Sohag Journal of Sciences
An International Journal

An Efficient skin lesion segmentation using deep fully convolutional neural network with gradient skin images

Rania R. Mohamed[1] , *Saleh K. Aly*[2], *and Mahmoud M. Abdel-Aty*[1]

[1]Mathematics Department, Faculty of Science, Sohag University, Sohag, 82524, Egypt

[2]Department of Electrical Engineering, Faculty of Engineering, Aswan University, Aswan, 81542, Egypt

Received: 21 June 2022, Revised: 2 Aug. 2021, Accepted: 18 Aug. 2022.
Published online: 1 Sept. 2022

Abstract: Skin lesion semantic segmentation is a vital process that aims to identify each pixel in the input image whether it belongs to the foreground (lesion skin) or background (normal skin). Skin lesion image segmentation is an essential step in the medical image analysis domain for use in radiotherapy to enhance diagnostic radiology. Misclassified border pixels cause a significant reduction in the global accuracy because they maybe belong to the foreground or background. The aim of this paper is to improve the skin segmentation results at border pixels by building a deep fully convolutional network fed with gradient skin images instead of traditional color images. The proposed segmentation network produces a binary predicted output image with efficient inference at all image pixels while giving extra attentions to border pixels. The appropriate gradient components of the input skin image are employed to train one of the famous deep convolutional neural networks called U-Net with some modifications. Dice loss function is utilized to train the network instead of cross entropy network in order to improve the performance segmentation results especially in the border pixels. Several experiments are conducted using the ISIC 2018 dataset to evaluate the performance of the proposed network compared to other state-of-the-art approaches.

Keywords: Color spaces, gradient U-Net, spatial attention unit, Skin lesion

.

1 Introduction

Recent medical statistics indicate a significant increase in the number of people diagnosed with skin cancer , , which poses a threat to their lives, especially if these infections are discovered in advanced stages. Therefore, early detection of this infection contributes significantly to avoid serious consequences and thus preserve the lives of patients. However, the manual diagnosis of such cases is considered ineffective method because it requires long time and high medical experience in this field. Therefore, researchers have resorted to developing a Computer-Aided Diagnostic (CAD) system capable of accomplishing this task [**1, 2, 3, 4**], but in a faster time with higher accuracy. Automatic segmentation for lesions is an essential step as it prepares the data for the lesion classification stage and help the dermatologist in making decision.

The skin lesion segmentation methods face many obstacles [5, 6, 7] that hinder the performance of the segmentation process with high accuracy. Investigating images in popular skin lesion databases[1] reveals that there is a clear color contrast between the portions that represent the lesions in the images. In addition to the differences in the colors of the portions that represent the background, there are also differences in the texture, shape, position, and borders of the lesions. Moreover, there are also some variations related to the patient himself, such as color, texture, position, and size of the skin lesion. It was also found that the database images contain some external obstacles such as body hair, air bubbles, ruler marks, markers sign, uneven shading, some dark corners, ink marks, blood vessels, and colored lights**.** Recent medical statistics indicate a significant increase in the number of people diagnosed with skin cancer , , which poses a threat to their lives, especially if these infections are discovered in advanced stages. Therefore, early detection of this infection contributes significantly to avoid serious consequences and thus preserve the lives of patients. However, the manual diagnosis of such cases is considered ineffective method because it requires long time and high medical experience in this field. Therefore, researchers have resorted to

* Corresponding author E-mail: rania abdrabou@science.sohag.edu.eg

developing a Computer-Aided Diagnostic (CAD) system capable of accomplishing this task [**1, 2, 3, 4**], but in a faster time with higher accuracy. Automatic segmentation for lesions is an essential step as it prepares the data for the lesion classification stage and help the dermatologist in making decision. shows samples of variations in skin lesions.

In an attempt to overcome the problems of skin lesion related to color contrast variations and irregularity of the lesion boundaries, we applied a preprocessing operation to the skin images before using them in the training of the proposed network architecture. The pre-processing operation computes the gradient representation of the RGB skin image and feeds it to the proposed network. The suggested network architecture, called gradient U-Net (GU-Net) is based on the well-known U-Net model. The gradient U-Net architecture (GU-Net) uses spatial attention units in the links between the encoder and decoder paths to focus on the location of the lesions in the images. The use of spatial attention units improves the network's ability to determine the correct boundaries of the lesions. In addition, the Atrous Spatial Pyramid Pooling (ASPP) block is used in the bottleneck stage to capture multi-scale skin lesion representation information. The following is a list of contributions used to address the obstacles faced in segmentation of skin lesion

1. Exploiting gradient skin images instead of color images to feed the proposed modified U-Net model.

2. Using spatial attention units in the links that pass form encoder to the decoder paths.

3. Using spatial Atrous convolution mechanism in the bottleneck stage of the suggested architecture.

4. Experiments were carried out on the ISIC 2018 dataset and the results are compared with the recent methods.

The organization of the paper is as follows. Section 2 outlines some related works. In Section 3, the original and the modified U-Net model is explained. The experimental results are presented in Section 4. Finally, the conclusion is presented in Section 5.

2 Related work

In this section, we review methods related to the skin segmentation field and based mainly on deep learning architectures and edge detection methods.

Figure 1: Example of obstacles in skin lesion image segmentation

2.1 Deep learning techniques for skin lesion segmentation

In 2015, Ronneberger et al. work [**8**] is considered an important achievement in the field of medical image segmentation. They presented a network with training strategy based on data augmentation. This network consists of two paths (1) encoder path to obtain spatial feature information and (2) similar decoder path to find exact object localization. Their network is fast and the structure outperforms the previous best method and achieved the best results in the 2015 challenge of ISBI cell tracking in electron microscopic stacks and in transmitted light microscopy images (phase contrast and DIC) categories. Later, U-Net network become the model that has been relied on by many methods recently.

