.59	71.20	77.77	96.01
RefineNet	88.28	74.05	63.64	73.88	94.36	91.52	75.31	67.82	74.92	95.57
FCA-Net	92.08	83.52	76.36	84.47	94.11	96.29	88.28	78.94	88.09	97.36

outputs still present some misled areas, the obtained degree of improvement empirically proves that the channel attention mechanism can significantly smooth the effect of artifacts since lesion features are consistently enhanced.

The FK variant obtains a much more compact coloring of the target areas, thanks to the local spatial coherence provided by the residual filters. Despite this good property, lesion regions in Conv2 maps contain different colors (red, pink, dark blue), which indicates that several feature channels are responsible for characterizing different areas of the lesions. In Conv3, however, the lesion features are more consistent, showing one single blue color. Segmentation outputs are better than outputs from BL and CA variants, which highlight the power of the residual branch.

Finally, the FCA-Net combines the good properties of CA and FK variants, since the activation maps from Conv2 and Conv3 tend to be both spatial and channel coherent within the lesion area. For the non-lesion area, however, in Conv3 there is a color gradation (halo) around the lesion, which may turn out into misleading boundary delineation, like the small green area on top of the lesion in the final segmentation of image 1. This inconvenience is largely compensated by the significant

reduction of false negatives in output 1. At the same time, there is also a significant reduction of false positives in output 2. In summary, the combination of the two branches of the FCA block helps each other to provide the best results of the four variants.

In our experiments, we added the proposed FCA block to different state-of-the-art image segmentation methods (FCN8 [33], U-Net [34], SegNet [35], LinkNet [43] and RefineNet [44]) and evaluated them on ISBI2016 and ISBI2017 dataset. Tables 3 and 4 present the results of all models with and without the proposed FCA block. It is important to note that all compared models have a multi-scale input layer (they have the same configuration of the multi-scale input layer of our model). We can see that all models give better results when adding the FCA block. Besides, our model gives the best results among all evaluated methods.

To demonstrate the effectiveness of the proposed model, we provide descriptive statistics of Dice and IoU scores. In Fig. 6 (a,b), we show the boxplots of the Dice and IoU scores of the proposed model, FCN8, U-Net, SegNet, LinkNet, and RefineNet models with the ISBI2016 dataset. As shown in Fig. 6 (a), among the tested models, the proposed

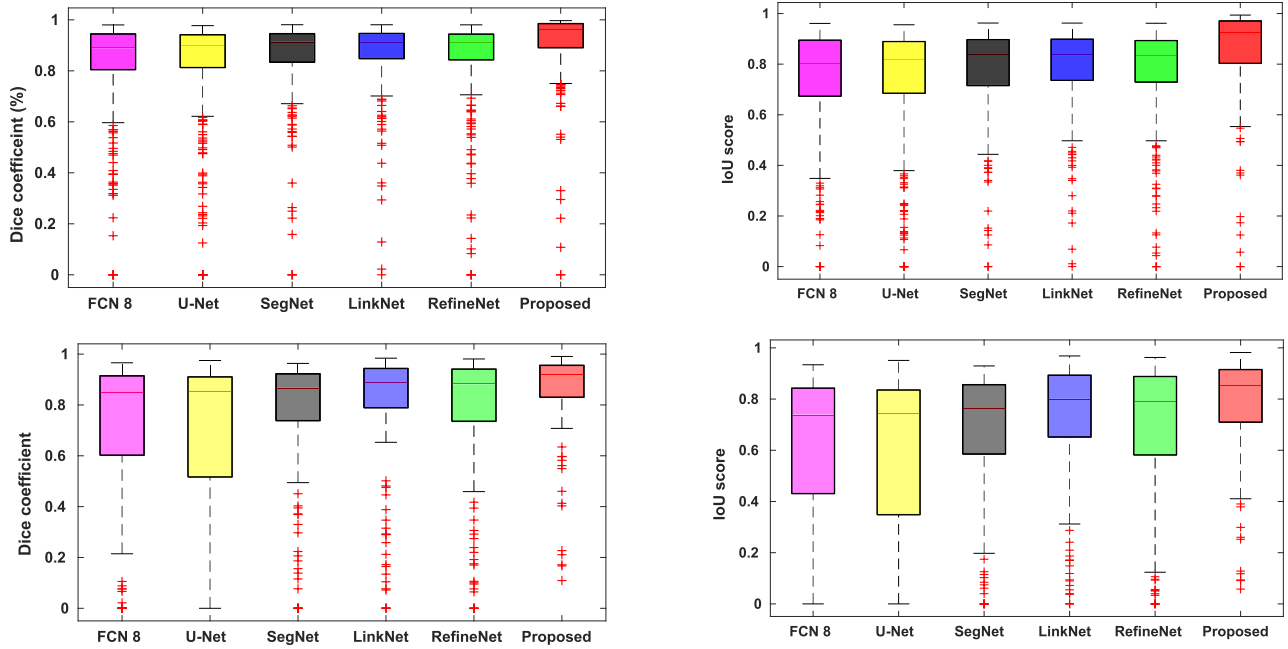


FIGURE 6. Boxplots of Dice and IoU scores for all test samples in ISBI2016 dataset (a,b) in the upper row and ISBI2017 dataset (a,b) in the bottom row. Different color boxes indicate the score range of several methods, the red line inside each box represents the median value, box limits include interquartile ranges Q2 and Q3 (from 25% to 75% of samples), upper and lower whiskers are computed as 1.5 times the distance of upper and lower limits of the box, and all values outside the whiskers are considered as outliers, which are marked with the (+) symbol.

