



A New Hybrid Model for Segmentation of the Skin Lesion Based on Residual Attention U-Net

Saleh Naif Almuayqil¹, Reham Arnous^{2,*}, Noha Sakr³ and Magdy M. Fadel³

¹Department of Information Systems, College of Computer and Information Sciences, Jouf University, Sakaka, Saudi Arabia

²Department of Communications and Electronics, Delta Higher Institute for Engineering and Technology (DHIET), Mansoura, Egypt

³Computers Engineering and Control Systems Department, Faculty of Engineering, Mansoura University, Mansoura, Egypt

*Corresponding Author: Reham Arnous. Email: dr.reham.arnous2022@gmail.com

Received: 21 December 2022; Accepted: 27 February 2023

Abstract: Skin segmentation participates significantly in various biomedical applications, such as skin cancer identification and skin lesion detection. This paper presents a novel framework for segmenting the skin. The framework contains two main stages: The first stage is for removing different types of noises from the dermoscopic images, such as hair, speckle, and impulse noise, and the second stage is for segmentation of the dermoscopic images using an attention residual U-shaped Network (U-Net). The framework uses variational Autoencoders (VAEs) for removing the hair noises, the Generative Adversarial Denoising Network (DGAN-Net), the Denoising U-shaped U-Net (D-U-NET), and Batch Renormalization U-Net (Br-U-NET) for removing the speckle noise, and the Laplacian Vector Median Filter (MLVMF) for removing the impulse noise. In the second main stage, the residual attention u-net was used for segmentation. The framework achieves (35.11, 31.26, 27.01, and 26.16), (36.34, 33.23, 31.32, and 28.65), and (36.33, 32.21, 28.54, and 27.11) for removing hair, speckle, and impulse noise, respectively, based on Peak Signal Noise Ratio (PSNR) at the level of (0.1, 0.25, 0.5, and 0.75) of noise. The framework also achieves an accuracy of nearly 94.26 in the dice score in the process of segmentation before removing noise and 95.22 after removing different types of noise. The experiments have shown the efficiency of the used model in removing noise according to the structural similarity index measure (SSIM) and PSNR and in the segmentation process as well.

Keywords: Skin tumor; speckle noise; impulse noise; hair noise; deep learning; segmentation

1 Introduction

Medical image segmentation is one of the most critical fields in computer vision. Segmentation means dividing the images into disjointed parts to simplify the processes of analysis, diagnosis, and



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

examination [1–3]. Medical images can be used to treat different organ tumors, such as brain, breast, liver, cardiac tumors, etc. Early detection of tumors is critical for clinicians in the examination and treatment of patients [4]. There are many types of medical images, such as magnetic resonance imaging (MRI), computed tomography (CT), positron emission tomography (PET), mammography, and dermoscopic, which can effectively participate in the process of segmentation and detection [5].

Skin lesion segmentation is one of the most critical types of segmentation. Lesions can be effectively segmented to aid in the detection of melanoma. Melanoma, the most common type of skin cancer, can occur on normal skin without warning or develop on a preexisting lesion. Therefore, lesions must be carefully monitored to avoid harmful consequences. The segmentation of the skin lesion images can aid in melanoma examination and diagnosis. The rate of skin cancer is extremely high worldwide for both males and females. Skin lesions are one of the most common causes of death worldwide for both males and females. Malignant melanoma is among the worst and most deadly cancers. Early melanoma segmentation and detection contribute significantly to treatment and reduce the high risk of skin cancer [6]. Recently, several frameworks based on deep learning architectures have presented several methods for segmenting skin medical images, including Convolutional Neural Networks (CNN), Fully Convolution Networks (FCN), Generative Adversarial Networks (GAN), U-shaped Networks (U-Net), and other [7–9]. Meta-heuristics also introduce significant methods in the segmentation process [10,11].

Removing the noise from medical images, such as skin images, can participate in the process of enhancing the results of segmentation. The noise in skin images that occurs during picture capture, transmission, and compression is known as noise. The noise is produced due to signal fluctuations or faulty imaging equipment utilized during the image capture process. This noise can negatively affect the quality of the image. Different types of noises can negatively affect the processing and segmentation of skin images, such as impulse noise, speckle noise, and hair noise [12,13]. Preprocessing removes different types of noise from the skin images [14]. The process of removing the noise leads to increased accuracy in segmentation.

The contributions of this paper can be summarized as follows:

1. The framework is the first framework that can eliminate three types of noise at the same time.
2. The framework enhances the accuracy of different types of U-Net architecture after removing different types of noise.
3. The framework has achieved better accuracy than the other framework for segmentation of the skin lesion and in removing different types of noise.

The rest of the paper is organized as follows: Section 2 introduces the related work, and Section 3 introduces the methodology, which describes each stage in the novel framework, such as importing the dataset, reducing different types of noise such as impulse, hair, and speckle noise, and the residual attention U-Net. Section 4 discusses the results that show the results of different sections of the framework and the comparison between the framework and another framework for removing noise or segmenting skin tumors. Section 5 introduces the discussion. Section 6's conclusion introduces the limitations and discussion, and Section 7 introduces future work.

2 Related Work

Many large amounts of scientific research have introduced skin lesion segmentation performed with image processing methods dependent on deep learning approaches. Segmented images of skin lesions have been used in scientific research to increase classification accuracy. However, image

classification, segmentation, and object recognition have all benefited from the increased attention given to deep learning (DL), especially by using U-Net. The model was developed for medical segmentation and can work with small datasets while providing high accuracy.