Abraham et al. [**9**] introduced a focal Tversky cost function to increase the balance between precision and recall semantic segmentation performance metrics. They developed an attention U-Net model [**10**] by merging each

scale of the input image pyramid into model construction. Their method outperformed the baseline U-Net in Dice scores and introduced low standard deviations for balanced precision-recall scores. Liu et al. [11] proposed an efficient skin lesion segmentation based on the improved U-net model and their method achieved the latest performance in the skin lesion segmentation task. Hashemi et al. [12] trained a 3D U-net with an asymmetric similarity loss layer. They applied wide-overlapping patches of the image as inputs for extrinsic and intrinsic data augmentation. They succeed to ease the data imbalance problem and reach a much better trade-off between recall and precision. Dash et al. [13] suggested a cascaded deep convolutional neural network based automated CAD procedure for psoriasis segmentation, recognition, and severity evaluation. The adjusted U-Net and modified VGG-16 models are used for the implementation and training of segmentation and classification tasks, respectively. Their modified VGG model produced higher classification accuracy using lower trainable parameters and shown superior performance for both binary and multiclass classification methods. In [14], authors proposed three variants of the U-Net model using multiple encoders and a single decoder architecture. Each encoder is fed with different color space of the input image to handle skin lesion color variation problem. Later, this work was extended to combine gradient and color information using the dual gradient-color U-Net (DGCU-Net) [15] architecture.

2.2 Edge detection-based methods for skin lesion segmentation

In this part, we will consider in some detail relevant works that adopted the edge detection technique in their proposed approaches. Barcelos et al. [16] presented a segmentation approach that merges non-linear diffusion equations and the Canny edge detector to automatically identify the boundary of the skin lesion. They obtained smooth skin image while keeping the boundaries of interest to create an accurate edge detection. The experimental results showed that the suggested approach is effective and can be implemented for images of skin lesions on grayscales and color. Their method localized lesion boundaries automatically from images involving hair and noise. Sheykhahmad et al. in [17] introduced an efficient method to detect borderline of skin lesion image. They introduced a narrative approach based on image processing that mixes edge detection and the thresholding procedure for skin lesions detection. Their technique produced efficient discrimination of skin lesions and the experimental results indicate the power of the suggested model. Yasmin and Sathik [18] have developed advanced iterative segmentation methods that use a canny edge detector to detect the borderline of certain skin lesions to accelerate the early detection of malignant melanomas. Their work is compared, and the experimental results indicate good border detection of noisy skin lesions by their suggested advanced iterative segmentation method utilizing the canny detector. Their model was extremely reliable compared to other segmentation methods that adopted Canny detector. Pereir et al. [19] has focused on segmentation of skin lesion images, using both histogram-stretching and clustering methods to reduce their limitations. A gradient-based technique was invented for optimized thresholding and ROI border quality parameter. The obtained segmentation results indicated that this approach is very appropriate for delineating parts of the related lesion that comprise a large in recognizing boundaries of the lesion and was more powerful for artifacts in the image. They compared their designed method with dermatologist's manual border drawings, the result was 87.7% for the average dice.

3 Methodology

3.1 Skin lesion image pre-processing

The aim of this work is to enhance the U-Net architecture segmentation performance, especially at the border pixel regions. To achieve this goal, we applied an image gradient conversion as a pre-processing step to the input images fed to the modified U-Net architecture. This conversion is described as follows:

An image gradient is a term that describes a continuous shift from light to dark, or dark to light for a single color, or from one color to another in an image. This shift refers specifically to the intensity. This means that a single color will go from maximum intensity to minimum intensity. For images of skin lesions, this change in color or intensity is important for lesion borderline detection. Usually, two-directional images can be generated to capture horizontal and vertical changes for any image. When talking about the image gradient, several kernels can be used to approximate the gradient vector. The gradient of any image has two components, which are the horizontal Gx and vertical Gy gradients. Gx is the partial derivative in the X direction, while Gy is the partial derivative in the Y direction. The gradient of skin image can be computed by convolving image I with two kernels Kx and Ky. These kernels are used to calculate Gx and Gy gradients by applying the following equation:

$$G_x = I \times K_x \quad \textbf{(1)}$$

$$G_y = I \times K_y \quad (2)$$

The kernels that are used to approximate these two components include Prewitt, Sobel, the central kernel, and the intermediate kernel. The Prewitt kernel was first proposed in 1970 and its values for the Gx component as:

$$K_x = \begin{bmatrix} 1 & 0 & -1 \\ 1 & 0 & -1 \\ 1 & 0 & -1 \end{bmatrix} \quad (2)$$

And for the *Gy* component as

$$K_y = \begin{bmatrix} 1 & 1 & 1 \\ 0 & 0 & 0 \\ -1 & -1 & -1 \end{bmatrix} \quad (3)$$

This type of kernel deal better with noise because both of these kernels are normal one-dimensional kernels that approximate the derivative in a certain direction convolved with a smoothing kernel to smooth the image. Another kernel is Sobel that was proposed in 1980. It is the most famous kernel and was used in most edge-detector applications. Sobel kernel also based on the same idea like Prewitt kernel (derivative part and smoothing part) but with different values.

$$K_x = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix} \quad (4)$$

And for the Gy component as

$$K_y = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix} \quad (5)$$

The smoothing part in the Sobel kernel is the Gaussian kernel, and this explains the superior performance of the Sobel kernel. The central operator or 'central difference gradient' is one of the gradient operators where a gradient of a pixel is computed as a weighted difference of neighboring pixels. Using this operator, Gx is given by:

$$\mathbf{G_x} = \mathbf{I(x+1)} - \frac{\mathbf{I(x-1)}}{\mathbf{2}} \quad (6)$$

while, Gy equals

$$\mathbf{G_y} = \mathbf{I(y+1)} - \frac{\mathbf{I(y-1)}}{\mathbf{2}} \quad (7)$$