model has the highest mean Dice score and the smallest standard deviation with few outliers. Indeed, with the 379 images of the test set of ISBI2016, the proposed model produces 13 outliers of the boxplot with IoU scores while the LinkNet and SegNet models have 18 and 17 outliers, respectively. In turn, Fig. 6 (c,d) shows the boxplots of Dice and IoU scores of the six models with ISBI2017 dataset. The FCN8 and U-Net models have no outliers but the deviation of IoU metric is much bigger than the one of our model as shown in Fig. 6 (d). With the 600 images of the test set of ISBI2017, our model has 10 outliers with IoU scores while the RefineNet model has 14 outliers.

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

In Table 5, the FCA-Net model is compared with 8 state-of-the-art skin lesion segmentation methods on ISBI2016 dataset. For fair comparisons, we have also trained and tested FCA-Net with the ISBI2016 training set and evaluate them with the ISBI2016 test set. As shown, FCA-Net gives a Dice score of 92.80% and an IoU score of 86.41% that is the second best value. Li *et al.* [45] achieve a sensitivity score higher than our model but we obtain better specificity and accuracy.

Table 6 shows a comparison between the results of FCA-Net and 11 state-of-the-art skin lesion segmentation methods on ISBI2017 dataset. For fair comparisons, we have also trained FCA-Net with the ISBI2017 training set and evaluate it with the ISBI2017 test set. The IoU score of [21] is slightly better than our FCA-Net method, our score is the second-best. Moreover, we achieve the best Dice and

TABLE 5. Comparing the proposed model with 8 state-of-the-art methods on ISBI2016 dataset. Best results are marked in bold.

Methods	ACC	Dice	IoU	SEN	SPE
ExB ⁴	95.30	91.00	84.30	91.00	96.50
Mirikharaji et al [20]	95.02	90.11	83.30	90.15	97.00
Hang Li et al. [45]	95.90	93.10	87.00	95.10	96.00
Lei Bi et al. [27]	95.80	91.70	85.90	93.10	96.00
Jahanifar et al. [26]	94.30	90.70	83.80	90.10	96.60
CUMED [17]	94.90	89.70	82.90	91.10	95.70
Rehman et al. [8]	95.20	89.50	82.20	88.00	96.90
Yuan et al. [6]	95.50	91.20	84.70	91.80	96.60
FCA-Net	95.93	92.80	86.41	91.63	97.07

accuracy scores. Also, [25] obtains a specificity 1% higher than FCA-Net but the values of other metrics are much lower than the ones of the later model. Note that the results of related methods mentioned in Table 5 and Table 6 are taken from the cited references.

The performance of the FCA-Net model is also assessed on the validation set of ISIC2018. The segmented images are submitted to the *Leaderboards* platform, which calculates an IoU_{th} score of the provided results. Note that this IoU_{th} score is computed as follows [41]:

$$IoU_{th} = \begin{cases} IoU, & \text{if } IoU > 0.65 \\ 0, & \text{otherwise.} \end{cases} \quad (9)$$

In Fig. 7, we show a screenshot of the live leaderboard. Our model has been ranked in the 11th position among 52 tested models, at the time of the submission. On the validation set, it has achieved an IoU score of 77.2%. On the test set (on

Rank	User	Title	Organization	Date	Score
1	rashika mishra	rcnn_superpixels	UTD-geobiolab	Wed, 20 Mar 2019, 9:59:19 am	0.830
2	c c (dense)	dense	xc	Fri, 26 Apr 2019, 8:10:34 am	0.802
3	Md. Mostafa Kamal Sarker	Ensemble SLSDeep	Md. Mostafa Kamal Sarker	Fri, 8 Mar 2019, 8:34:19 am	0.794
4	Vinicius Ribeiro (DeepLab V3+ model trained for 150 epochs with 20 epochs patience using gaussian noise, color and contrast degradation as data augmentation (last model))	DeepLab V3+	RECOD Titans	Fri, 10 May 2019, 4:45:59 pm	0.793
5	Umasethi Sivasarani (mask-rcnn)	test_mask-rcnn_real	test	Mon, 24 Jun 2019, 2:14:21 am	0.788
6	Md. Mostafa Kamal Sarker (GAN)	GAN	IRCV	Fri, 17 May 2019, 1:33:27 am	0.784
7	rashika mishra (mrcnn with colorspace)	mrcnn with colorspace	GeoBioblab-UTD	Fri, 8 Mar 2019, 8:09:48 pm	0.782
8	rashika mishra (contour maskrcnn)	contour maskrcnn	UTD	Fri, 8 Mar 2019, 6:15:48 pm	0.778
9	Vinicius Ribeiro (DeepLab V3+ model trained for 100 epochs and infinite patience using gaussian noise, color and contrast degradation as data augmentation (best model))	DeepLab V3+ model	RECOD Titans	Sun, 12 May 2019, 3:40:07 pm	0.774
10	Vinicius Ribeiro (DeepLab V3+ model trained for 150 epochs using gaussian noise, color and contrast degradation as data augmentation)	DeepLab V3+ model	RECOD Titans	Tue, 30 Apr 2019, 3:37:48 am	0.773
11	Vivek Kumar Singh	FCA-Net	IRCV	Thu, 11 Jul 2019, 11:59:12 am	0.772
12	Vinicius Ribeiro (DeepLab V3+ model trained for 100 epochs and infinite patience using gaussian noise, color and contrast degradation as data augmentation (last model))	DeepLab V3+ model	RECOD Titans	Sun, 12 May 2019, 3:48:51 am	0.770
13	Vivek Singh	GAN	IRCV	Thu, 11 Jul 2019, 1:20:48 pm	0.765
14	Zhiye Wang (sdu_baseline)	sdu_baseline	Shandong	Mon, 7 Jan 2019, 0:00:00 am	0.762

