DFP-Net: Melanoma Segmentation based on Deep Feature Pyramid Network

HONGZAN WANG[1], ZHANLIN JI[1], IVAN GANCHEV[2,3,4]
[1]College of Mathematics and Computer Science,
Zhejiang A&F University,
Hangzhou 311300,
CHINA

[2]Faculty of Mathematics and Informatics,
University of Plovdiv "Paisii Hilendarski",
Plovdiv 4000,
BULGARIA

[3]Institute of Mathematics and Informatics,
Bulgarian Academy of Sciences,
Sofia 1040,
BULGARIA

[4]TRC/ECE,
University of Limerick,
Limerick, V94 T9PX,
IRELAND

**Corresponding Author*

Abstract: - The skin is the body's largest organ. Melanoma is a highly malignant skin tumor for which there is currently no effective treatment. Therefore, the best approach is to perform early diagnosis and treatment in initial stages. Thus, with an attitude of early detection and treatment to increase post-operative survival time, melanoma detection in its initial stages becomes especially important. In the era of advanced deep learning (DL), combining DL with medical imaging for computer-aided diagnosis (CAD) and lesion segmentation can not only increase the efficiency of doctors' work but also make it more convenient for the general public to pay attention to and monitor their own health. In this paper, a novel Deep Feature Pyramid Network (DFP-Net) DL model is proposed. It uses a Feature Pyramid Network (FPN) as a basis, with ResNet-101 as the backbone network, and incorporates newly designed Residual Receptive Field Blocks (RRFBs), to increase the receptive field while also enhancing its feature extraction capabilities. A novel Squeeze-and-Excitation Atrous Spatial Pyramid Pooling (SEASPP) module is added to capture and filter contextual information better after the feature extraction process. Newly proposed Depthwise Over-parameterized Convolutional Block Attention Modules (DOCBAMs) are introduced after each top-down step to further refine the features, which are then fused to produce the final output. Experiments, conducted on two datasets, demonstrate the superiority of the proposed DFP-Net model compared to state-of-the-art models, based on multiple metrics.

Key-Words: - melanoma; medical imaging; segmentation; computer-aided diagnosis (CAD); deep learning (DL); Feature Pyramid Network (FPN); Deep Feature Pyramid Network (DFP-Net).

Received: April 11, 2026. Revised: May 13, 2026. Accepted: June 14, 2026. Published: November 12, 2026.

1 Introduction

The skin is the body's largest organ, with functions such as protection, sensation, and temperature regulation. However, when skin cells lose control due to diseases, they may develop into skin cancer, [1]. There are primarily three types of skin cancer – basal cell carcinoma, squamous cell carcinoma, and melanoma, which often occur in areas exposed to

direct sunlight. However, this does not mean that areas not exposed to sunlight, like the palms and soles of the feet, are without risk. Melanoma is the most concerning type of skin cancer, making early detection and diagnosis crucial to increasing a patient's chances of survival, [2]. According to statistics from the United States in 2023, early-stage melanoma is entirely curable, [3]. Early melanomas are characterized by irregular shapes, unclear borders, and non-uniform color. Typically, melanomas have a diameter greater than 6 mm and may change over time, [4]. The difficulty in diagnosing melanoma arises from its resemblance to moles and other skin features, and the task is complicated by variations in skin lesions and skin diversity. Malignant melanoma often has irregular borders and can appear in different locations, making it challenging for non-specialist physicians to diagnose it, [5].

In the past decades, diagnosing melanoma required medical experts to rely on visual observation and auxiliary screening methods such as the ABCDE rule [6]. However, having doctors diagnose melanoma in every patient is clearly challenging, making it essential to use computers for assisted diagnosis to alleviate physicians' workload and improve efficiency, [7].

Research has introduced Computer-Aided Diagnosis (CAD) systems to enhance the accuracy and speed of melanoma diagnosis. However, existing algorithms for segmenting skin lesions perform poorly when handling images containing body hair and boundary edges, limiting the effectiveness of CAD systems. Although there are many skin lesion segmentation methods, some of them are only suitable for images without body hair and boundary edges, leading to incorrect segmentation results. This is because these methods often incorrectly classify boundary edges as part of the lesion, [8], [9].

Feature extraction plays a crucial role in image analysis, especially for CAD. Through computer analysis, one can effectively capture key features that may enable early detection. The advantage of early melanoma diagnosis lies in reducing the risk of fatality, with the ABCD criteria –serving as a reference in dermatoscopy– emphasizing the need for special attention when melanoma diameter (D) exceeds 6 mm. In the initial stages, minor surgeries can enhance a patient's chances of recovery. Therefore, automated lesion segmentation holds a significant position in computer-aided melanoma diagnosis [10].

In recent years, deep learning (DL)-based methods have achieved excellent results in many semantic segmentation tasks, with architectures like the fully convolutional neural network (CNN) U-Net demonstrating outstanding performance [11].

In the field of image analysis, particularly the analysis of histopathological images, segmentation-based approaches have been explored. Creating datasets for these tasks typically involves manual annotation by expert physicians, who label melanomas in images for computer training. However, a major challenge lies in the need for extensive and precise annotations [12], [13].

Some recent studies on medical segmentation models have found that traditional CNN architectures are not efficient and do not capture sufficient contextual information [14]. In the segmentation of medical skin lesions, lower resolution can provide better receptive fields for detecting larger objects, while higher resolution is needed to capture finer details, such as edge information. Therefore, fusing features of different scales has a significant impact on the effectiveness of medical skin lesion segmentation [15].

Several outstanding models have emerged in this domain. U-Net [16] introduced the encoder-decoder structure and achieved multiple first places in the 2015 IEEE International Symposium on Biomedical Imaging cell tracking competitions. UNet++ [17] improved upon U-Net by fusing features from different layers, having different receptive fields, allowing the model to extract features of various sizes. DoubleU-Net [18] stacked two U-Net models and used a pre-trained VGG-19 as the backbone network, incorporating Atrous Spatial Pyramid Pooling (ASPP) to gather diverse contextual information. Additionally, the Feature Pyramid Network (FPN) [19], originally from the object detection domain, has also performed well in the segmentation field. UNet 3+ [20] employed full-scale skip connections to combine different semantic information from feature maps of different scales. SegNet [21], proposed by the University of Cambridge, adopted an encoder–decoder structure, removing fully connected layers and introducing other enhancements to improve performance.