Likewise, Intermediate or 'Intermediate difference gradient. The gradient of a pixel is calculated as the difference between an adjacent pixel and the current pixel. So the vertical gradient is given by the following formula:

$$G_y = I(y+1) - I(y-1) \quad (8)$$

while, the horizontal gradient Gx equals

$$G_x = I(x+1) - I(x-1) \quad (9)$$

When dealing with a gradient image, sometimes we need more gradient information like, how quickly the image is changing? and in which direction the image is changing most

high-level features, and in the decoder, the size of the feature map rapidly? To answer these questions, we should calculate the gradient magnitude Gm and direction Gdir for the image. Using Gx and Gy, the magnitude of the gradient Gm can be written as the following formula:

$$G_m = \sqrt{G_x^2 + G_y^2} \quad (10)$$

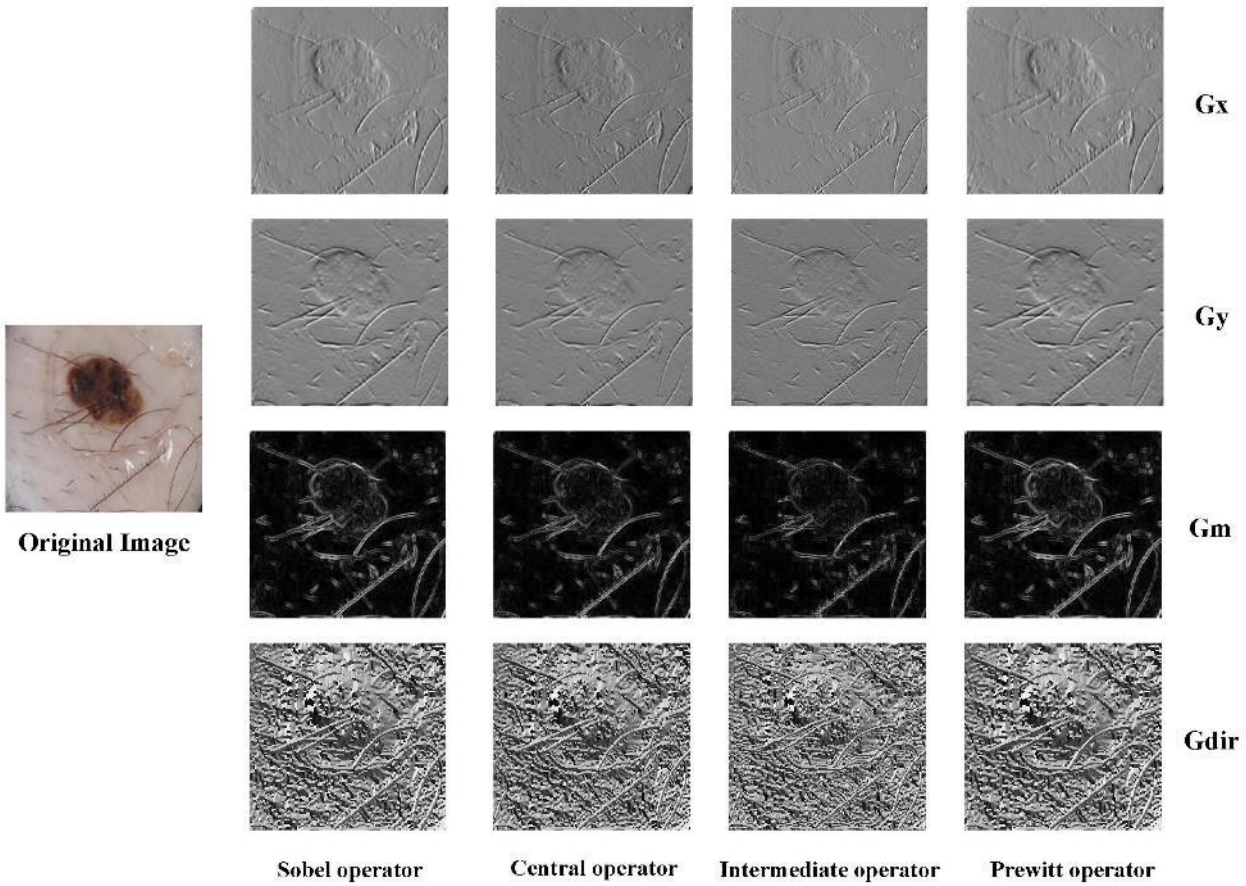
and the gradient direction Gdir

$$G_{dir} = tan^{-1}\left(\frac{G_y}{G_x}\right) \quad (11)$$

Figure 2 shows the gradient components using different gradient operators for an image of skin lesion.