Ahmadi [15] has presented a modified U-Net for the segmentation of skin lesions. The authors introduced a Multi-Scale Attention U-Net for skin lesion segmentation. In particular, the author has improved the classical version of the U-Net by adding an attention gate. The attention module aggregates the multi-level representation in a non-linear manner. A Bidirectional Convolutional Long Short-Term Memory (BDC-LSTM) structure is then used to retrieve the common discriminatory characteristics and suppress the less useful ones. The model has achieved accuracy near 0.89 in dice score. The model has some limitations, such as the value of the dice score and the noisy images. Kartal et al. [16] introduced the U-Net model with some modifications to the backbone by adding a model called Efficient Net. The model uses the ISIC 2017 dataset, which contains 2750 dermoscopic images. The model has achieved a dice score near 0.889. To improve the dice score, this model requires a sequence of enhancements and preprocessing in the dataset. Abraham et al. [17] introduced a focal Tversky loss function to maximize the balance between the precision and recall value of segmentation metrics. By combining the sizes of the image stacks used as input during model training, they created an attention U-Net model. They introduced low standard deviations for balanced precision-recall scores and surpassed the baseline U-Net in dice scoring. A cascaded deep convolutional neural network-based automated procedure for segmenting and assessing psoriasis was introduced by Dash et al. [18]. For the implementation and training of segmentation and classification tasks, respectively, the modified Visual Geometry Group (VGG) and U-Net models are used. Their modified VGG model demonstrated greater performance for both binary and multiclass classification approaches, producing higher classification accuracy with fewer trainable parameters. Nisar et al. [9] used the classical variant of U-net to segment atopic eczema skin lesions. The authors have used only the limited number of images during the training process. They have achieved a validation accuracy of 85.16%. Table 1 shows the comparison between the different architecture of U-Net in segmentation and detection of the skin tumor.

Table 1: Segmentation and detection of skin using U-net

Reference	DL model	Task	Advantages	Disadvantages
[15]	modified U-Net	Segmentation	Types of images in dataset	Lack of preprocessing
[16]	U-Net	Segmentation and detection	Efficient learning process	Noisy images
[17]	U-Net	segmentation and augmentation	Loss function	Noisy images
[18]	U-Net and VGG	Segmentation	greater performance for both binary and multiclass classification	Accuracy and noisy images
[9]	U-Net	Segmentation of atopic eczema skin lesions	Simple architecture of the model	Accuracy and limited images

Even though many different kinds of U-Net architectures have been made to segment different skin images, there are still some problems, such as how to remove different types of noise from the skin images and how to judge them.

3 Methodology

This section of the paper presents each step in the proposed framework, starting with removing different noises such as impulse, hair, and speckle noise, and also presents the steps of the segmentation process using the residual attention U-Net.

Steps of methodology:

Step 1: image acquisition: importing the dermoscopic images.

Step 2: Image Preprocessing: This step contains three stages.

- Stage A: The framework uses MLVMF to remove the impulse noise.
- Stage B: The framework uses DGAN-Net, D-U-Net, and Br-U-Net for removing the speckle noise.
- Stage C: Removing the hair noise using

Step 3: segmentation: using residual attention for the segmentation step for the freest image from the previous step [Fig. 1](#) and Peusdo.1 show the block diagram of the steps.

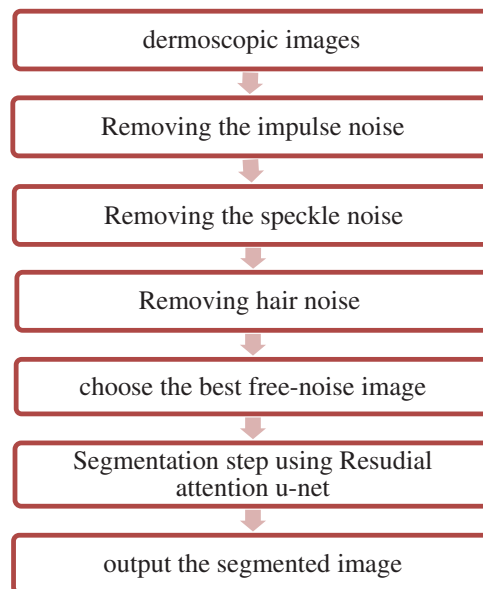


Figure 1: Methodology block diagram

Pseudocode 1: Pseudocode of the framework

Input: dermoscopic images Output: segmented image

-
- Importing the dermoscopic images
 - Insert the dermoscopic image of the patient
 - Uses MLVMF for removing the impulse noise.
-

(Continued)

Pseudocode 1: Continued

- Reduce speckle noise with different modified methods
- Removing the hair noise using variational autoencoder for removing the hair noises
- Compile Residual attention U-Net after removing different types of noises from the dataset images
- Segmentation process for the patient image
- Output the segmented image

3.1 Dermoscopic Dataset

The novel framework uses a dataset from Kaggle called PH2-resized2. The images in the dataset have been collected from dermatoscopic medical images of different populations of natural people. The dataset contains data for 200 different patients. The dataset contains two folders; the first folder is for the data images, and the second folder contains the masks for the skin images. Fig. 2 shows different images from the dataset [19].