FIGURE 7. Screenshot of the live leaderboard on ISIC2018 dataset challenge.

TABLE 6. Comparing the proposed model with 11 the state-of-the-art methods on ISBI2017 datasets. Best results are marked with the bold. Dashes (–) indicate that results are not reported in the cited papers.

Methods	ACC	Dice	IoU	SEN	SPE
Bisla et al. [19]	–	–	77.00	–	–
Almasani et al. [5]	94.03	87.08	77.11	85.40	96.69
Xiaomeng et al. [21]	94.30	87.40	79.80	87.90	95.30
Xue et al. [14]	94.10	86.70	78.50	–	–
Yuan et al. [6]	93.40	84.90	76.50	82.50	97.50
Berseth et al. [7]	93.20	84.70	76.20	82.00	97.80
Jahanifar et al. [26]	93.00	83.90	74.90	81.0	98.10
Venkatesh et al. [22]	93.60	85.60	76.40	83.00	97.60
Galdran et al. [23]	92.30	82.40	73.50	81.30	96.80
Sarker et al. [25]	93.60	87.80	78.20	81.60	98.30
Sulaiman et al. [24]	93.20	85.10	76.70	93.00	90.50
FCA-Net	94.95	87.80	78.65	87.91	97.05

1000 skin images), our method has achieved an average Dice score of 85.8% and IoU score of 78.2%.

An additional comparison has been performed by comparing the FCA-Net model with FCN, U-Net, SegNet, FrCN (reported in [46]), GAN-FCN [47], and hand-crafted [48]. In the case of GAN-FCN model, GAN is used to derive additional training data from ISIC2018 dataset, and then this data is combined with the original training data to train the FCN model for skin lesion segmentation. The authors of [48] transform skin images into the RGB space and trained a color classifier to discriminate between lesions and normal skin tissue based only on RGB color vectors. Then, they use a Gaussian mixture model (GMM) to model the probability density functions of skin lesions and a support vector machine regression algorithm to segment the images. As shown in

TABLE 7. The performance of FCA-Net on the ISIC2018 validation dataset. The proposed model has been evaluated on skin lesion leaderboard (<https://submission.challenge.isic-archive.com/>).

Methods	IoU _{th} (%)
FCN	74.70
U-Net	54.40
SegNet	69.50
FrCN	74.60
GAN-FCN [47]	77.80
Hand-crafted [48]	66.30
FCA-Net	77.20

Table 7, our model achieves an IoU score close to the one of GAN-FCN. In turn, the U-Net model yields the worst results with an IoU of 54%.

Fig. 8 presents qualitative results of skin lesion segmentation that include a variety of challenging conditions: hair presence, blurriness, illumination variations (intensity, chromaticity, fading, etc.), fuzzy borders, irregular borders, sharp or straight borders, big and small lesions, and two-color lesions. Each row of Fig. 8 includes a skin lesion image along with its segmentation obtained with FCN8, U-Net, SegNet, Refine-Net, LinkNet and the proposed model. To visualize the accuracy of each model, we compare the ground truth mask with the generated mask and use four different color codes to mark up the classification result for each pixel. Note that yellow refers to TP, red refers to FN, green refers to FP and black refers to TN. An ideal segmentation model will assign yellow to skin lesion pixels and black to the background pixels.

For all examples, our model achieves the best Dice and IoU scores with all images (Fig. 8). Moreover, it produces a tiny