Proposed Gradient U-Net (GU-Net) Architecture

The structure of original U-Net model is a symmetric design and from that its name was derived (i.e., U letter). It consists of two corresponding paths, an encoder path, which extracts the features, and the decoder path, which builds up a mask from the features. These two paths are interconnected with each other to enrich features used for classification. Both paths have four convolution blocks. In the encoder subnetwork, every block involves two convolution layers that use the ReLU activation function and a max-pooling layer that shrinks the feature map size by half. In the decoding procedure, every block is initiated with a deconvolution layer to double the dimensions of each feature map while reducing the collection of feature maps by half. Going through the encoder path, the Going through the encoder path, the number of filters is doubled; conversely, the number of filters shrinks as the decoder reaches the prediction layer. Therefore, in the U-Net, there is a concatenation layer that joins feature maps of the same volume in the two paths. This merger contributes to obtain extra details that are required for perfect segmentation. shows the main structure of our designed GU-Net. The encoder path progressively reduces the size of the feature map to derive gradually increases to fit the mask size. However, the suggested structure involves some improvements in the original U-Net to establish a better convenient for detecting and extracting the boundary pixels that represent the lesion edge, hence improving its performance and accuracy. The proposed GU-Net network accepts the pre-processed input image and passes it through the four levels of the encoder, where the number of applied filters is duplicated at each level, starting from 64, 128... until reaches 512 at the end of last (fourth) level of the encoder. The resulting feature maps of each level pass through a spatial attention unit to focus on the spatial features of the lesion images. Then, the resulted maps are concatenated with their corresponding feature maps on the decoder path. The final level of encoder feature maps passes through the Atrous Spatial Pyramid Pooling (ASPP) block with different rates before reaching to the first level of the decoder where there is a deconvolution layer at the beginning of each block. After each deconvolution layer, two convolution layers are implemented to cut down the number of feature maps. The last layer is a convolution layer with kernel 1×1 and the total number of feature maps The last layer is a convolution layer with kernel 1×1 and



2.1 Figure 2: Example of skin lesion image gradient components (Gx, Gy, Gm, and Gdir) using Sobel kernel

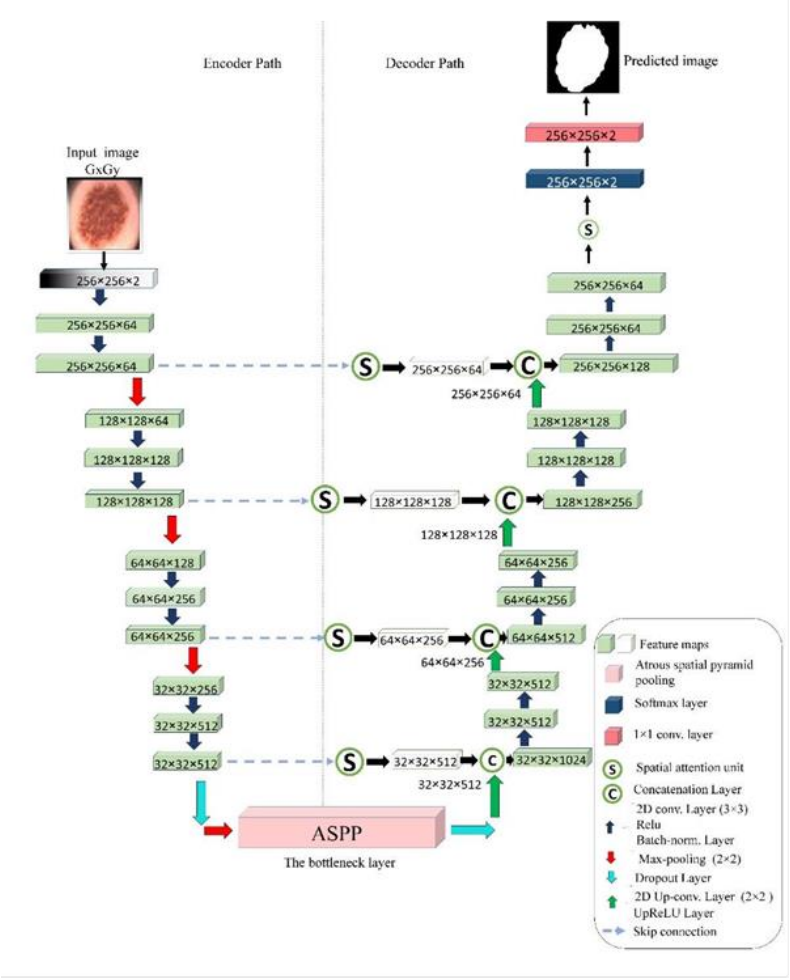
the total number of feature maps equals 2, which is proper for two-class segmentation process. In the following, we will summarize our modification on the traditional U-Net.

- At the beginning, the input image to the network should be pre-processed by calculating the gradient components (Gx, Gy) of the image.
- Replacing the bottleneck stage in the traditional U-Net by five successive atrous spatial pyramid pooling layers with different dilated rates.
- Adding a spatial attention unit before each concatenation layer in each level of the connection between encoder and decoder paths.
- Finally, adding a spatial attention unit before the soft- max layer at the end of decoder path.

3.2.1 Spatial Attention Unit

At the lower layers on the encoder path of the U-Net, we have more spatial information which provides the context features of the image. Thus, when the concatenation is done through the skip connection, it combines the spatial information from the down-sampling path with the corresponding up-sampling features path. This mechanism is very helpful for providing rich spatial information. Therefore, including spatial attention units before doing concatenation is the most suitable location to get the best context information extraction.

The spatial attention unit was presented by Bahdanau et al. [21] to enhance the representation strength of context features. This is done by passing the encoder feature maps $F^{c\times h\times w}$to max-pooling layer to get $F_{max}^{1\times h\times w}$ and the average pooling layer to get $F_{avg}^{1\times h\times w}$. Then a concatenation is applied to both of the pooled features Fmax, avg before sending them to the convolution layer that is followed by the sigmoid activation function. The final spatial attention features FSFM are obtained by applying the element-wise multiplication between the output of the sigmoid function SAM and the encoder feature maps entered F as shown in Using spatial attention units helps to create highly accurate and logical semantic features and pick out the most relevant



information to feed into the decoder path. The following equations describe the steps used to calculate spatially attentioned feature maps:

$$F_{max,avg} = concat\big(maxpool(F), averagepool(F)\big) \quad (12)$$

that we use the same number of parameters and the same amount of computations while keeping the data resolution of the output. The first presentation of the idea of atrous convolution or dilated convolution was in [22, 23]. Traditional Convolution is similar to the Atrous convolution with a rate ratio of r=1. Figure 5 shows how the atrous convolution technique appears at various rates. ASPP block consists of successive atrous convolution layers with