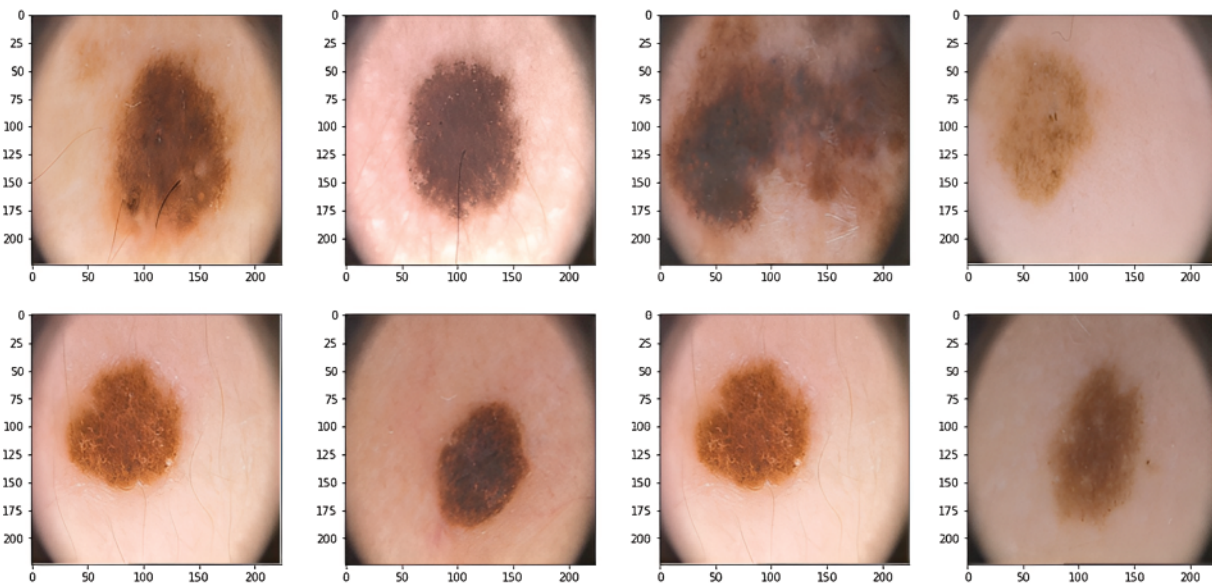


Figure 2: Dermoscopic images

3.2 Removing Speckle Noise

The framework uses DGAN-Net, D-U-Net, and Br-U-Net to remove speckle noise. The framework uses many modified methods to remove the speckle noise and chooses the model between them according to the noise level and the value of PSNR or SSIM.

3.2.1 Generative Adversarial Denoising Network

There are two components to this model: a generator and a discriminator. The generator is used to generate the fake skin images, and the discriminator is used to differentiate between the actual images and the fake images. The generator is regarded as a counterfeiter. A noisy picture is provided as generator input to the denoising GAN network. The discriminator determines if the image is actual or fraudulent, while the generator creates the denoised image. The 15 levels of the U-shaped

generator network model. The discriminating process is carried out by the convolutional layers of the discriminator network, which has 15 layers in total. The architecture consists of two different and connected parts of the generator deep network in the upper part of the figure, which reads a noise image of 128×128 and generates a generated image with the same pixel. The architecture of the upper part contains a 3×3 convolution transpose layer, a leaky rectifier, and dropout and batch renormalization. The lower part of the image is called the discriminator deep network, which is used to differentiate between the generated image and the actual image [20,21]. The architecture was trained using an optimizer called Adamx, and the learning rate was set to 0.0005 with a batch size of 256.

3.2.2 Denoising U-shaped U-Net

The D-U-NET model is similar to the traditional U-Net model, with a few exceptions. The model contains two stages: contraction and expansion [22]. The contraction stage of the model uses the Max Pool layer to reduce the size of the original noisy image. The expansion stage works in the opposite direction and resizes the reduced image to its original size. Transpose convolution layers are also employed in our D-U-Net network in place of the upsampling layer in the network's expansion phase. As is known, feature extraction procedures in deep learning networks involve the convolution layer. This procedure causes the skin image size to be reduced by the convolution window. The structural characteristics of the layer's reducing picture size are recovered with little distortion thanks to the transposed convolution process, which is the opposite of the convolution process. On the other hand, the upsampling layer pairs the image to restore its original dimensions. However, when the image is reconstructed, structural deformation takes place. The architecture was trained using an optimizer called Adamx, and the learning rate was set to 0.0005 with a batch size of 256.

3.2.3 Batch Renormalization U-Net

The U-shaped model is the same as D-U-NET with a set of differences. The key distinction between Br-U-Net and D-U-Net is that batch renormalization layers have taken the place of the batch normalization layers in this network. The leaky rectifier layer employed by Br-U-Net is the other distinction between them. The internal variable shifting problem was addressed by Tian et al. by using the batch renormalization layer to conduct scroll and scale operations. The batch renormalization procedure fixes the gamma and beta parameters to values of 1 and 0, respectively. The architecture was trained using an optimizer called Adamx, and the learning rate was set to 0.0005 with batch size 256.

3.3 Removing Impulse Noise

For real-time denoising of lesion skin images tainted by "salt and pepper" impulsive noise, use the MLVMF [23–25]. The framework uses two rounds of three steps each: beginning with the identification of pixels in the first round. It can be filtered with a modified Laplacian filter to remove noise. Corrupted pixels then pass a validation test based on their locality. Finally, noisy pixels are changed using the Vector Median Filter. While the traditional Laplacian filters frequently utilize rotation steps of $\pi/2$ or $\pi/4$, the MLVMF uses a 5×5 window to observe the intensity variations surrounding each pixel of the image. The rotation step adjustment leads to improved noise-corrupted pixel identification. Despite this benefit, a substantial portion of the impulsive noise can cause two or more corrupted pixels (of the same intensity) to collide, making it impossible to distinguish between noise-corrupted and uncorrupted pixels. Then, using a second set of filters that are still based on the Laplacian operator but allow exclusive focus on the collision phenomenon, a second round is required. To begin, MLVMF

is tested on standard photos with noise levels ranging from 3% to 30% to validate the approach used to remove impulse noise from skin lesion images.

3.4 Removing Hair Noise Using Variational Auto Encoders

The framework uses variational autoencoders to remove hair noise from skin images [26]. The architecture of VAEs contains two-part encoders and decoders that work oppositely. The decoder compresses the input data into a lower-dimensional representation vector. It reconstructs the representation vector into the original domain using the representation vector. After going through both the encoder and the decoder, the VAE searches for the weights that minimize the calculated loss between the input and output pictures during training. By allowing the encoder to develop noise-free space in the latent space and learn to disregard the noise, the variational AE may successfully be used to denoise pictures, as shown in Fig. 3.