In this paper, a new melanoma segmentation model, named DFP-Net, is proposed, based on FPN. It combines novel attention mechanisms and incorporates multi-scale information. Experiments, conducted on the public ISIC2018 dataset and a combined public–private dataset, demonstrate that the proposed model outperforms state-of-the-art models, based on majority of evaluation metrics used.

The key innovations of the proposed model could be summarized as follows:

- In the feature extraction stage, newly designed Residual Receptive Field Blocks (RRFBs) are proposed for use, consisting of convolutional layers with different-size kernels and residual blocks in a multi-branch structure to increase the receptive field. This allows the model to extract multi-scale, multi-level information, preserving crucial details from the original images.
- A novel Squeeze-and-Excitation Atrous Spatial Pyramid Pooling (SEASPP) module is proposed to absorb semantic information from different scales during the upsampling stage, facilitating the suppression of unwanted information and directing the model's attention to the lesion area [22].
- Newly designed attention modules, called Depthwise Over-parameterized Convolutional Block Attention Modules (DOCBAMs), are proposed for use before the feature fusion stage, enabling weighted extraction of features, retaining important semantic information, and eliminating unnecessary information, such as the adverse effects of body hair, as to further enhance the model's capabilities.

The rest of the paper is organized as follows. The next section presents related work in this field, followed by a section dedicated to the structure of the proposed model. Description of the conducted segmentation performance experiments and obtained results comes next, followed by conclusions and presentation of future directions for research.

2 Related Work

With advanced research performed by experts and scholars in the DL area, the performance of image segmentation models is getting better and better. For example, the model proposed in [23] combines stacked CNNs with GrabCut-stacked CNNs (GC-SCNN) and support vector machines (SVM) for improved results. [24] studied the impact of image preprocessing on segmentation outcomes. [25] developed software for automatic detection and segmentation of skin moles, and [26] used an object detection CornerNet model to detect melanoma lesions, whereby local moles are passed as input to the fuzzy K-means (FKM) clustering method to perform segmentation tasks, making it easy for individual users to operate. [27] introduced a model that utilizes ResNeXt50 as a pretrained model and their own developed channel and spatial attention block (CSAB) and multi-scale capture block (MSCB) to obtain more information for enhanced robustness. [28] incorporated a hybrid residual attention (HRA) module into their self-developed MHAU-net model to address the issue of unclear image boundaries and achieved good results. The team in [29] proposed a CNN framework, based on atrous convolution, for automatic segmentation of lesion areas. [30] introduced a new attention-guided filter module and embedded it into the skip connections of attention-guided network with densely connected convolution for skin lesion segmentation (CSAG and DCCNet), improving the model's performance. [31] combined edge-guided modules and weighted aggregation modules in their network and proposed a fusion model using a combination of balanced binary cross-entropy (BBCE) and Focused Tversky Loss (FTL) as the loss function. [32] developed a novel feature extraction technique based on K-means clustering, considering skin color and texture features from five different color spaces for analysis. [33] used a pretrained MobileNetv2 for the backend of the DeepLabv3+ model and achieved high performance by extracting features from the pretrained N×1000 DesneNet201 and feeding them into SVM and K-Nearest Neighbors (KNN) classifiers. [1] introduced Multi-Scale Attention U-Net (MSAU-Net), which incorporates attention mechanisms at the bottom of U-Net and utilizes the Bidirectional Convolutional Long Short-term Memory (BDC-LSTM) structure for feature extraction, enhancing the utilization of useful information. [34] proposed a multimodal-based approach that effectively quantifies feature differences between different colors. They then employed a Gaussian Mixture Model for lesion edge detection and summarized the collected feature information. [35] presented a model capable of boundary contour detection, addressing the issue of regional confusion between normal and lesioned skin by incorporating selective kernel modules at skip connections. [36] developed a risk prediction system for mortality based on Positron Emission Tomography/Magnetic Resonance Imaging (PET/MRI) and PET/Computed Tomography (CT), including a CNN for metastatic melanoma patients before the initiation of systemic treatment. [37] proposed a model that eliminates the impact of factors like body hair through techniques such as maximum pooling and contrast adjustment. They then used a Fully Convolutional Network (FCN) for segmentation, followed by a series of post-processing steps, including resolution reduction, for melanoma classification. [38] introduced a learning approach based on super-pixel labels. They

combined super pixels formed by the Simple Linear Iterative Clustering (SLIC) algorithm with a thermal label constraint, defined a distance function, and converted it into a soft probability distribution. The labels were subsequently utilized in training the segmentation network. [39] classified annotation noise in image segmentation datasets and proposed two new methods for automatically identifying inconsistencies within and between annotators. [40] presented the O-Net model with an O-shaped architecture. They incorporated attention-based feature modules to iteratively refine segmentations in the wood board region, enhancing feature extraction and merging fine-grained features. [41] introduced the computational intelligence-based melanoma detection and classification technique using dermoscopic images (CIMDC-DI) model, which utilizes FKM clustering-based bilateral filtering for image preprocessing and Neural Architecture Search Net (NasNet) model for feature extraction. [42] proposed an anti-aliasing attention spatial convolution method capable of overcoming limitations posed by skin mirrors, achieving good results in hair-dense areas and cases with unclear contrast and color.

3 Proposed model: DFP-Net

The overall structure of the proposed DFP-Net model is shown in Figure 1 (Appendix). The left part of the model is used for feature extraction. Fused feature information is shared through skip connections in the middle. After multi-dimensional feature fusion, the obtained information is integrated and extracted through a structure similar to a decoder. Finally, multiple outputs are generated to obtain the final result.

DFP-Net was built upon the original FPN model by using Res101 as the backbone network. Receptive Field Blocks (RFBs) are fused in each layer as part of newly designed Residual RFBs (RRFBs). After the last RRFB, a newly designed SEASPP module is utilized. After each combination of a skip connection and upsampling, a newly designed DOCBAM is used to enhance effective information and suppress unwanted interference. Finally, multiple outputs, containing different features, are integrated into the final result.

3.1 SEASPP

The SEASPP module, utilized by the proposed DFP-Net model, is shown in Figure 2 (Appendix). It was designed based on ASPP [43] to better suit the model. More specifically, in the ASPP pooling stage, we introduced a Squeeze-and-Excitation (SE) attention [44], preceded by adaptive average pooling, which compresses the feature map of each channel to extract features from each channel, followed by a 1×1 convolution. The incorporation of SE, allows the SEASPP module to distinguish extracted feature information through attention weighting, thus emphasizing valuable information while disregarding interfering information.