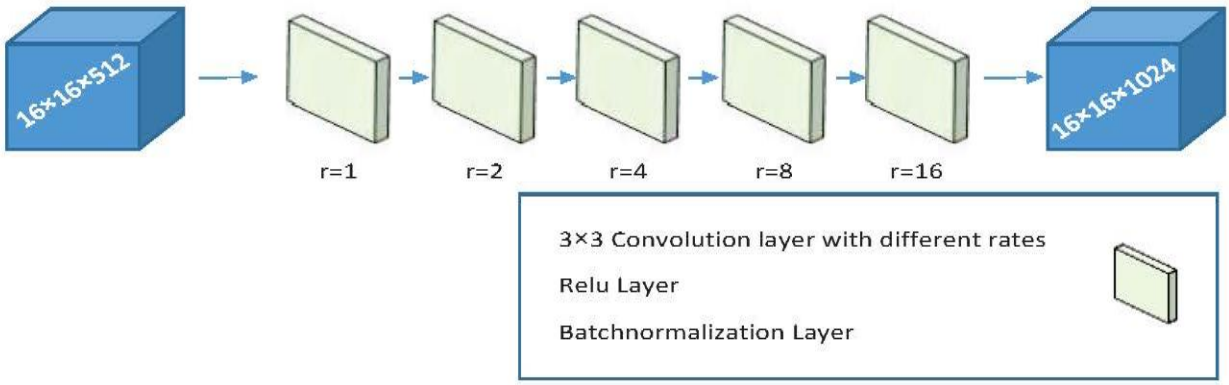


Figure 3: The structure of ASSP used in the proposed GUNet.

$$SAM = sigmoid(ReLU(Conv.(F_{max,avg}))) \quad (13)$$

$$FSFM = F \otimes SAM \quad (14)$$

3.2.2 Atrous Spatial Pyramid Pooling (ASPP)

In deep learning, it is necessary to transform the input data and reduce the dimensionality so that the model can learn more high-level features. Most of the convolutional neural networks used for segmentation/classification tasks perform multilevel convolution with a stride operation to create high-level features. Another way of down-sampling feature maps is by using max grouping with a window size 2×2 and stride 2. However, subsequent pooling processes cause the feature map to lose a lot of information. To avoid this problem, a dilated convolution using Atrous filters can be employed instead of regular convolutions. In regular convolution, we take a patch of an image that may be of size 3×3 and then multiply it by the kernel of the same size to obtain one output value in the output image. But with the dilated convolution, the image that we get is equally sparse. This is done by putting zeros in-between pixels of the kernel, which leads to an increase in the field of view. Actually, increasing the field of view is useful to capture a larger context of the image if we want to accomplish classification or segmentation of large objects. In the dilated convolution, we stack multiple 3×3 convolutions with various dilation rates. Therefore, with the dilated convolution, we can adjust the field of view with the dilation rate parameter. Increasing the dilation rate will add more zeros in the kernel and hence increase its effective field of view. One advantage of using dilated convolution is different rates connected in serial. Figure 6 displays the structure of the ASSP used in the bottleneck stage in our proposed GU-Net. It consists of five blocks connected in a serial manner; each block has a 3 × 3 convolution layer with different rates followed by a ReLU activation function and batch normalization layer with a number of filters equal to 1024. the rate r takes the values 1,2,4,8,16 respectively to output feature map of size 16 × 16 × 1024.

3.2.3 Dice loss function

Dice coefficient is a very important statistical metric to evaluate segmentation results with a large set of images. It is defined as 2 multiplied by the overlapped area between predicted and ground truth mask images divided by the total number of pixels in both predicted mask images (X) and ground truth (Y). Mathematically it is given by:

$$Dice = \frac{2|X \cap Y|}{|X|+|Y|} \quad (15)$$

where the $|X \cap Y|$ is the common pixels between X and Y. While X + Y are all pixels in X and all pixels in Y. The first version of Dice coefficient was Sørensen–Dice coefficient. The metric value ranges from 0 which indicate no overlap to 1 which indicates perfect overlap.

When using Boolean data and TP, FP, FN metrics, the equation 16 can be reformulated as:

$$Dice = \frac{2TP}{2TP+FP+FN} \quad (16)$$

where FN, TP, and FP denote false negatives, true positives,

and false positives, respectively. From Eq. $Dice = \frac{2|X \cap Y|}{|X|+|Y|}$ **(15)**, we conclude that FP and FN have equal weights for the precision and recall trade-off. The dice loss function is calculated from the following formula:

$$L_{Dice} = 1 - Dice \quad \textbf{(17)}$$

$$Dice = \frac{2\sum_{i=1}^{N}\sum_{c=1}^{C}(X_{ic}Y_{ic})}{\sum_{i=1}^{N}\sum_{c=1}^{C}X_{ic}^2+\sum_{i=1}^{N}\sum_{c=1}^{C}Y_{ic}^2} \textbf{(19)}$$

$$L_{Dice} = 1 - \frac{2\sum_{i=1}^{N}\sum_{c=1}^{C}(X_{ic}Y_{ic})}{\sum_{i=1}^{N}\sum_{c=1}^{C}X_{ic}^2+\sum_{i=1}^{N}\sum_{c=1}^{C}Y_{ic}^2} \quad \textbf{(18)}$$

In which *Yic* and *Xic* represents *i* values of the expected and actual values at a specific class c, respectively. Dice loss is used efficiently to address the data imbalance situations between the foregrounds and background pixels and achieved good results in the semantic segmentation tasks. So, we adopted it as a loss function in our proposed model and used the available Matlab function2.