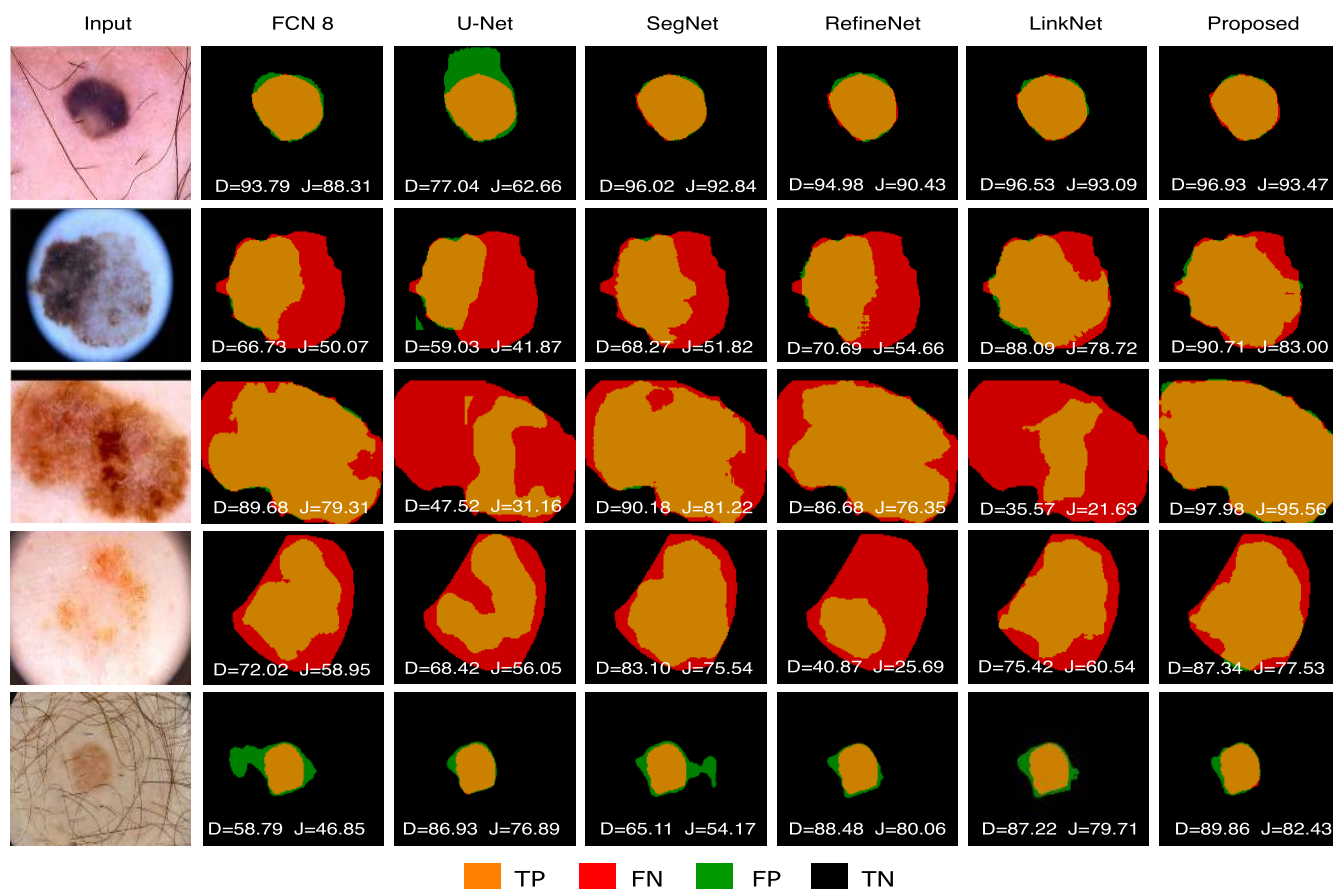


FIGURE 8. Skin lesion segmentation using the FCN8, U-Net, SegNet, RefineNet, LinkNet and FCA-Net models. Note that D and J represent the Dice coefficient and the IoU score, respectively. Further visualization for the segmentation results of the proposed method can be found at <https://youtu.be/GeUM8FghFA>.

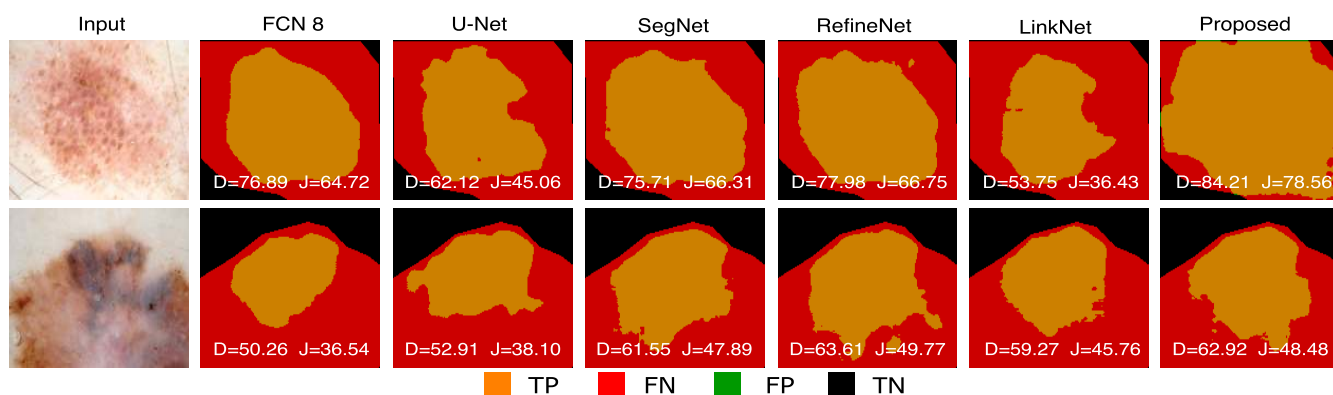


FIGURE 9. Examples of inaccurately segmented lesions with the proposed FCA-Net model and compared with other baseline segmentation models. Note that D and J represent the Dice coefficient and the IoU score, respectively.

amount of red or green pixels, which are usually distributed around the border of the skin lesion, while the other methods have a higher number of FP and FN.