Given an input feature map X with dimensions $C \times H \times W$, where C is the number of channels, H is the height, and W is the width, for various dilated convolutions to capture information at different scales, the coordinate of the k-th convolution's output feature value $Y_k(p)$ is

$$Y_k(p) = \sum_{i,j} X\,(p + r_k \cdot (i,j)) \cdot W_k(i,j), \quad (1)$$

where r_k denotes the dilation rate of the k-th convolution, and $W_k(i,j)$ denotes the weight of convolutional kernel in the k-th convolution operation.

The output feature maps of different atrous convolutions are concatenated to obtain a stacked feature map as follows:

$$Y(p) = [Y_1(p), Y_2(p), \dots, Y_K(p)]\,. \quad (2)$$

Convolutional operations are performed on the stacked feature map to fuse contextual information across multiple scales. Typically, a 1×1 convolutional kernel is used for feature mapping as follows:

$$Z(p) = \sum_k W_{out}\,(k) \cdot Y_k(p), \quad (3)$$

where $Z(p)$ denotes the fused feature value, and $W_{out}(K)$ denotes the weight for the output channel.

As a result of the above steps, the output feature map of the SEASPP module can be represented as:

$$Z(p) = \sum_k W_{out}\,(k) \cdot (\sum_{i,j} X\,(p + r_k \cdot (i,j)) \cdot W_k(i,j)). \quad (4)$$

The SE attention module [44] utilizes an attention mechanism designed to enhance the performance of CNNs. Its goal is to adaptively adjust the weights between feature channels, allowing the network to focus more on important feature channels and thereby improving the model's feature representation capability. The fundamental idea behind the SE module is to adjust the importance of channels adaptively through "squeeze" and "excitation" operations.

The "squeeze" operation captures global information for each channel through global average pooling. First, the input feature map is subjected to

global average pooling, resulting in a feature map $F_{squeeze}$ with dimensions $1 \times 1 \times C$, representing the global statistical information for each channel.

The "excitation" operation adjusts the importance of channels using learned weights. This operation consists of two key steps – a linear transformation and a sigmoid activation. During the first step, $F_{squeeze}$ is fed into a fully connected layer (or a convolutional layer), producing a feature map F_{excite} with dimensions $1 \times 1 \times C/r$, where r is a hyperparameter referred to as the "reduction ratio" which determines the extent of dimensionality reduction of weights to reduce computation. During the second step, a sigmoid activation function is applied to each element of F_{excite}, mapping the values to the range of 0 to 1, resulting in a feature map $F_{sigmoid}$ with dimensions $1 \times 1 \times C/r$. Then, a rescaling operation is performed by multiplying the values of each channel in $F_{sigmoid}$ by the corresponding channel in the input feature map X, achieving adaptive weight adjustment between channels. At the end, the output feature map of the SE module can be represented as:

$$X_{SE} = X \otimes F_{sigmoid}\ , \quad (5)$$

where $\otimes$ denotes an element-wise multiplication.

The newly designed SEASPP module has a multi-scale receptive field, effectively capturing multi-scale contextual information, which enables it to better handle objects of varied sizes. Additionally, it can learn the overall image background and semantic information, providing a better understanding of the context. It filters and integrates information within it, allowing useful features to be expressed more effectively while suppressing irrelevant features. Moreover, it can be easily incorporated into various network architectures as needed.

3.2 DOCBAM

DOCBAM is built upon the Convolutional Block Attention Module (CBAM) [45] and incorporates Depthwise Over-parameterized Convolution (DOConv) to accelerate convergence and improve network performance. DOCBAM consists of two key components – channel attention and spatial attention, as shown in Figure 3 (Appendix).

The channel attention submodule ('DOChannel att'), shown in Figure 4 (Appendix), is used to adjust the weights between different channels so as to emphasize important channel information.

The computation steps for channel attention are the following. First, global average pooling is performed on the input feature map X, resulting in a feature map F_{avg} (with dimensions $1 \times 1 \times C$), which is passed through two separate fully connected layers to obtain channel attention weights, which are then applied to each pixel. The feature map adjusted by channel attention can be represented as:

$$X_c = X \otimes \sigma\left(W_c^{(2)}(\delta(W_c^{(1)}(F_{avg})))\right), \quad (6)$$

where σ denotes the sigmoid activation function, and $W_c^{(1)}(.)$ and $W_c^{(2)}(.)$ denote the two fully connected layers for channel attention.

The spatial attention submodule ('Spatial att' in Figure 3, Appendix) is used to adjust the weights of different spatial positions, focusing on important spatial regions. Its computation steps are the following. First, two separate convolution operations are performed on the input feature map X – one in the horizontal direction and the other in the vertical direction – to capture spatial information in different directions. Then, the two feature maps obtained by the convolution operations are added together and passed through a convolutional layer, resulting in a feature map F_{max} (with dimensions $1 \times H \times W$), which is passed through a fully connected layer to obtain spatial attention weights, which are then applied to each channel. The feature map adjusted by spatial attention can be represented as:

$$X_s = X \otimes \sigma\big(g(\delta(W_s(\mathrm{Conv}(F_{max}))))\big), \quad (7)$$

where $Conv$ denotes the convolution operation, g denotes the fully connected layer, and δ denotes the ReLU activation function.

Finally, the results of channel attention and spatial attention are combined, as follows:

$$X_{DOCBAM} = X_c \otimes X_s\ . \quad (8)$$

By combining CBAM with DOConv, DOCBAM can further extract important features that are needed and reduce the risk of model overfitting that can occur after adding linear layers. Therefore, DOCBAM plays a crucial role in each top-down process, enhancing network representation and improving the model performance.

3.3 RRFB

The newly designed RRFB component of the proposed DFP-Net model is built upon RFB [46], which intends to enhance the network's feature extraction capabilities by simulating human visual perception, drawing inspiration from the issue of

object scale variation in object detection. Different objects might appear at different scales within an image, necessitating the model's ability to capture features at various object scales. RFB exhibits two main characteristics: (1) having a multi-branch structure composed of convolutional layers with different-size kernels, drawing inspiration from the Inception structure; and (2) using dilated convolutional layers, which were previously applied in the segmentation algorithm Deeplab, primarily to increase the receptive field. Different rates are used to represent the parameters of dilated convolutional layers. The outputs of convolutional layers with varied sizes and rates are concatenated, achieving the goal of fusing different features.