[4]https://www.mathworks.com/help/vision/ref/nnet.cnn.layer dicepixelclassificationlayer.html

[5]https://isic-challenge-data.s3.amazonaws.com/2018/ISIC2018 Task1-2 Training Input.zip

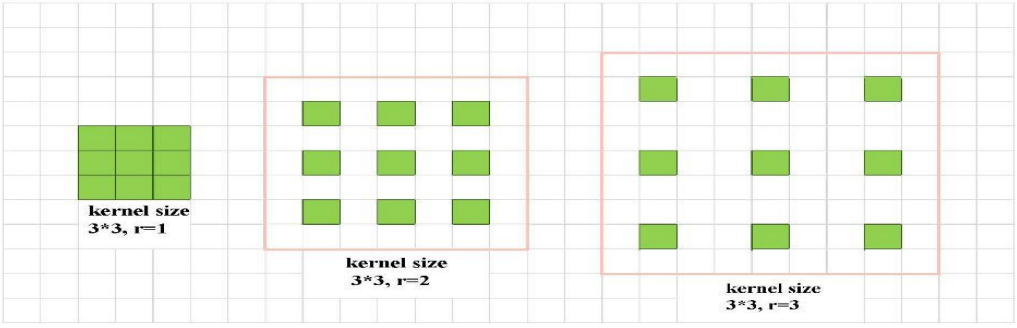


Figure 4: Example of three 3×3 dilated convolutional kernel

4 Experimental results

This section explains in details the results of the experiments conducted using our introduced (GU–Net) on ISIC 2018 dataset and the comparisons with other recent algorithms that used the same dataset to make fair comparisons.

4.1 Dataset

The performance of GU-Net model is verified using ISIC 2018 dataset[3] that was produced by the International Skin Imaging Collaboration (ISIC) as a large-scale dermoscopy images dataset. It contains a 2594 JPG RGB images with ranged dimensions from (576×768) to (6748×4499). We separated the available training set images into (2076) images (80%) for training and (518) images 20% for testing with size (256×256). Figure 5 shows some sample images from ISIC 2018 dataset.

4.2 Implementation details

In our implementation, we used MATLAB software package running on desktop with 16 GB RAM, Intel Core i7 processor, and NVIDIA RTX 2080 Ti GPU card with 11 GB memory. The implemented network structure was the same as explained in the previous section. The training parameters used to train GU-Net model include initial learning rate = 0.05, 4 for mini-batch size, SGDM optimizer with 0.9 momenta, and 30 epochs using the dice loss function, without data augmentation. All these parameters were chosen after performing various experiments to find the optimal values for each parameter.

4.3 Evaluation metrics

Six evaluation metrics are used to evaluate the performance of the introduced GU-Net, including Dice Similarity Coefficient score (Dice) (16, Sensitivity (Se), Specificity (Sp), Accuracy (Acc), Jaccard Coefficient (Jac), and Area under curve (AUC). Sensitivity measures the capability of a model to accurately determine lesions. Specificity measures the ability of a model to correctly identify the background without the lesions. Accuracy measures the ability of the model to differentiate infected pixels from healthy pixels correctly. Jaccard Coefficient measures the similarity for the predicted images and the corresponding ground truth, with a range from 0% to 100%. AUC measures the area under the receiver operating characteristic (ROC) curve. These metrics are defined with the following formulas:

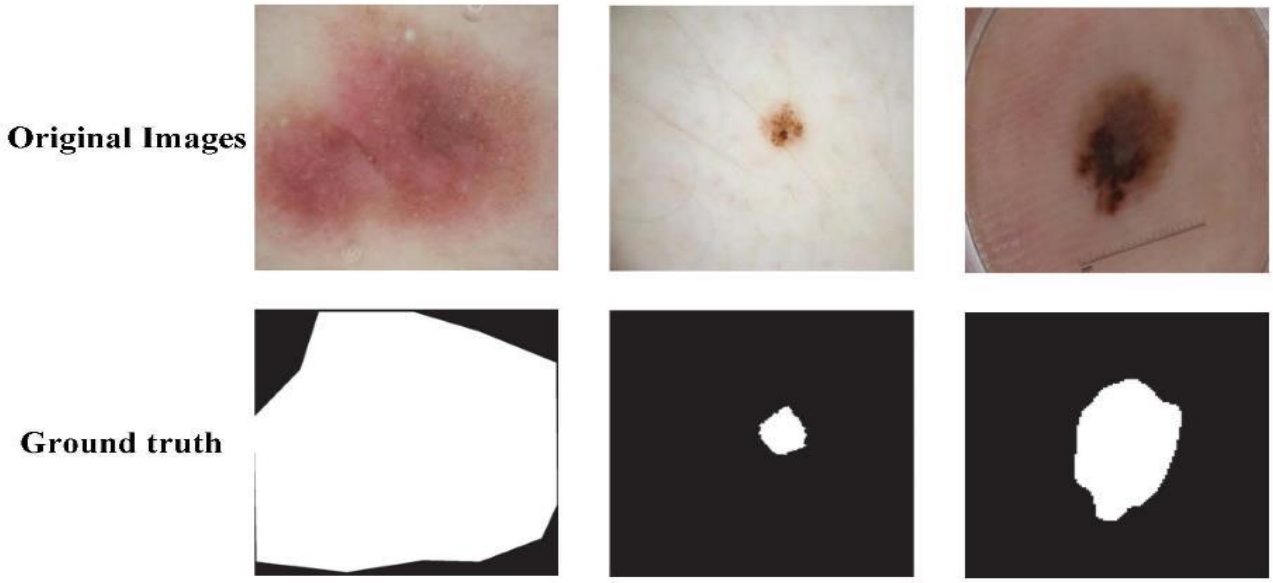


Figure 5: Sample skin images from ISIC 2018 dataset and their corresponding ground truth