The sample in the first row has a small lesion with thick hair in the background. The U-Net model failed to properly segment by over-extending the lesion area, while the rest of the models give accurate segmentation and high Dice and

IoU scores. The sample of the second row has low contrast with bad illumination conditions, along with a framing effect due to the optic lenses. In this case, FCN8, U-Net, SegNet, RefineNet, and LinkNet give a low Dice and IoU scores ($\leq 88\%$), while our method outputs a decent result. In the third row, the pixels inside the lesion region have inhomogeneous colors, and some of them are similar to the ones of the

background. The U-Net and LinkNet method give bad segmentation results while the proposed model achieves the most accurate fit of the lesion, with very high Dice and IoU scores. In the fourth row, the lesion region is not homogeneous and has low contrast. With this case, our model achieves a Dice of 87.34% and an IoU of 77.53%, which correspond to the best matching of the lesion. The last row has a small skin lesion with very dense hair in the background. Here, the proposed model obtains promising segmentation results compared to the other models.

C. LIMITATIONS

Although the proposed FCA-Net outperforms several deep learning-based models (FCN8, UNet, SegNet, RefineNet and LinkNet), it may produce inaccurate results with some cases as shown in Figure 9. As we can see, it is difficult to manually segment such images. The skin image of the first row has fuzzy boundaries, low contrast and intensity inhomogeneity. With this case, our model achieves a Dice score of 84.21% and an IoU score of 78.56%. The other five models provide less accurate segmentation results, while the FCA-Net almost fit the lesion area. The second sample has two distinct color shades in the lesion. All models have failed to accurately segment the lesion because of this dual shading, selecting the stronger one as the lesion and misclassifying the weaker one as background. It leads to IoU scores less than 50% in all segmentations, although FCA-Net has achieved the second best result.

V. CONCLUSION AND FUTURE WORK

In this paper, we have proposed an accurate skin lesion segmentation model based on a generative adversarial network with the proposed FCA block that integrates a channel attention mechanism with residual 1-D factorized kernel convolutions. The FCA block noticeably improves the performance of our cGAN model as well as other well-known architectures (FCN8, U-Net, etc.). We have run several qualitative and quantitative experiments that show how both channel attention and 1-D residual convolution mechanisms contribute to the segmentation improvement. Those mechanisms boost similar features across all channels or in neighboring regions of the encoder activation maps. As expected, those similar features tend to correspond more robustly with relevant patterns of lesion/non-lesion areas.

Our model is fully automated and fully self-contained, in the sense that we did not use any data augmentation techniques in the training phase or pre-processing steps. The efficiency of the proposed model is assessed on three publicly available skin lesion segmentation challenge datasets: ISBI2016, ISBI2017, and ISIC2018. Our model outperforms several state-of-the-art methods, such as FCN8, U-Net, SegNet, ExB, CUMED, MResNet-Seg, and FrCN, in terms of Dice and IoU metrics.

In future work, we will focus on adapting the proposed model to segment lesions in different organs (e.g., breast, colon and eye) using multi-modal medical images.