The newly designed RRFB component of the proposed DFP-Net model combines each layer of ResNet101 with RFB (Figure 5, Appendix). This combination enhances the grasp of global features by increasing the receptive field for the deeper information extracted through each residual block, allowing the model to extract as much information as possible.

4 Experiments and Results

4.1 Datasets

In the experiments, three datasets were used – two public datasets (ISIC2018 and PH2) and a private dataset. The International Skin Imaging Collaboration Challenge 2018 (ISIC2018) public dataset [47] contains 2594 skin lesion images from both clinical and non-clinical environments, covering various skin types and disease conditions, with the goal of facilitating CAD and research of skin lesions. The PH2 public dataset [48] originates from a research project at a university medical center in Portugal and contains 200 skin-lesion images of diverse types and severity levels, widely used for skin lesion analysis, primarily focusing on pigmentary lesions. The private dataset [22], provided by Peking Union Medical College Hospital, includes 1010 skin lesion images of acne and lupus erythematosus.

4.2 Evaluation Metrics

For performance comparison of the proposed DFP-Net model with state-of-the-art models, multiple widely accepted metrics, each representing distinct aspects of the segmentation performance, were used. These are briefly described below.

The Intersection over Union (IoU) is a commonly used evaluation metric in segmentation tasks, measuring the degree of overlap between the predicted region and the ground-truth region. IoU is typically used to evaluate the performance of object detection and image segmentation models, as follows:

$$IoU = \frac{TP}{TP+FP+FN}\ , \tag{9}$$

where *TP* (True Positive) represents the number of correctly identified pixels as part of the object (in our case, skin lesions), *FP* (False Positive) represents the number of incorrectly identified pixels as part of the object, and *FN* (False Negative) represents the number of incorrectly identified pixels as not belonging to the object.

The Dice coefficient is another commonly used metric in image segmentation tasks. It measures the similarity between the model's prediction and the ground-truth region. It is a more sensitive metric, particularly suitable for sparse objects, and calculated as follows:

$$Dice = \frac{2TP}{2TP+FP+FN}\,. \tag{10}$$

Precision is a metric used in classification problems to measure how many of the predicted positive samples are truly positive. It calculates the ratio of the number of samples correctly predicted as positive to the total number of samples predicted as positive, as follows:

$$Pre = \frac{TP}{TP+FP} \tag{11}$$

Accuracy is a metric used in classification problems to measure the proportion of samples that are correctly classified by the model among all samples, as follows:

$$Acc = \frac{TP+TN}{TP+FP+FN+TN} \tag{12}$$

where *TN* (True Negative) represents healthy skin regions in our case.

4.3 Experimental Setup

The Python version 3.9 and Windows 10 operating system were used on a computer equipped with a 12th generation Intel® Core™ i5-13490 CPU, and an NVIDIA GeForce RTX 3060 (12GB memory). The initial learning rate was set to 1e-4, with a minimum learning rate of 1e-5. The batch size was set to 8, and the number of epochs was 100. The input images' dimension was 256 x 256. A combined BCE-Dice loss function [49] was used.

4.4 Results and Analysis

4.4.1 On ISIC2018 dataset

First, melanoma segmentation performance comparison experiments were conducted on the ISIC2018 public dataset [47], divided into training, validation, and test sets using a ratio of 7:1:2.

In the conducted experiments, the proposed model was compared to state-of-the-art models, namely U-Net [16], UNet++ [17], UNet 3+ [20], AttentionUNet [50], DoubleU-Net [18], SegNet [21], and UResnet [51]. The results were obtained through ten rounds of testing, where each validation image was shuffled randomly for each test, and the average values were recorded as shown in Table 1 (the best result on each metric is shown in **bold**).

Table 1. Segmentation performance comparison with state-of-the-art models, based on experiments conducted on the ISIC2018 dataset

Models	IoU, %	Acc, %	Dice, %	Pre, %
U-Net	0.8169	0.9576	0.8838	0.9268
UNet++	0.8203	0.9587	0.8881	**0.9272**
UNet 3+	0.8215	0.9585	0.8867	0.9249
AttentionUNet	0.8217	0.9588	0.8863	0.9182
DoubleU-Net	0.8173	0.9573	0.8834	0.9125
SegNet	0.8258	0.9598	0.8911	0.9269
UResnet	0.8287	0.9607	0.8940	0.9193
DFP-Net (*proposed model*)	**0.8501**	**0.9653**	**0.9074**	0.9169

Source: created by the authors

As evident from Table 1, the proposed DFP-Net model achieves an IoU value of 0.8501, which is 0.0214 higher than the second-best performing model (UResnet) on this metric. The DFP-Net accuracy reaches 0.9653, with a margin of 0.0046 over the second-best performing model (UResnet). The Dice coefficient of the proposed model reaches a value of 0.9074, surpassing the second-best performing model (UResnet) by 0.0134. Only according to precision, the proposed model does not perform so well, taking seventh place.

The loss curves of the compared models are depicted in Figure 6 (Appendix). Loss reflects the discrepancy between the model's predictions and the ground truth values. From Figure 6 (Appendix), one can observe that the proposed DFP-Net model consistently exhibits lower loss compared to other models, both during training and validation. This closeness to the ground truth values further validates the superior performance of the proposed model.

Figure 7 (Appendix) visualizes sample segmentation results of the compared models, based on the ISIC2018 dataset.

Next, the proposed model was compared to other state-of-the-art models, based on their results reported in the literature, as summarized in [22]. Results, shown in Table 2 (the best result on each metric is shown in **bold**), demonstrate that the proposed DFP-Net model surpasses all other models in terms of IoU and accuracy, and takes third place according to the Dice coefficient and precision.