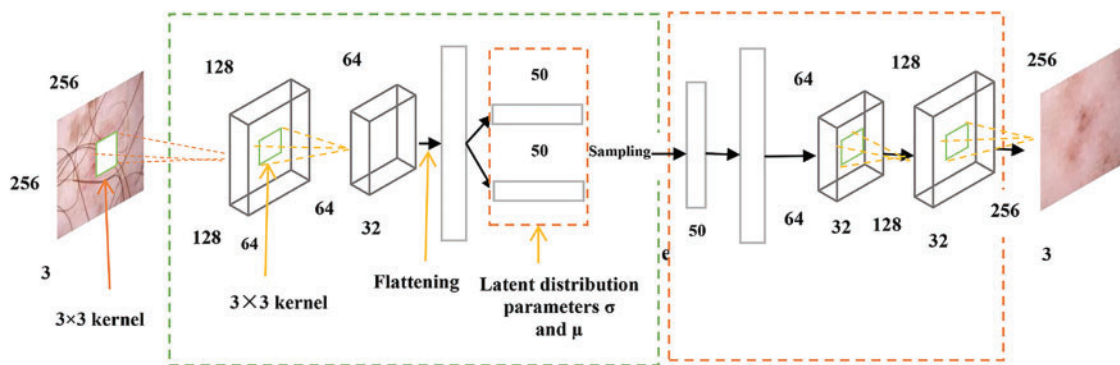


Figure 3: Architecture of hair removal [26]

The detailed architecture of the autoencoder model, which is used for removing the hair from the image, contains two layers. The first layer is based on three kernels (3×3) with flattening and rectifier activation functions and 64 filters. This part is called the encoder, and the second part is called the decoder. The decoder contains 64 filters with two Rectifier activation functions. The architecture uses two loss functions: square error and mean average error. In contrast to the input photos, the generative model produces hazy images that are unsuitable for the human visual system. This is because it builds the model using per-pixel loss [27].

The architecture was trained using an optimizer called Adamx, and the learning rate was set to 0.0005 in the first stage of training and 0.00001 in the second stage. The batch size is set to 256, while the number of epochs is 100 in the first stage of training and 50 in the second stage.

3.5 Segmentation Using Residual Attention U-Net

The framework uses the residual attention U-Net architecture to segment the skin images. This architecture has the same as the traditional U-Net but with a residual attention mechanism. The classic version of U-Net depends on the encoder and decoder parts; the encoder is used to extract features from the input skin image, while the decoder constructs the features. Skip-connections combine different-scale features from encoder and decoder parts. Residual blocks are presented in this type of network to change the initial convolutional layers and update the depth of the network [28–30]. Fig. 4 shows the spatial attention block of residual U-Net. The spatial attention block provides a lot of necessary background information since low-level characteristics lack semantic meaning. The

segmentation of the target object may be complicated by this information. The augmented attention module is created to take high-level semantic information and emphasize target aspects to solve this issue. By upsampling, the decoder retrieves the position information. However, this causes location data to be lost and causes edge blurring. Skip connections are used in some existing work to combine low-level characteristics with high-level features, which helps to refresh the position details. But this approach is naive. It provides unnecessary background information since low-level characteristics lack semantic meaning. The segmentation of the target object may be complicated by this information. The augmented attention module is created to capture high-level semantic information and accentuate target aspects in to solve this issue. The framework is trained over 200 epochs with a 0.0001 learning rate and 256 batch sizes.

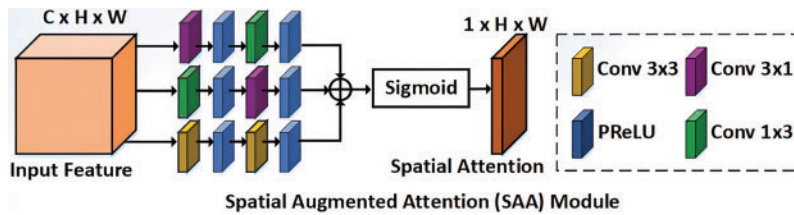


Figure 4: Architecture of spatial augmented attention module [30]

3.6 Hardware and Software Specification

The numerical experiments are carried out with an Intel® Core i7-10750H CPU running at 2.60 GHz on the 10th and an NVIDIA GeForce RTX 3050Ti laptop GPU, as shown in [Table 2](#).

Table 2: Hardware and software specification

Item	Content
Central processing unit (CPU)	Intel Core i7-10750H CPU running at 2.60 GHz
Graphical processing unit (GPU)	NVIDIA GeForce RTX 3050Ti laptop GPU
Random access memory (RAM)	128 GB
Operating system	Windows 10
Programming language	Python
Deep learning library	Tensorflow

4 Results and Discussion

This section shows the results of different stages and the comparison between the method used and other available methods that accomplish the same task.

4.1 Results of Removing Noise

This section presents the results of removing different types of noise from the lesion skin images in the three different stages of the framework. All the compared models in the experiments were trained over 100 epochs with a 0.0001 learning rate, 256 batch sizes, and the Admax optimizer.

Table 3 shows the results of removing the hair noise and the comparison between the used method and the other available deep learning models. The results in the table show the efficiency of the used method compared with other available deep learning models at different levels of noise, such as 0.1, 0.25, 0.5, and 0.75 in both PSNR and SSIM values. Table 3 also shows the runtime performances of the proposed networks for removing the hair noise. The last column shows the time per epoch based on the previous hardware specification, learning rate, and optimizer. The time in the column is calculated in minutes for each epoch. The proposed model outperforms the majority of the other models in the comparison in terms of running time. The model employs a 10% validation rate, a 20% dropout rate, and a mini-batch size of 32. Fig. 5 shows the comparison chart for removing hair noise.

Table 3: Removing hair noise methods

Method name	PSNR				SSIM				Time elapsed per epoch
	0.1	0.25	0.5	0.75	0.1	0.25	0.5	0.75	
Original mage	16.25	14.17	12.31	11.41	62.34%	56.38%	54.03%	52.84%	
CNN [31]	26.34	21.34	17.22	15.41	79.31%	75.66%	70.31%	64.21%	26.2
FCN [32]	27.21	24.24	21.33	18.31	84.36%	79.31%	77.43%	69.31%	23.1
Auto encoder [33]	27.31	23.64	21.35	19.39	84.31%	77.34%	76.97%	69.39%	24.3
GANs [34]	28.45	26.34	24.10	20.89	86.12%	82.41%	76.31%	73.16%	24.34
The used method	35.11	31.26	27.101	26.16	90.11%	86.90%	81.67%	78.12%	24