$$Jac(Weighted\ IOU) = \frac{TP}{TP+FP+FN} \quad (19)$$

$$ACC = \frac{TP+TN}{TP+FP+FN+TN} \quad (20)$$

$$Sp = \frac{TN}{FP+TN} \quad (21)$$

$$Se = \frac{TP}{TP+FN} \quad (22)$$

$$AUC = \frac{Se+Sp}{2} \quad (23)$$

4.4 Studying the effect of changing the input gradient component on the proposed GU-Net:

Gradient component	Acc%	TPR%	TNR%	Dice%	Jacc%	AUC%
Gx	92.87	81.25	96.64	84.81	86.82	88.95
Gy	93.20	84.08	96.16	85.84	87.46	90.12
GxGy	93.75	83.04	97.23	86.68	88.32	90.13
Gm	91.88	76.33	96.92	82.15	85.04	86.62
Gdir	89.82	66.66	97.33	76.23	81.40	81.99
GmGdir	90.83	70.14	97.54	78.94	83.12	83.84

In this experiment, we present the results obtained by altering the components of the gradient image on the designed GU-Net. We conduct this experiment by applying

varied combination of Sobel kernel gradient components including Gx, Gy, Gm, Gdir, GxGy, and GmGdir. Table 1 demonstrates the results of the evaluation of each gradient component and the results of the combination of two complementary components. Results indicate that combining Gx and Gy components performs better than the alternative components in all evaluation metrics.

Table 1: Segmentation results by changing the input gradient component of the proposed GU-Net

4.5 Studying the effect of changing Gradient kernel on the proposed GU-Net:

This experiment examines the effect of varying gradient kernel on the performance of GU-Net model. This experiment is conducted by applying GxGy gradient images on various kernel types. shows the evaluation results for each kernel. The results indicate that the Prewitt kernel gives the best results in comparison to other kernels. These results conclude that the combination of GxGy gradients of the Prewitt kernel improves the segmentation results of skin lesion.

Table 2: Evaluation metrics with different kernels of GxGy gradient images

4.6 Comparison with other state-of-the-art Approaches

The performance of several state-of-the-art methods is compared with our proposed GU-Net method. shows the results of various evaluation metrics for our proposed method. The table demonstrates the superiority of our results over other compared methods in the Jacc index, where it achieved 88.50%. Other evaluation metrics achieve competitive results with state-of-the-art approach in [24].

Table 3: comparison with some recent methods

Method	Year	Acc%	TPR%	TNR%	Dice%	Jacc%
U-Net [8]	2015	90.70	73.12	98.17	82.42	70.10
AttU-Net [25]	2019	93.76	86.00	98.26	85.66	77.64
ResUNet++ [26]	2019	93.82	87.35	97.21	85.36	77.21
FTL [9]	2018	94.12	87.54	96.32	86.93	78.25
ERU [27]	2020	94.35	90.32	96.92	88.12	80.56
[28]	2020	93.24	90.72	95.88	88.07	81.13
CKD-Net [29]	2021	94.92	90.55	97.01	87.79	80.41
FAT-Net [24]	2021	95.78	91.00	96.99	89.03	82.02
GU-Net(GxGy)	2022	93.81	85.35	96.56	87.11	88.50

5 Conclusion

In this work, we introduced a novel gradient model based on the U-Net architecture called (GU-Net) for skin lesion image segmentation. GU-Net takes full advantage of the characteristics of the U-Net architecture to extract the high-level features of the image. Additionally, selecting the optimum gradient kernel and components (GxGy) and the modifications applied to the traditional U-Net structure have enhanced the performance of the network. Proposed GU-Net can distinguish the boundaries of the lesions within images well and differentiate them from the healthy pixels. Our work achieved state-of-the-art performance on the ISIC 2018 dataset in Jaccard coefficient. The use of gradient components of the image and the use of spatial attention units helped GU-Net to make network better focusing on the borderline of skin lesions.

References

[1] P. Schmid-Saugeona, J. Guillodb, and J.-P. Thirana, "Towards a computer-aided diagnosis system for pigmented skin lesions," *Computerized Medical Imaging and Graphic*s, 2003, vol. 27, no. 1, pp. 65–78.

[2] N. Razmjooy, M. Ashourian, M. Karimifard, V. V. Estrela, H. J. Loschi, D. Do Nascimento, R. P. Franc¸a, and M. Vishnevski, "Computer-aided diagnosis of skin cancer: a review," Current Medical Imaging, vol. 16, no. 7, pp. 781–793, 2020.

[3] A. Masood and A. Ali Al-Jumaily, "Computer aided diagnostic support system for skin cancer: a review of techniques and algorithms," International journal of biomedical imaging, vol. 2013, 2013.

[4] S. Pathan, K. G. Prabhu, and P. Siddalingaswamy,

Filter	Acc%	TPR%	TNR%	Dice%	Jacc%	AUC%
Sobel	**93.75**	**83.04**	**97.23**	**86.68**	**88.32**	**90.13**
Central	**92.85**	**88.68**	**94.21**	**85.88**	**87.05**	**90.45**
Inter.	**92.98**	**87.16**	**94.87**	**85.87**	**87.19**	**90.01**
Prewitt	**93.81**	**85.35**	**96.56**	**87.11**	**88.50**	**91.95**

"Techniques and algorithms for computer aided diagnosis of pigmented skin lesions—a review," Biomedical Signal Processing and Control, vol. 39, pp. 237–262, 2018.

[5] A. A. Adegun and S. Viriri, "Deep learning techniques for skin lesion analysis and melanoma cancer detection: a survey of state-of-the-art," Artificial Intelligence Review, pp. 1–31, 2020.