REFERENCES

- [1] B. W. Stewart and C. P. Wild, Eds., *World Cancer Report 2014*. 2014. [Online]. Available: https://www.who.int/cancer/publications/WRC_2014/en/
- [2] G. Argenziano, H. P. Soyer, S. Chimenti, R. Talamini, R. Corona, F. Sera, M. Binder, L. Cerroni, G. De Rosa, G. Ferrara, and R. Hofmann-Wellenhof, "Dermoscopy of pigmented skin lesions: Results of a consensus meeting via the Internet," *J. Amer. Acad. Dermatol.*, vol. 48, no. 5, pp. 679–693, May 2003.
- [3] G. R. Day and R. H. Barbour, "Automated melanoma diagnosis: Where are we at?" *Skin Res. Technol.*, vol. 6, no. 1, pp. 1–5, 2000.
- [4] L. Bi, J. Kim, E. Ahn, and D. Feng, "Automatic skin lesion analysis using large-scale dermoscopy images and deep residual networks," 2017, *arXiv:1703.04197*. [Online]. Available: <https://arxiv.org/abs/1703.04197>
- [5] 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.
- [6] Y. Yuan, "Automatic skin lesion segmentation with fully convolutional-deconvolutional networks," 2017, *arXiv:1703.05165*. [Online]. Available: <https://arxiv.org/abs/1703.05165>
- [7] M. Berseeth, "ISIC 2017—Skin lesion analysis towards melanoma detection," 2017, *arXiv:1703.00523*. [Online]. Available: <https://arxiv.org/abs/1703.00523>
- [8] M. Rahman, N. Alpaslan, and P. Bhattacharya, "Developing a retrieval based diagnostic aid for automated melanoma recognition of dermoscopic images," in *Proc. IEEE Appl. Imag. Pattern Recognit. Workshop (AIPR)*, Oct. 2016, pp. 1–7.
- [9] X. Du, W. Zhang, H. Zhang, J. Chen, Y. Zhang, J. C. Warrington, G. Brahm, and S. Li, "Deep regression segmentation for cardiac bi-ventricle MR images," *IEEE Access*, vol. 6, pp. 3828–3838, 2018.
- [10] X. He, H. Zhang, M. Landis, M. Sharma, J. Warrington, and S. Li, "Unsupervised boundary delineation of spinal neural foramina using a multi-feature and adaptive spectral segmentation," *Med. Image Anal.*, vol. 36, pp. 22–40, Feb. 2017.
- [11] V. K. Singh, S. Romani, H. A. Rashwan, F. Akram, N. Pandey, M. M. K. Sarker, S. Abdulwahab, J. Torrents-Barrena, A. Saleh, M. Arquez, M. Arenas, and D. Puig, "Conditional generative adversarial and convolutional networks for X-ray breast mass segmentation and shape classification," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. Granada, Spain: Springer*, 2018, pp. 833–840.
- [12] V. K. Singh, H. A. Rashwan, F. Akram, S. Abdulwahab, N. Maarroof, J. T. Barrena, S. Romani, and D. Puig, "Retinal optic disc segmentation using conditional generative adversarial network," in *Artificial Intelligence Research and Development: Current Challenges, New Trends and Applications* (Frontiers in Artificial Intelligence and Applications), vol. 308, Z. Falomir, K. Gibert, and E. Plaza, Eds. Amsterdam, The Netherlands: IOS Press, 2018, p. 373.
- [13] N. C. F. Codella, D. Gutman, M. E. Celebi, B. Helba, M. A. Marchetti, S. W. Dusza, A. Kalloo, K. Liopyris, N. Mishra, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the 2017 international symposium on biomedical imaging (ISBI), hosted by the international skin imaging collaboration (ISIC)," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 168–172.
- [14] 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)*, Apr. 2018, pp. 859–863.
- [15] J. Fu, J. Liu, H. Tian, Y. Li, Y. Bao, Z. Fang, and H. Lu, "Dual attention network for scene segmentation," 2018, *arXiv:1809.02983*. [Online]. Available: <https://arxiv.org/abs/1809.02983>
- [16] E. Romera, J. M. Álvarez, L. M. Bergasa, and R. Arroyo, "ERFNet: Efficient residual factorized ConvNet for real-time semantic segmentation," *IEEE Trans. Intell. Transp. Syst.*, vol. 19, no. 1, pp. 263–272, Jan. 2018.
- [17] L. Yu, H. Chen, Q. Dou, J. Qin, and P.-A. Heng, "Automated melanoma recognition in dermoscopy images via very deep residual networks," *IEEE Trans. Med. Imag.*, vol. 36, no. 4, pp. 994–1004, Apr. 2017.
- [18] A. Bissoto, F. Perez, V. Ribeiro, M. Fornaciari, S. Avila, and E. Valle, "Deep-learning ensembles for skin-lesion segmentation, analysis, classification: RECOD titans at ISIC challenge 2018," 2018, *arXiv:1808.08480*. [Online]. Available: <https://arxiv.org/abs/1808.08480>
- [19] D. Bisla, A. Choromanska, J. A. Stein, D. Polsky, and R. Berman, "Towards automated melanoma detection with deep learning: Data purification and augmentation," 2019, *arXiv:1902.06061*. [Online]. Available: <https://arxiv.org/abs/1902.06061>