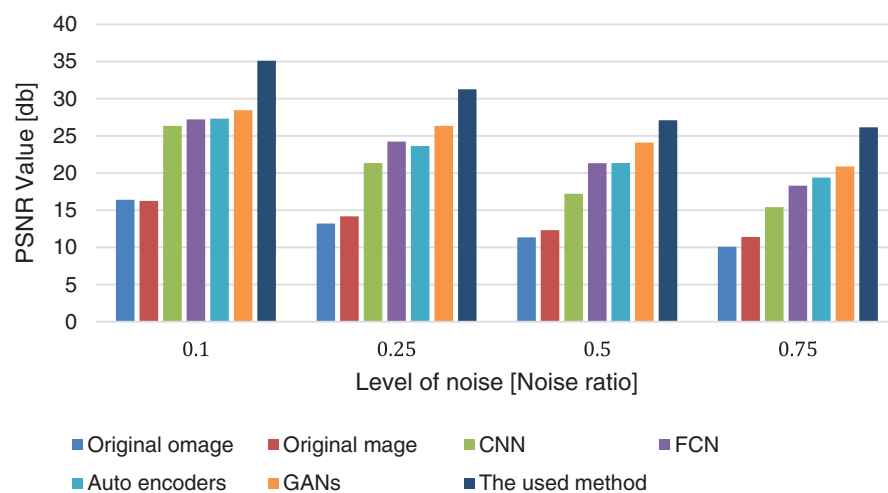


Figure 5: Hair removal comparison chart

The results of removing speckle noise from lesion skin images are shown in Table 4. The paper compares the modified method with other available deep-learning models for removing speckle noise. All models were trained in 50 epochs with the Adamx optimizer at a learning rate of 0.001. The time in the table is by the minute for each epoch, and the results in the table show the efficiency of the framework methods compared with other available methods at different levels of noise according to the values of PSNR and SSIM. The three models used for removing the speckle noise use a 10%

validation rate, a 20% dropout rate, a mini-batch size of 32, a sigmoid activation function, and an admax optimizer.

Table 4: Removing speckle noise methods

Method name	PSNR				SSIM				Time elapsed per epoch
	0.1	0.25	0.5	0.75	0.1	0.25	0.5	0.75	
Original skin image	15.45	12.34	11.43	10.99	66.22%	54.32%	44.65%	38.21%	
U-Net [14]	25.23	23.21	21.76	18.43	81.21%	77.53%	72.12%	65.44%	23.55
FCN [31]	22.21	20.76	16.43	15.61	84.22%	78.32%	69.39%	64.21%	22.54
CNN [32]	26.45	23.54	21.76	18.32	83.21%	80.01%	76.21%	68.21%	22.4
Auto encoder [33]	24.43	21.19	18.77	16.93	80.33%	77.76%	69.59%	63.44%	11.4
Br-U-NET	35.32	31.34	29.56	25.88	91.21%	89.73%	87.31%	84.21%	25.3
D-U-NET	36.34	33.23	31.32	28.65	92.32%	90.21%	88.54%	85.54%	26.21
DGAN-Net	34.32	33.21	31.54	29.22	90.21%	90.13%	89.21	86.21	25.3

Fig. 6 shows the comparison chart for removing the speckle noise. The figure contains four sections for different levels of noise, and each color refers to a different model for removing the speckle noise. The comparison shows the efficiency of the three methods compared with other deep learning models. When compared to CNN, FCN, autoencoders, and GANs with the same number of epochs and learning rate, the used method achieves high PSNR values at the level (0.1, 0.25, 0.5, and 0.7), as well as in SSIM.

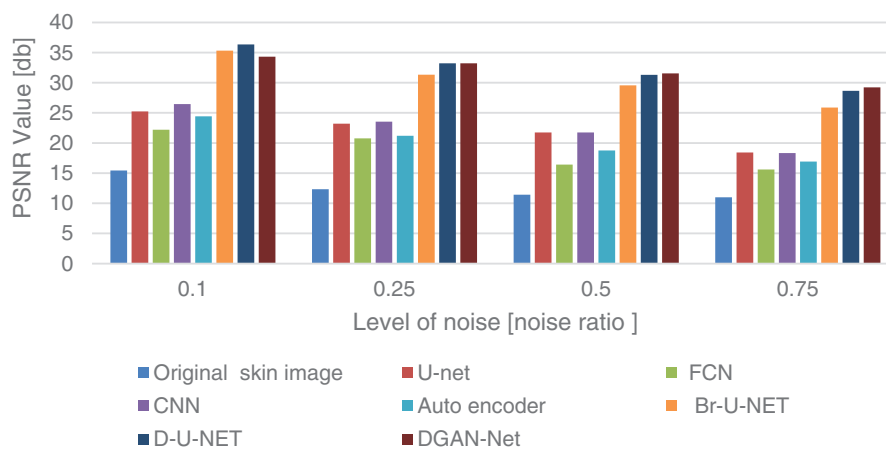


Figure 6: PSNR comparison chart for speckle noise removal

Table 5 and Fig. 7 show the results of removing impulse noise from the skin using the used method and the other available filters. The last column in the table shows the time per epoch for each model and the three proposed models for removing the impulse noise. All deep learning models have been used in the framework using Adamx as an optimizer, the Rectifier activation function, 100 epochs.00005 as

the learning rate, and 128 as the batch size. The proposed model for removing the hair noise uses a 10% validation rate, a 20% dropout rate, a mini-batch size of 32, and a sigmoid activation function.