[6] Z. Wei, H. Song, L. Chen, Q. Li, and G. Han, "Attention-based denseunet network with adversarial training for skin lesion segmentation," IEEE Access, vol. 7, pp. 136 616– 136 629, 2019.

[7] H. M. U¨ nver and E. Ayan, "Skin lesion segmentation in dermoscopic images with combination of yolo and grabcut algorithm," Diagnostics, vol. 9, 2019.

[8] O. Ronneberger, P. Fischer, and T. Brox, "U- net: Convolutional networks for biomedical image segmentation," Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics), vol. 9351, pp. 234–241, 2015.

[9] N. Abraham and N. M. Khan, "A Novel Focal Tversky Loss Function With Improved Attention U-Net for Lesion Segmentation," in 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019). IEEE, 2019, pp. 683–687. [Online]. Available: https://ieeexplore.ieee.org/document/8759329/

[10] O. Oktay, J. Schlemper, L. L. Folgoc, M. Lee, M. Heinrich, K. Misawa, K. Mori, S. McDonagh, N. Y. Hammerla, B. Kainz et al., "Attention u-net: Learning where to look for the pancreas," arXiv preprint arXiv:1804.03999, 2018.

[11] L. Liu, L. Mou, X. X. Zhu, and M. Mandal, "Skin lesion segmentation based on improved u-net," in 2019 IEEE Canadian Conference of Electrical and Computer Engineering (CCECE). IEEE, 2019, pp. 1–4.

[12] 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, 2019.

[13] M. Dash, N. D. Londhe, S. Ghosh, R. Raj, and R. S. Sonawane, "A cascaded deep convolution neural network based CADx system for psoriasis lesion segmentation and severity assessment," Applied Soft Computing Journal, vol. 91, p. 106240, 2020. [Online]. Available: https: //doi.org/10.1016/j.asoc.2020.106240

[14] R. Ramadan and S. Aly, "CU-Net: A New Improved Multi-Input Color U-Net Model for Skin Lesion Semantic Segmentation," IEEE Access, vol. 10, pp. 15 539–15 564, 2022.

[15] S. Aly and R. Ramadan, "DGCU – Net : A new dual gradient-color deep convolutional neural network for efficient skin lesion segmentation," Biomedical Signal Processing and Control, vol. 77, p. 103829, 2022.

[16] C. A. Barcelos and V. B. Pires, "An automatic based nonlinear diffusion equations scheme for skin lesion

segmentation," Applied Mathematics and Computation, vol. 215, no. 1, pp. 251–261, 2009. [Online]. Available: http://dx.doi.org/10.1016/j.amc.2009.04.081

[17] F. Rashid Sheykhahmad, N. Razmjooy, and M. Ramezani, "A novel method for skin lesion segmentation," International Journal of Information, Security and Systems Management, vol. 4, no. 2, pp. 458–466, 2015.

[18] J. Yasmin and M. Sathik, "An improved iterative segmentation algorithm using canny edge detector for skin lesion border detection," International Arab Journal of Information Technology, vol. 12, no. 4, pp. 325–332, 2015.

[19] "Image Segmentation using Gradient-based Histogram Thresholding for Skin Lesion Delineation," in Proceedings of the 12th International Joint Conference on Biomedical Engineering Systems and Technologies, no. Biostec. SCITEPRESS - Science and Technology Publications, 2019, pp. 84–91. [Online]. Available: http://www.scitepress.org/DigitalLibrary/Link.aspx?doi=10.5220/0007354100840091

[20] A. R. Ali, J. Li, S. J. O'Shea, G. Yang, T. Trappenberg, and X. Ye, "A Deep Learning Based Approach to Skin Lesion Border Extraction with a Novel Edge Detector in Dermoscopy Images," Proceedings of the International Joint Conference on Neural Networks, vol. 2019-July, 2019.

[21] D. Bahdanau, K. Cho, and Y. Bengio, "Neural machine translation by jointly learning to align and translate," arXiv preprint arXiv:1409.0473, 2014.

[22] M. Holschneider, R. Kronland-Martinet, J. Morlet, and P. Tchamitchian, "A real-time algorithm for signal analysis with the help of the wavelet transform," in Wavelets. Springer, 1990, pp. 286–297.

[23] M. J. Shensa et al., "The discrete wavelet transform: wedding the a trous and mallat algorithms," IEEE Transactions on signal processing, vol. 40, no. 10, pp. 2464– 2482, 1992.

[24] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, and Z. Wen, "FAT-Net: Feature adaptive transformers for automated skin lesion segmentation," Medical Image Analysis, vol. 76, p. 102327, 2022. [Online]. Available: https://doi.org/10.1016/j.media.2021.102327

[25] 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," Medical Image Analysis, vol. 53, pp. 197–207, 2019. [Online]. Available: https://doi.org/10.1016/j.media.2019.01.012

[26] D. Jha, P. H. Smedsrud, M. A. Riegler, D. Johansen, T. De Lange, P. Halvorsen, and H. D. Johansen, "ResUNet++: An Advanced Architecture for Medical Image Segmentation," Proceedings - 2019 IEEE International Symposium on Multimedia, ISM 2019, pp. 225–230, 2019.

[27] 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," Proceedings of 2020 5th International Conference on Green Technology and Sustainable Development, GTSD 2020, pp. 366–371, 2020.

[28] 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," Medical Image Analysis, vol. 64, 2020.

[29] Q. Jin, H. Cui, C. Sun, Z. Meng, and R. Su, "Cascade knowledge diffusion network for skin lesion diagnosis and segmentation," Applied Soft Computing Journal, vol. 99, p. 106881, 2021. [Online]. Available: https://doi.org/10.1016/j.asoc.2020.106881