Table 2. Segmentation performance comparison with state-of-the-art models, based on results reported in literature for ISIC2018 dataset

Models	IoU, %	Acc, %	Dice, %	Pre, %
Improved U-Net++ [38]	0.8250	-	0.8680	0.9030
MALUNet [52]	0.8025	0.9462	0.8904	-
Ms RED [53]	0.8345	0.9619	0.8999	0.9147
ICL-Net [54]	0.8390	0.9440	0.9030	-
TransCeption [55]	-	0.9628	0.9124	-
M-CSAFN [56]	0.8364	0.9645	0.9031	0.9171
CACDU-Net [22]	0.8427	0.9641	**0.9134**	**0.9252**
DFP-Net (*proposed model*)	**0.8501**	**0.9653**	0.9074	0.9169

Source: created by the authors

4.4.2 On combined Public–Private Dataset

To validate the generalizability of the proposed DFP-Net model, additional training was performed on a combined dataset, containing the images of the PH2 public dataset [48] and the private dataset [22]. (Since the PH2 dataset contains only 200 images and is insufficient for standalone training, we combined it with the private dataset containing 1010 images.) We then split the combined dataset into training, validation, and test sets in a ratio of 871:164:175.

In these experiments, the proposed model was compared to the same set of state-of-the-art models, participating in the first experiment. The results were obtained through ten rounds of testing, where each validation image was shuffled randomly for each test, and the average values were recorded as shown in Table 3 (the best result on each metric is shown in **bold**).

As one can observe from Table 3, the proposed DFP-Net model is superior to all other models according to IoU, accuracy, and Dice coefficient, which demonstrates its exceptional generalization ability. More specifically, DFP-Net achieves an IoU value of 0.7579, which is 0.0143 higher than the second-best performing model (U-Net) on this

metric. The DFP-Net accuracy reaches 0.9249, with a margin of 0.0009 over the second-best performing model (U-Net). The Dice coefficient of the proposed model reaches a value of 0.8406, surpassing the second-best performing model (U-Net) by 0.0098. Only, according to precision, the proposed model is not performing well.

Figure 8 (Appendix) visualizes sample segmentation results of the models, compared on the combined dataset.

Table 3. Segmentation performance comparison with state-of-the-art models, based on experiments conducted on the combined public–private dataset

Models	IoU, %	Acc, %	Dice, %	Pre, %
U-Net	0.7436	0.9240	0.8308	0.9062
UNet++	0.7052	0.9097	0.8030	0.8842
UNet 3+	0.7039	0.9107	0.8080	0.9013
AttentionUNet	0.7100	0.9130	0.8076	0.8984
DoubleU-Net	0.7330	0.9191	0.8227	0.8936
SegNet	0.7185	0.9154	0.8177	0.9015
UResnet	0.7225	0.9163	0.8183	**0.9119**
DFP-Net (*proposed model*)	**0.7579**	**0.9249**	**0.8406**	0.8751

Source: created by the authors

4.4.3 Ablation Study

To assess the impact of each main module on the segmentation performance of the proposed DFP-Net model, we conducted ablation study experiments on the ISIC2018 dataset using FPN as a baseline. The obtained results are presented in Table 4 (the best result on each metric is shown in **bold**).

Table 4. Ablation study results

Steps	IoU, %	Acc, %	Dice, %	Pre, %
FPN	0.8319	0.9611	0.9009	0.9099
FPN + SEASPP	0.8380	0.9627	0.9041	0.9184
FPN + DOCBAM	0.8349	0.9621	0.8998	**0.9328**
FPN + RRFB	0.8429	0.9641	0.9019	0.9193
FPN + SEASPP + RRFB	0.8391	0.9627	0.9065	0.9124
FPN + SEASPP + DOCBAM	0.8436	0.9635	0.9071	0.9065
FPN + RRFB + DOCBAM	0.8406	0.9630	0.9069	0.9067
SEASPP + RRFB + DOCBAM (i.e., DFP-Net model)	**0.8501**	**0.9653**	**0.9074**	0.9169

Source: created by the authors

Even though there were fluctuations in some metrics at certain stages of the ablation study, overall, the process of iterative addition of each new module was on an upward trajectory. Generally, each module contributed to improving the segmentation performance. While some improvements may not be substantial, the performance of the final model shows significant enhancement when compared to the baseline.

5 Conclusion and Future Research Directions

This paper has proposed a novel Deep Feature Pyramid Network (DFP-Net) model for melanoma segmentation, based on the Feature Pyramid Network (FPN), which demonstrated outstanding performance compared to state-of-the-art models on a public and a combined public–private datasets, confirming its generalization ability. DFP-Net can significantly ease the workload of medical professionals and empower individuals to conduct self-examinations anytime-anywhere-anyhow, thus increasing awareness of their own conditions and enabling earlier treatment. The fusion of deep learning (DL) and medical image processing holds great promise for the future, particularly with the application of DFP-Net in promoting the early detection and treatment of melanoma.

Future plans involve refining the connections of the model's long-range dependencies and innovatively combining it with U-Net to achieve higher segmentation accuracy. We will also explore the use of more advanced network architectures as backbones, replacing ResNet101 with networks like TransUNet to further improve the model's performance. We will persist in enhancing precision values comprehensively. In addition, we will aim to enhance and adapt the proposed model not only for skin melanoma segmentation but also for use in other domains such as liver tumor segmentation, heart segmentation, lung segmentation, etc.

We should approach this with the spirit of scientific inquiry, step by step, applying each innovation to real-world medical applications. Through generations of effort, we can eventually achieve excellence and make meaningful advancements in the field of healthcare.

Declaration of Generative AI and AI-assisted Technologies in the Writing Process

During the preparation of this work, the author(s) utilized artificial intelligence (AI) technologies solely for the purpose of linguistic refinement and grammatical correction. Following the application of these tools, the author(s) comprehensively

reviewed, verified, and edited the resulting text. The author(s) maintain full accountability and responsibility for the accuracy, integrity, and final content of the publication.

References:

[1] M. D. J. I. A. Alahmadi, "Multiscale attention U-Net for skin lesion segmentation," vol. 10, pp. 59145-59154, 2022.

[2] M. F. Jojoa Acosta, L. Y. Caballero Tovar, M. B. Garcia-Zapirain, and W. S. J. B. M. I. Percybrooks, "Melanoma diagnosis using deep learning techniques on dermatoscopic images," vol. 21, no. 1, pp. 1-11, 2021.

[3] A. K. Nambisan *et al.*, "Improving automatic melanoma diagnosis using deep learning-based segmentation of irregular networks," vol. 15, no. 4, p. 1259, 2023.

[4] A. Pennisi, D. D. Bloisi, V. Suriani, D. Nardi, A. Facchiano, and A. R. J. J. o. D. I. Giampetruzzi, "Skin lesion area segmentation using attention squeeze U-Net for embedded devices," vol. 35, no. 5, pp. 1217-1230, 2022.