Table 5: Removing impulse noise methods

Method name	PSNR				SSIM				Time elapsed per epoch
	0.1	0.25	0.5	0.75	0.1	0.25	0.5	0.75	
Original mage	16.41	13.21	11.34	10.09	61.21%	57.21%	51.21%	50.21%	
CNN [31]	24.43	19.64	15.35	13.45	77.36%	74.21%	69.19%	57.24%	22.2
FCN [32]	26.14	23.12	19.26	17.48	81.42%	77.33%	75.12%	66.21%	24.1
Auto encoder [33]	26.12	23.52	20.31	18.41	82.31%	77.96%	75.21%	67.14%	23.1
GANs [34]	28.23	26.23	23.45	21.09	85.41%	83.91%	74.21%	73.21%	24.56
The used method	36.33	32.21	28.54	27.11	91.21%	88.32%	82.21%	79.21%	23.12

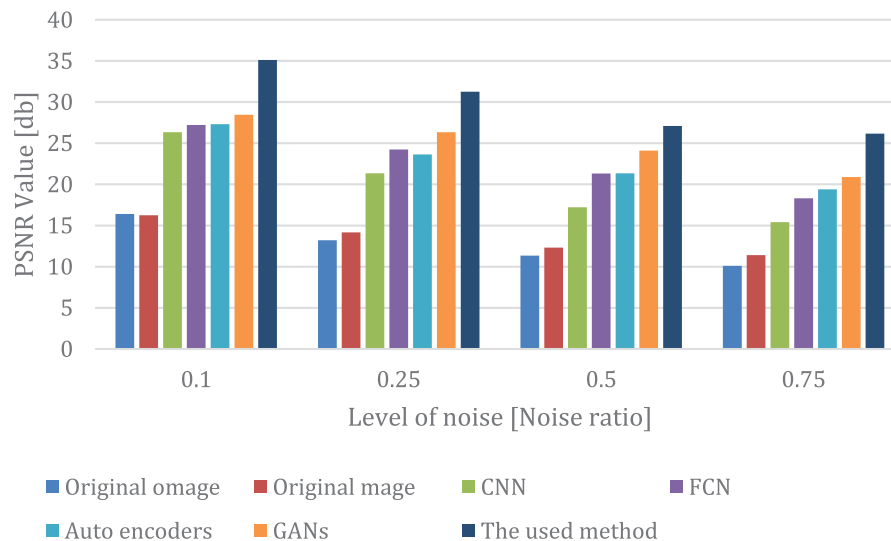


Figure 7: PSNR comparison chart for impulse noise removal

Fig. 7 contains four sections for different levels of noise, and each color refers to a different model in the comparison. The comparison shows the efficiency of the used method compared with other deep learning models. When compared to CNN, FCN, autoencoders, and GANs with the same number of epochs and learning rate, the used method achieves high PSNR values at the level (0.1, 0.25, 0.5, and 0.7) as well as in SSIM with the same level.

4.2 Results of Segmentation Step

This section shows the results of the segmentation of the dermoscopic images after removing the impulse, hair, and speckle noise in the previous stage of the framework.

Table 6 shows the results of the segmentation process before removing the noises, and Table 7 shows the results after removing the different types of noises. The results in the table show the efficiency of residual attention U-Net compared with other architectures of U-Net.

Table 6: Comparison between different methods of segmentation before reducing different types of noise

Model	Dice score	Accuracy	Precision	Recall	Time per epoch
Attention gate U-Net [35]	93.11	93.18	92.11	91.11	24.56
U-Net [14]	91.87	91.81	91.61	91.25	24.69
U-Net++ [36]	91.11	91.09	91.18	91.19	26.31
Attention U-Net++ [37]	91.12	91.15	91.21	91.21	27.44
CNN [31]	88.23	88.11	87.21	88.23	24.87
FCN [32]	89.91	89.44	89.43	89.23	24.71
Residual Attention U-Net	94.26	94.11	94.02	94.11	23.19

Table 7: Comparison between different methods of segmentation after reducing different types of noise

Model	Dice score	Accuracy	Precision	Recall	Time per epoch	Inference time per images in ms
Attention gate U-Net [33]	93.21	93.51	93.16	92.31	24.3	86
U-Net [12]	92.31	92.33	92.12	92.03	24.65	87.3
U-Net++ [34]	91.93	91.81	92.01	92.39	26.21	88
Attention U-Net++ [35]	92.01	91.49	91.88	91.48	27.43	89
CNN [29]	91.21	91.65	91.03	91.26	24.87	88.3
FCN [30]	91.86	91.36	91.61	91.73	24.65	88.44
Residual Attention U-Net	95.22	94.19	94.12	94.61	23.1	83

The last column in Tables 6 and 7 shows the time per epoch for each model. The time in the column shows the efficiency of the running time factors and all used evaluation metrics before and after removing the noise. The time of the epoch has been calculated per minute in the same hardware specification. All deep learning models have been used in the framework using Adamx as an optimizer, the Rectifier activation function, 100 epochs.00005 as the learning rate, and 128 as the batch size. The last column in the Table 7 shows the final inference time of each image in Mille Seconds (MS). All the models for segmentation use a 10% validation rate, a 20% dropout rate, a mini-batch size of 32, and a sigmoid activation function.

Fig. 8 shows the segmented image and the predicted mask. The image shows the efficiency of the segmented image after using residual attention U-Net after removing noise in the previous stage.