- [20] Z. Mirikharaji, S. Izadi, J. Kawahara, and G. Hamarneh, "Deep auto-context fully convolutional neural network for skin lesion segmentation," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 877–880.
- [21] X. Li, L. Yu, H. Chen, C.-W. Fu, and P.-A. Heng, "Semi-supervised skin lesion segmentation via transformation consistent self-ensembling model," 2018, *arXiv:1808.03887*. [Online]. Available: <https://arxiv.org/abs/1808.03887>
- [22] G. M. Venkatesh, Y. G. Naresh, S. Little, and N. E. O'Connor, "A deep residual architecture for skin lesion segmentation," in *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis*. Springer, 2018, pp. 277–284.
- [23] A. Galdran, A. Alvarez-Gila, M. I. Meyer, C. L. Saratxaga, T. Araújo, E. Garrote, G. Aresta, P. Costa, A. M. Mendonça, and A. Campilho, "Data-driven color augmentation techniques for deep skin image analysis," 2017, *arXiv:1703.03702*. [Online]. Available: <https://arxiv.org/abs/1703.03702>
- [24] S. Vesal, N. Ravikumar, and A. Maier, "Skinnet: A deep learning framework for skin lesion segmentation," 2018, *arXiv:1806.09522*. [Online]. Available: <https://arxiv.org/abs/1806.09522>
- [25] M. M. K. Sarker, H. A. Rashwan, F. Akram, S. F. Banu, A. Saleh, V. K. Singh, F. U. Chowdhury, S. Abdulwahab, S. Romani, P. Radeva, and D. Puig, "SLSDep: Skin lesion segmentation based on dilated residual and pyramid pooling networks," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Granada, Spain: Springer, 2018, pp. 21–29.
- [26] M. Jahanifar, N. Z. Tajeddin, B. M. Asl, and A. Gooya, "Supervised saliency map driven segmentation of lesions in dermoscopic images," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 509–518, 2019.
- [27] 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.
- [28] S. Izadi, Z. Mirikharaji, J. Kawahara, and G. Hamarneh, "Generative adversarial networks to segment skin lesions," in *Proc. IEEE 15th Int. Symp. Biomed. Imag. (ISBI)*, Apr. 2018, pp. 881–884.
- [29] A. Wong, J. Scharcanski, and P. Fieguth, "Automatic skin lesion segmentation via iterative stochastic region merging," *IEEE Trans. Inf. Technol. Biomed.*, vol. 15, no. 6, pp. 929–936, Nov. 2011.
- [30] B. Erkol, R. H. Moss, R. J. Stanley, W. V. Stoecker, and E. Hvatum, "Automatic lesion boundary detection in dermoscopy images using gradient vector flow snakes," *Skin Res. Technol.*, vol. 11, no. 1, pp. 17–26, 2005.
- [31] H. Zhou, X. Li, G. Schaefer, M. E. Celebi, and P. Miller, "Mean shift based gradient vector flow for image segmentation," *Comput. Vis. Image Understand.*, vol. 117, no. 9, pp. 1004–1016, 2013.
- [32] M. Silveira, J. C. Nascimento, J. S. Marques, A. R. S. Marcal, T. Mendonca, S. Yamauchi, J. Maeda, and J. Rozeira, "Comparison of segmentation methods for melanoma diagnosis in dermoscopy images," *IEEE J. Sel. Topics Signal Process.*, vol. 3, no. 1, pp. 35–45, Feb. 2009.
- [33] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jun. 2015, pp. 3431–3440.
- [34] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent.* Munich, Germany: Springer, 2015, pp. 234–241.
- [35] V. Badrinarayanan, A. Kendall, and R. Cipolla, "SegNet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 12, pp. 2481–2495, Dec. 2017.
- [36] V. K. Singh, H. A. Rashwan, S. Romani, F. Akram, N. Pandey, M. M. K. Sarker, A. Saleh, M. Arenas, M. Arquez, D. Puig, and J. Torrents-Barrena, "Breast tumor segmentation and shape classification in mammograms using generative adversarial and convolutional neural network," *Expert Syst. Appl.*, vol. 139, Jan. 2020, Art. no. 112855. [Online]. Available: <http://www.sciencedirect.com/science/article/pii/S0957417419305573>
- [37] 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.
- [38] N. van Noord and E. Postma, "Learning scale-variant and scale-invariant features for deep image classification," *Pattern Recognit.*, vol. 61, pp. 583–592, Jan. 2017.
- [39] S. Baker, D. Scharstein, J. P. Lewis, S. Roth, M. J. Black, and R. Szeliski, "A database and evaluation methodology for optical flow," *Int. J. Comput. Vis.*, vol. 92, no. 1, pp. 1–31, 2011.
- [40] D. Gutman, N. C. F. Codella, E. Celebi, B. Helba, M. Marchetti, N. Mishra, and A. Halpern, "Skin lesion analysis toward melanoma detection: A challenge at the international symposium on biomedical imaging (ISBI) 2016, hosted by the international skin imaging collaboration (ISIC)," 2016, *arXiv:1605.01397*. [Online]. Available: <https://arxiv.org/abs/1605.01397>
- [41] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, D. Gutman, B. Helba, A. Kalloo, K. Liopyris, M. Marchetti, H. Kittler, and A. Halpern, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (ISIC)," Feb. 2019, *arXiv:1902.03368*. [Online]. Available: <https://arxiv.org/abs/1902.03368>
- [42] A. Lalande, M. Garreau, and F. Frouin, "Evaluation of cardiac structure segmentation in cine magnetic resonance imaging," *Multi-Modality Cardiac Imag., Process. Anal.*, pp. 169–215, Dec. 2015.
- [43] A. Chaurasia and E. Culurciello, "LinkNet: Exploiting encoder representations for efficient semantic segmentation," in *Proc. IEEE Vis. Commun. Image Process. (VCIP)*, Dec. 2017, pp. 1–4.
- [44] G. Lin, A. Milan, C. Shen, and I. Reid, "RefineNet: Multi-path refinement networks for high-resolution semantic segmentation," in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, Jul. 2017, pp. 1925–1934.
- [45] H. Li, X. He, F. Zhou, Z. Yu, D. Ni, S. Chen, T. Wang, and B. Lei, "Dense deconvolutional network for skin lesion segmentation," *IEEE J. Biomed. Health Inform.*, vol. 23, no. 2, pp. 527–537, Mar. 2019.
- [46] M. Al-masni, M. Al-antari, P. Rivera, E. Valarezo, G. Gi, T. Kim, H. Park, and T. Kim, "Automatic skin lesion boundary segmentation using deep learning convolutional networks with weighted cross entropy," Tech. Rep., 2018.
- [47] L. Bi, D. Feng, and J. Kim, "Improving automatic skin lesion segmentation using adversarial learning based data augmentation," 2018, *arXiv:1807.08392*. [Online]. Available: <https://arxiv.org/abs/1807.08392>
- [48] R. C. Hardie, R. Ali, M. S. De Silva, and T. M. Kebede, "Skin lesion segmentation and classification for ISIC 2018 using traditional classifiers with hand-crafted features," 2018, *arXiv:1807.07001*. [Online]. Available: <https://arxiv.org/abs/1807.07001>



VIVEK KUMAR SINGH received the bachelor's degree in electronics and communication engineering from Gautam Buddha Technical University, Lucknow, India, and the master's degree in signal processing from the Sam Higginbottom Institute of Agriculture, Technology and Sciences (SHIATS), Allahabad, India, in 2011 and 2014, respectively. He is currently pursuing the Ph.D. degree with the Intelligent Robotics and Computer Vision Group, Department of Computer Engineering and Mathematics Security, Rovira i Virgili University, Tarragona, Spain, where he has been a Predoctoral Researcher, since December 2016. His research interests include image processing, medical image analysis, machine learning, deep learning, and signal processing.