[5] X. Bian, H. Pan, K. Zhang, C. Chen, P. Liu, and K. J. E. Shi, "NeDSeM: Neutrosophy domain-based segmentation method for malignant melanoma images," vol. 24, no. 6, p. 783, 2022.

[6] M. A. Khan, M. Sharif, T. Akram, R. Damaševičius, and R. J. D. Maskeliūnas, "Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization," vol. 11, no. 5, p. 811, 2021.

[7] S. Mustafa *et al.*, "Entropy and Gaussian Filter-Based Adaptive Active Contour for Segmentation of Skin Lesions," vol. 2022, 2022.

[8] M. Rehman *et al.*, "Machine learning based skin lesion segmentation method with novel borders and hair removal techniques," vol. 17, no. 11, p. e0275781, 2022.

[9] R. Ramadan and S. J. I. A. Aly, "CU-net: a new improved multi-input color U-net model for skin lesion semantic segmentation," vol. 10, pp. 15539-15564, 2022.

[10] F. Yalcinkaya and A. J. E. i. E. Erbas, "Convolutional Neural Network and Fuzzy Logic-based Hybrid Melanoma Diagnosis System," vol. 27, no. 2, pp. 55-63, 2021.

[11] Y. Jiang, H. Qiao, Z. Zhang, M. Wang, W. Yan, and J. J. I. I. P. Chen, "MDSC- Net: A multi- scale depthwise separable convolutional neural network for skin lesion segmentation," 2023.

[12] S. Nofallah *et al.*, "Improving the diagnosis of skin biopsies using tissue segmentation," vol. 12, no. 7, p. 1713, 2022.

[13] S. Nofallah *et al.*, "Segmenting skin biopsy images with coarse and sparse annotations using U-Net," vol. 35, no. 5, pp. 1238-1249, 2022.

[14] S. K. Singh, V. Abolghasemi, and M. H. J. A. S. Anisi, "Fuzzy Logic with Deep Learning for Detection of Skin Cancer," vol. 13, no. 15, p. 8927, 2023.

[15] D. Shao, L. Ren, and L. J. B. Ma, "MSF-Net: A Lightweight Multi-Scale Feature Fusion Network for Skin Lesion Segmentation," vol. 11, no. 6, p. 1733, 2023.

[16] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015: 18th International Conference, Munich, Germany, October 5-9, 2015, Proceedings, Part III 18*, 2015: Springer, pp. 234-241.

[17] Z. Zhou, M. M. Rahman Siddiquee, N. Tajbakhsh, and J. Liang, "Unet++: A nested u-net architecture for medical image segmentation," in *Deep Learning in Medical Image Analysis and Multimodal Learning for Clinical Decision Support: 4th International Workshop, DLMIA 2018, and 8th International Workshop, ML-CDS 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 20, 2018, Proceedings 4*, 2018: Springer, pp. 3-11.

[18] D. Jha, M. A. Riegler, D. Johansen, P. Halvorsen, and H. D. Johansen, "Doubleu-net: A deep convolutional neural network for medical image segmentation," in *2020 IEEE 33rd International symposium on computer-based medical systems (CBMS)*, 2020: IEEE, pp. 558-564.

[19] A. Kirillov, R. Girshick, K. He, and P. Dollár, "Panoptic feature pyramid networks," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2019, pp. 6399-6408.

[20] H. Huang *et al.*, "Unet 3+: A full-scale connected unet for medical image segmentation," in *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing (ICASSP)*, 2020: IEEE, pp. 1055-1059.

[21] V. Badrinarayanan, A. Kendall, R. J. I. t. o. p. a. Cipolla, and m. intelligence, "Segnet: A deep convolutional encoder-decoder architecture for

image segmentation," vol. 39, no. 12, pp. 2481-2495, 2017.
[22] S. Hao *et al.*, "CACDU-Net: A novel DoubleU-Net based semantic segmentation model for skin lesions detection in images," 2023.
[23] U. Bhimavarapu and G. Battineni, "Skin Lesion Analysis for Melanoma Detection Using the Novel Deep Learning Model Fuzzy GC-SCNN," in *Healthcare*, 2022, vol. 10, no. 5: MDPI, p. 962.
[24] S. Joseph and O. O. Olugbara, "Preprocessing effects on performance of skin lesion saliency segmentation," vol. 12, no. 2, p. 344, 2022.
[25] S. Liu *et al.*, "DiaMole: Mole detection and segmentation software for mobile phone skin images," vol. 2021, pp. 1-10, 2021.
[26] M. Nawaz *et al.*, "MSeg-Net: A Melanoma Mole Segmentation Network Using CornerNet and Fuzzy-Means Clustering," vol. 2022, 2022.
[27] Z. Song, W. Luo, and Q. J. E. Shi, "Res-CDD-net: A network with multi-scale attention and optimized decoding path for skin lesion segmentation," vol. 11, no. 17, p. 2672, 2022.
[28] Y. Li *et al.*, "MHAU-net: Skin lesion segmentation based on multi-scale hybrid residual attention network," vol. 22, no. 22, p. 8701, 2022.
[29] R. Kaur, H. GholamHosseini, R. Sinha, and M. J. B. M. I. Lindén, "Automatic lesion segmentation using atrous convolutional deep neural networks in dermoscopic skin cancer images," vol. 22, no. 1, pp. 1-13, 2022.
[30] R. Katmah, F. Al-Shargie, U. Tariq, F. Babiloni, F. Al-Mughairbi, and H. J. S. Al-Nashash, "A review on mental stress assessment methods using EEG signals," vol. 21, no. 15, p. 5043, 2021.
[31] M. Nour, H. Öcal, A. Alhudhaif, and K. J. M. P. i. E. Polat, "Skin lesion segmentation based on edge attention vnet with balanced focal tversky loss," vol. 2022, 2022.
[32] S. Oukil, R. Kasmi, K. Mokrani, B. J. S. R. García- Zapirain, and Technology, "Automatic segmentation and melanoma detection based on color and texture features in dermoscopic images," vol. 28, no. 2, pp. 203-211, 2022.
[33] M. Zafar, J. Amin, M. Sharif, M. A. Anjum, G. A. Mallah, and S. J. M. Kadry, "DeepLabv3+-based segmentation and best features selection using slime mould algorithm for multi-class skin lesion classification," vol. 11, no. 2, p. 364, 2023.
[34] X. Bian, H. Pan, K. Zhang, P. Liu, and C. J. I. I. P. Chen, "Malignant melanoma dermoscopy image classification method based on multi- modal medical features," 2023.
[35] T.-D.-T. Phan, S.-H. Kim, H.-J. Yang, and G.-S. J. A. s. Lee, "Skin lesion segmentation by u-net with adaptive skip connection and structural awareness," vol. 11, no. 10, p. 4528, 2021.
[36] T. Küstner *et al.*, "Development of a Hybrid-Imaging-Based Prognostic Index for Metastasized-Melanoma Patients in Whole-Body 18F-FDG PET/CT and PET/MRI Data," vol. 12, no. 9, p. 2102, 2022.
[37] F. Alenezi, A. Armghan, and K. J. D. Polat, "A Novel Multi-Task Learning Network Based on Melanoma Segmentation and Classification with Skin Lesion Images," vol. 13, no. 2, p. 262, 2023.
[38] Q. Zhou, T. He, and Y. J. D. Zou, "Superpixel-oriented label distribution learning for skin lesion segmentation," vol. 12, no. 4, p. 938, 2022.
[39] M. P. Schilling *et al.*, "Automated annotator variability inspection for biomedical image segmentation," vol. 10, pp. 2753-2765, 2022.
[40] P. Chen, S. Huang, and Q. J. I. A. Yue, "Skin lesion segmentation using recurrent attentional convolutional networks," vol. 10, pp. 94007-94018, 2022.
[41] T. Vaiyapuri, P. Balaji, H. Alaskar, Z. J. C. I. Sbai, and Neuroscience, "Computational Intelligence-Based Melanoma Detection and Classification Using Dermoscopic Images," vol. 2022, 2022.
[42] P. T. Le, C.-C. Chang, Y.-H. Li, Y.-C. Hsu, J.-C. J. W. C. Wang, and M. Computing, "Antialiasing attention spatial convolution model for skin lesion segmentation with applications in the medical IoT," vol. 2022, 2022.
[43] S. C. Yurtkulu, Y. H. Şahin, and G. Unal, "Semantic segmentation with extended DeepLabv3 architecture," in *2019 27th Signal Processing and Communications Applications Conference (SIU)*, 2019: IEEE, pp. 1-4.
[44] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132-7141.
[45] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3-19.
[46] S. Liu and D. Huang, "Receptive field block net for accurate and fast object detection," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 385-400.