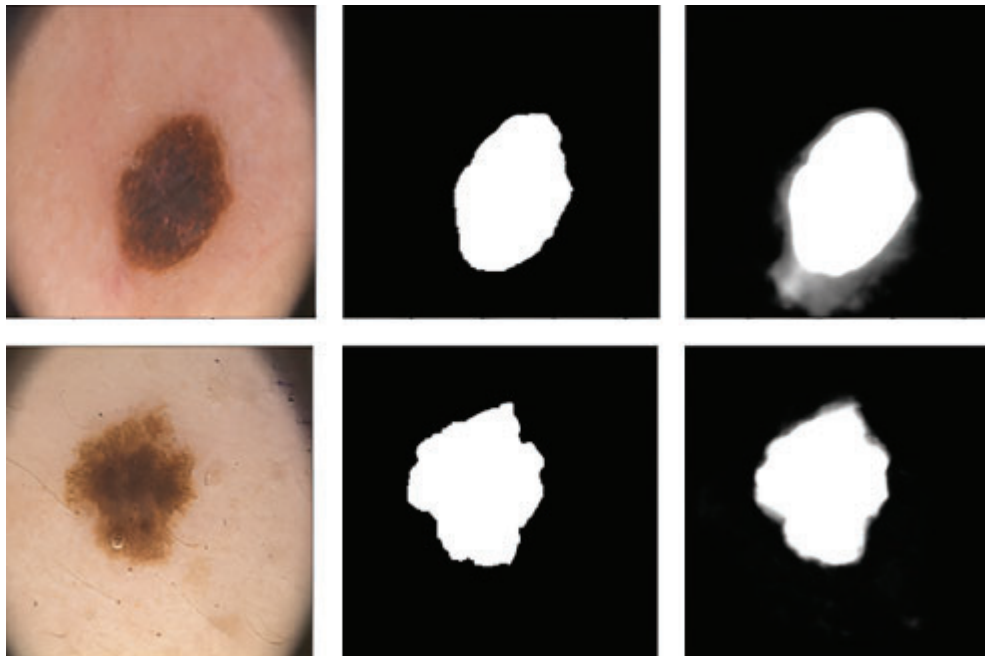


Figure 8: Original image, mask, and predicted mask using the framework

5 Discussion

According to the results of the experiments in both removing different types of noise and in the segmentation process, it seems clear that the efficiency of the proposed hybrid framework is evident at different stages of the model. The proposed hybrid framework has achieved an accuracy of nearly 94.26 in the dice score in the process of segmentation before removing noise and 95.22 after removing different types of noise using Attention residual U-Net. According to PSNR and SSIM values at noise levels of 0.1, 0.25, 0.5, and 0.6, the framework outperformed the other types in removing different types of noise. The results also show high values of PSNR and SSIM for the models used for different types of noise. The experiments have also shown the efficiency of the used model when compared with other available models.

6 Conclusion

Segmentation of skin lesion tumors is one of the most vital tasks for dermatologists. Segmentation helps dermatologists detect the tumor in its early stages. This paper has presented a novel framework for the segmentation of skin tumors after removing the main types of noise, such as impulse, hair, and speckle. The framework has presented two stages: the first stage is for removing noise, and the second stage is for segmenting the free images after removing the noise. The framework uses a variational autoencoder for removing the hair noises, DGAN-Net, D-U-NET, and Br-U-NET for removing the speckle noise; and the Laplacian Vector Median Filter MLVMF for removing the impulse noise. Using residual attention U-Net, the framework achieved efficient results in both noise removal and segmentation processes. The experiments done in the paper show the efficiency of the used models compared to other available models in removing different types of noise based on PSNR values and also show the efficiency of the segmentation process when compared with other available models.

Although the paper is significant in terms of noise removal and segmentation, it does have some limitations, including the computational time required when dealing with high-resolution images and the inability to deal with different types of noise.

7 Future Work

In the future, such as enhancing the detection of new skin tumor lesions, particularly in the presence of such challenging lesions. This might involve, for instance, bringing into the system each rater's unique delineations. To address the data scarcity and class imbalance issues encountered in this work, transfer learning from a similar lesion segmentation task with a larger dataset was used to improve the segmentation and detection processes. In the future, the smoothing function and different evaluation metrics will also be taken into consideration.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] Y. Jiang, S. Cao, S. Tao and H. Zhang, "Skin lesion segmentation based on multi-scale attention convolutional neural network," *IEEE Access*, vol. 8, pp. 122811–122825, 2020.
- [2] A. F. Subah, O. I. Khalaf, Y. Alotaibi, R. Natarajan, N. Mahadev *et al.*, "Modified self-adaptive bayesian algorithm for smart heart disease prediction in IoT system," *Sustainability*, vol. 14, no. 21, pp. 14208, 2022.
- [3] P. Jayapradha, Y. Alotaibi, O. I. Khalaf and S. Alghamdi, "Heap bucketization anonymity-An efficient privacy preserving data publishing model for multiple sensitive attributes," *IEEE Access*, vol. 10, pp. 28773–28791, 2022.
- [4] H. Wu, J. Pan, Z. Li, Z. Wen and J. Qin, "Automated skin lesion segmentation via an adaptive dual attention module," *IEEE Transactions on Medical Imaging*, vol. 40, no. 1, pp. 357–370, 2021.
- [5] Y. Qiu, J. Cai, X. Qin and J. Zhang, "Inferring skin lesion segmentation with fully connected CRFs based on multiple deep convolutional neural networks," *IEEE Access*, vol. 8, pp. 144246–144258, 2020.
- [6] R. Rout, P. Parida, Y. Alotaibi, S. Alghamdi and O. I. Khalaf, "Skin lesion extraction using multiscale morphological local variance reconstruction based watershed transform and fast fuzzy c-means clustering," *Symmetry*, vol. 13, no. 11, pp. 2085, 2021.
- [7] A. R. H. Ali, J. Li and G. Yang, "Automating the ABCD rule for melanoma detection: A survey," *IEEE Access*, vol. 8, pp. 83333–83346, 2020.
- [8] M. Goyal, A. Oakley, P. Bansal, D. Dancy and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2019.
- [9] H. Nisar, Y. Tan and Y. Kimho, "Segmentation of eczema skin lesions using U-Net," in *Proc. IEEE Conf. on Biomedical Engineering and Sciences (IECBES 2021)*, Malaysia, Langkawi Island, pp. 362–366, 2021.
- [10] L. Wei, S. Xiao, Y. Nanekaran and V. Rajinikanth, "An optimized method for skin cancer diagnosis using modified thermal exchange optimization algorithm," *Computational and Mathematical Methods in Medicine*, vol. 2021, no. 1, pp. 1–15, 2021.
- [11] Y. Nanekaran, Z. Licai, J. Chen, Q. Zhongpan, Y. Xiaofeng *et al.*, "Diagnosis of chronic diseases based on patients health records in IoT healthcare using the recommender system," *Wireless Communications and Mobile Computing*, vol. 2022, no. 1, pp. 1–14, 2022.
- [12] A. A. Khan and F. A. Khan, "A cost-efficient radiation monitoring system for nuclear sites: Designing and implementation," *Intelligent Automation & Soft Computing*, vol. 32, no. 1, pp. 1357–1367, 2022.