MOHAMED ABDEL-NASSER received the Ph.D. degree in computer engineering from Universitat Rovira i Virgili, Spain, in 2016, where he is currently a Postdoctoral Researcher and also an Assistant Professor with the Electrical Engineering Department, Aswan University, Egypt. He has participated in several projects funded by the Government of Spain and the Government of Egypt. He has published more than 50 articles in international journals and conferences. His research interests include the application of machine learning and deep learning to several real-world problems, including medical image analysis, smart grid analysis, and time-series forecasting.



HATEM A. RASHWAN received the B.S. degree in electrical engineering and the M.Sc. degree in computer science from South Valley University, Aswan, Egypt, in 2002 and 2007, respectively, and the Ph.D. degree in computer vision from the Rovira i Virgili University, Tarragona, Spain, in 2014. From 2004 to 2009, he was with the Electrical Engineering Department, Aswan Faculty of Engineering, South Valley University, Egypt, as a Lecturer. In 2010, he joined the Intelligent

Robotics and Computer Vision Group, Department of Computer Science and Mathematics, Rovira i Virgili University, where he was a Research Assistant with the IRCV Group, in 2014. From 2014 to 2017, he was a Researcher in vortex with IRIT-CNRS, INP-ENSEEIH, University of Toulouse, Toulouse, France. Since 2017, he has been a Beatriu de Pinós Researcher with DEIM, Universitat Rovira i Virgili. His research interests include image processing, computer vision, pattern recognition, and machine learning.



FARHAN AKRAM received the B.Sc. degree in computer engineering from the COMSATS Institute of Information Technology, Islamabad, Pakistan, in 2010, the M.Sc. degree in computer science with a major in application software from Chung-Ang University, Seoul, South Korea, in 2013, and the Ph.D. degree in computer engineering and mathematics from the Rovira i Virgili University, Tarragona, Spain, in July 2017. From October 2017 to July 2019, he was a Postdoctoral

Research Fellow with the Imaging Informatics Division, Bioinformatics Institute, A*STAR, Singapore. In August 2019, he joined the Department of Electrical and Computer Engineering, Khalifa University of Science and Technology, Abu Dhabi, United Arab Emirates, where he is currently a Postdoctoral Research Fellow. His current research interests include medical image analysis, image processing, computer vision, and deep learning.



NIDHI PANDEY received the bachelor's degree in medicine and surgery from India in June 2015. She is currently pursuing the master's degree in nervous system with the Department of Medicine and Health Sciences, Rovira i Virgili University. From 2015 to October 2016, she has been a Junior Resident Doctor. Her research interests include oncology, neuroscience, and radiology.



ALAIN LALANDE is currently an Associate Professor in biophysics and medical image processing with the University of Burgundy, Dijon, France, where he is also with the ImViA Laboratory (Image and Artificial Intelligence). His research interests concern the medical imaging domain, initially mainly the imaging of the cardiovascular system (particularly MRI), and cover the topics of image acquisition, image post-processing, and protocol design. His current research interests

include more diversified with also augmented reality in robotic surgery, post-processing of CT-scans in LDR prostate cancer brachytherapy, and artificial intelligence in medical imaging.



BENOIT PRESLES received the Ph.D. degree in image processing and analysis from the École des Mines de Saint-Étienne in Saint-Étienne, France, in 2011. After specializing in medical imaging, he was a Postdoctoral Researcher with the Centre de Recherche en Acquisition et Traitement de l'Image pour la Santé (CREATIS) Laboratory, Lyon, France, and also with the Translational Imaging Group (TIG), Centre for Medical Image Computing (CMIC), University College

London (UCL), London, U.K. In 2016, he joined the Imagerie et Vision Artificielle (ImViA) Laboratory, Dijon, France, as a Lecturer. His research interests include image registration, simulation, and artificial intelligence.



SANTIAGO ROMANI received the M.S. degree in computer science from the Technical University of Catalonia, Spain, in 1991, and the Ph.D. degree in computer science from the Technical University of Catalonia, in 2006. Since 1991, he has been an Associate Professor of computer engineering with the University Rovira i Virgili, Tarragona, Spain. His research interests include image analysis, color, and other perceptual cues representation, and computer applications based on such representation.



DOMÈNEC PUIG received the M.S. and Ph.D. degrees in computer science from the Polytechnic University of Catalonia, Barcelona, Spain, in 1992 and 2004, respectively. In 1992, he joined the Department of Computer Science and Mathematics, Rovira i Virgili University, Tarragona, Spain, where he has been a Professor, since 2017. He has been the Head of the Intelligent Robotics and Computer Vision Group, Rovira i Virgili University, since 2006. His research interests include

image processing, texture analysis, perceptual models for image analysis, scene analysis, and mobile robotics.

...