[47] N. Codella *et al.*, "Skin lesion analysis toward melanoma detection 2018: A challenge hosted by the international skin imaging collaboration (isic)," 2019.
[48] T. Mendonça, P. M. Ferreira, J. S. Marques, A. R. Marcal, and J. Rozeira, "PH 2-A dermoscopic image database for research and benchmarking," in *2013 35th annual international conference of the IEEE engineering in medicine and biology society (EMBC)*, 2013: IEEE, pp. 5437-5440.
[49] X. Huang, L. Li, R. Liu, C. Xu, and M. J. O. Ye, "Binarization of degraded document images with global-local U-Nets," vol. 203, p. 164025, 2020.
[50] O. Oktay *et al.*, "Attention U-net: Learning where to look for the pancreas," 2018.
[51] M. F. Rachmadi, M. del C. Valdés-Hernández, and T. Komura, "Transfer learning for task adaptation of brain lesion assessment and prediction of brain abnormalities progression/regression using irregularity age map in brain mri," in *PRedictive Intelligence in MEdicine: First International Workshop, PRIME 2018, Held in Conjunction with MICCAI 2018, Granada, Spain, September 16, 2018, Proceedings 1*, 2018: Springer, pp. 85-93.
[52] J. Ruan, S. Xiang, M. Xie, T. Liu, and Y. Fu, "MALUNet: A multi-attention and light-weight unet for skin lesion segmentation," in *2022 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*, 2022: IEEE, pp. 1150-1156.
[53] D. Dai *et al.*, "Ms RED: A novel multi-scale residual encoding and decoding network for skin lesion segmentation," vol. 75, p. 102293, 2022.
[54] W. Cao *et al.*, "ICL-Net: Global and local inter-pixel correlations learning network for skin lesion segmentation," vol. 27, no. 1, pp. 145-156, 2022.
[55] R. Azad, Y. Jia, E. K. Aghdam, J. Cohen-Adad, and D. J. a. p. a. Merhof, "Enhancing Medical Image Segmentation with TransCeption: A Multi-Scale Feature Fusion Approach," 2023.
[56] J. Mu *et al.*, "M-CSAFN: Multi-color Space Adaptive Fusion Network for Automated Port-wine Stains Segmentation," 2023.

Contribution of Individual Authors to the Creation of a Scientific Article (Ghostwriting Policy)

The authors equally contributed to the present research, at all stages from the formulation of the problem to the final findings and solution.

Sources of Funding for Research Presented in a Scientific Article or Scientific Article Itself

This publication has emanated from research conducted with the financial support of the Zhejiang A&F University Research Startup Funds (203402004501) and realized in part with the help of the infrastructure purchased under the Bulgarian National Roadmap for Scientific Infrastructure, financially coordinated by the Ministry of Education and Science of the Republic of Bulgaria under Grant D01-98/26.06.2025

Conflict of Interest

The authors have no conflicts of interest to declare that are relevant to the content of this article.

Creative Commons Attribution License 4.0 (Attribution 4.0 International, CC BY 4.0)

This article is published under the terms of the Creative Commons Attribution License 4.0
https://creativecommons.org/licenses/by/4.0/deed.en_US