- [13] A. Mahmoud, S. Rabaie, T. Taha and F. Abd El-Samie, "Mixed curvelet and wavelet transforms for speckle noise reduction in ultrasonic b-mode images," *Information Sciences and Computing*, vol. 3, no. 9, pp. 1–21, 2015.
- [14] L. Wei, S. Xiao, Y. A. Nanekaran and V. Rajinikanth, "An optimized method for skin cancer diagnosis using modified thermal exchange optimization algorithm," *Computational and Mathematical Methods in Medicine*, vol. 2, no. 1, pp. 1–11, 2021.
- [15] M. D. Alahmadi, "Multiscale attention U-Net for skin lesion segmentation," *IEEE Access*, vol. 10, pp. 59145–59154, 2022.
- [16] M. S. Kartal and O. Polat, "Segmentation of skin lesions using U-Net with efficientNetB7 backbone," in *Proc. Innovations in Intelligent Systems and Applications Conf. (ASYU 2022)*, Antalya, Turkey, pp. 1–5, 2022.
- [17] N. Abraham and N. M. Khan, "A novel focal tversky loss function with improved attention U-Net for lesion segmentation," in *Proc. IEEE 16th Int. Symp. on Biomedical Imaging (ISBI 2019)*, Venice, Italy, pp. 683–687, 2019.
- [18] 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, no. 1, pp. 106240, 2020.
- [19] P. Brahmabhatt, "PH2_resized2," 2018. [Online]. Available: <https://www.kaggle.com/datasets/paramaggargwal/fashion-product-images-dataset>
- [20] O. Karaoglu, H. S. Bilge and I. Uluer, "Removal of speckle noises from ultrasound images using five different deep learning networks," *Engineering Science and Technology an International Journal*, vol. 29, no. 1, pp. 101030–101049, 2022.
- [21] Z. Chen, Z. Zeng, H. Shen, X. Zheng and P. Dai, "DN-GAN: Denoising generative adversarial networks for speckle noise reduction in optical coherence tomography images," *Biomedical Signal Processing and Control*, vol. 55, no. 1, pp. 101632–101648, 2020.
- [22] S. Ai and J. Kwon, "Extreme low-light image enhancement for surveillance cameras using attention U-Net," *Sensors Multidisciplinary Digital Publishing Institute*, vol. 20, no. 2, pp. 495–511, 2020.
- [23] J. Jaculin Femil and T. Jaya, "An efficient hybrid optimization for skin cancer detection using pnn classifier," *Computer Systems Science and Engineering*, vol. 45, no. 3, pp. 2919–2934, 2023.
- [24] A. Mirbeik-Sabzevari and N. Tavassolian, "Ultrawideband, stable normal and cancer skin tissue phantoms for Millimeter-wave skin cancer imaging," *IEEE Transactions on Biomedical Engineering*, vol. 66, no. 1, pp. 176–186, 2018.
- [25] M. Goyal, A. Oakley, P. Bansal, D. Dancy and M. H. Yap, "Skin lesion segmentation in dermoscopic images with ensemble deep learning methods," *IEEE Access*, vol. 8, pp. 4171–4181, 2019.
- [26] M. Hamraz, D. M. Khan, N. Gul, A. Ali, Z. Khan *et al.*, "Regulatory genes through robust-snr for binary classification within functional genomics experiments," *Computers, Materials & Continua*, vol. 74, no. 2, pp. 3663–3677, 2023.
- [27] S. Gopikha and M. Balamurugan, "Regularised layerwise weight norm based skin lesion features extraction and classification," *Computer Systems Science and Engineering*, vol. 44, no. 3, pp. 2727–2742, 2023.
- [28] A. A. Alhadad, O. Tarawneh, R. R. Mostafa and H. M. El-Bakry, "Residual attention deep svdd for COVID-19 diagnosis using ct scans," *Computers, Materials & Continua*, vol. 74, no. 2, pp. 3333–3350, 2023.
- [29] J. Qu, J. Liu and C. Yu, "Adaptive multi-scale hypernet with bi-direction residual attention module for scene text detection," *Journal of Information Hiding and Privacy Protection*, vol. 3, no. 2, pp. 83–89, 2021.
- [30] Y. Wang, Y. Yang, H. Chen, H. Zheng and H. Chang, "End-to-end handwritten chinese paragraph text recognition using residual attention networks," *Intelligent Automation & Soft Computing*, vol. 34, no. 1, pp. 371–388, 2022.
- [31] S. Albawi, T. A. Mohammed and S. Al-Zawi, "Understanding of a convolutional neural network," in *Proc. Int. Conf. on Engineering and Technology (ICET 2017)*, Antalya, Turkey, pp. 1–6, 2017.

- [32] S. Sadrizadeh, H. Otroshi-Shahreza and F. Marvasti, "Removing impulsive noise from color images via a residual deep neural network enhanced by post-processing," in *Proc. European Signal Processing Conf. (EUSIPCO 2021)*, Dublin, Ireland, pp. 656–660, 2021.
- [33] A. Pawar, "Noise reduction in images using autoencoders," in *Proc. Int. Conf. on Intelligent Sustainable Systems (ICISS 2020)*, Thoothukudi, India, pp. 987–990, 2020.
- [34] H. Ashraf, Y. Jeong and C. H. Lee, "Underwater ambient noise removing GAN based on magnitude and phase spectra," *IEEE Access*, vol. 9, pp. 24513–24530, 2021.
- [35] J. Zhang, Z. Jiang, J. Dong, Y. Hou and B. Liu, "Attention gate ResU-Net for automatic MRI brain Tumor segmentation," *IEEE Access*, vol. 8, pp. 58533–58545, 2020.
- [36] L. Liu, L. Mou, X. X. Zhu and M. Mandal, "Skin lesion segmentation based on improved U-net," in *Proc. IEEE Canadian Conf. of Electrical and Computer Engineering (CCECE 2019)*, Edmonton, AB, Canada, pp. 1–4, 2019.
- [37] Z. Zhou, M. M. R. Siddiquee, N. Tajbakhsh and J. Liang, "UNet++: Redesigning skip connections to exploit multiscale features in image segmentation," *IEEE Transactions on Medical Imaging*, vol. 39, no. 6, pp. 1856–1867, 2020.