APPENDIX

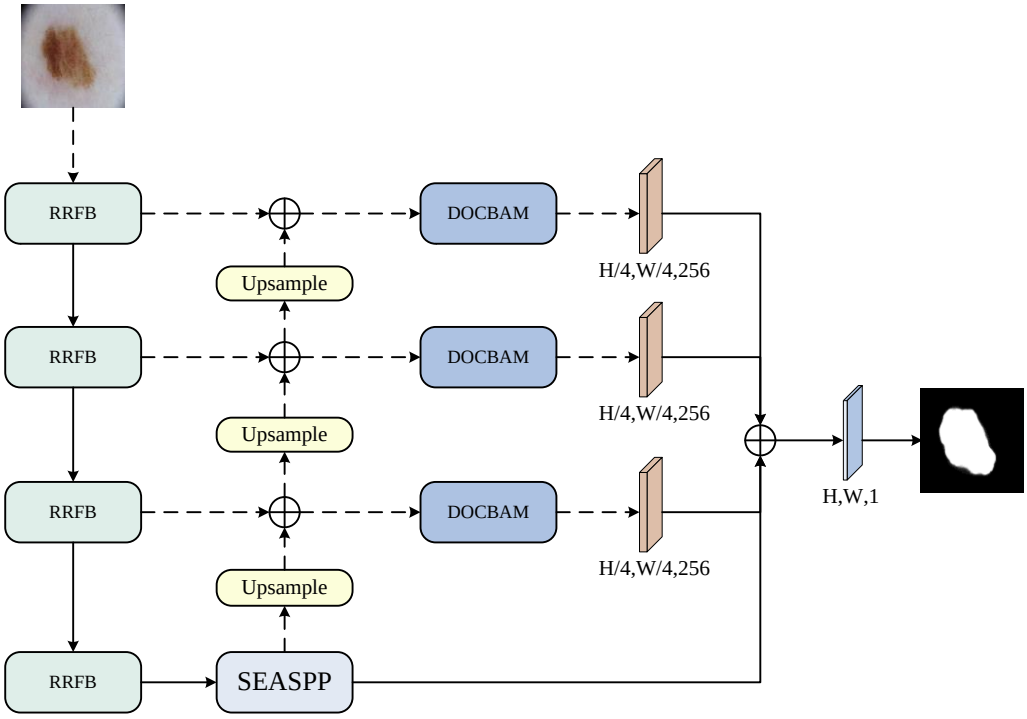


Fig. 1: The proposed DFP-Net model
Source: [47] *(top-left image); created by the authors (remaining parts)*

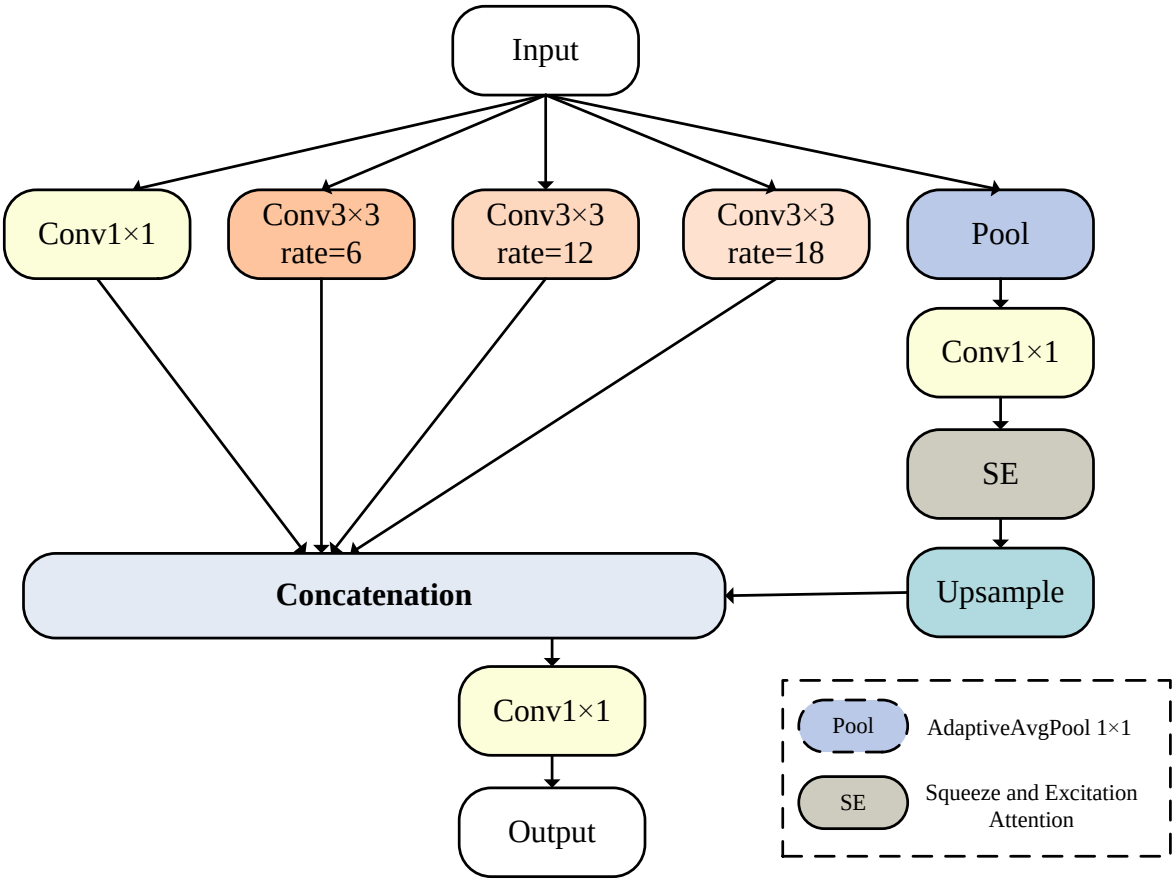


Fig. 2: The newly designed SEASPP module, utilized by the proposed DFP-Net model
Source: created by the authors

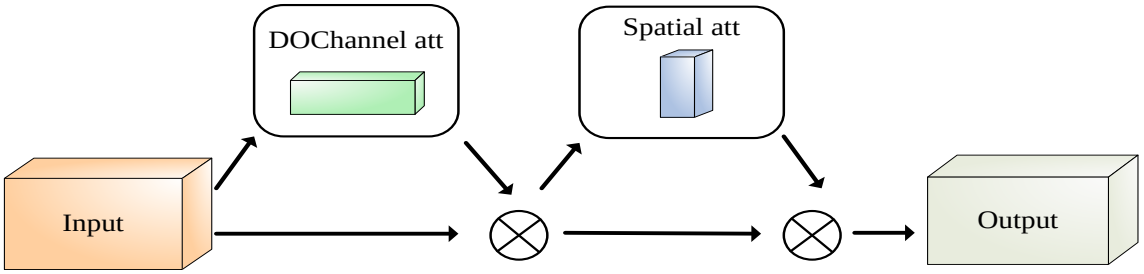


Fig. 3: The newly designed DOCBAM component, utilized by the proposed DFP-Net model
Source: created by the authors

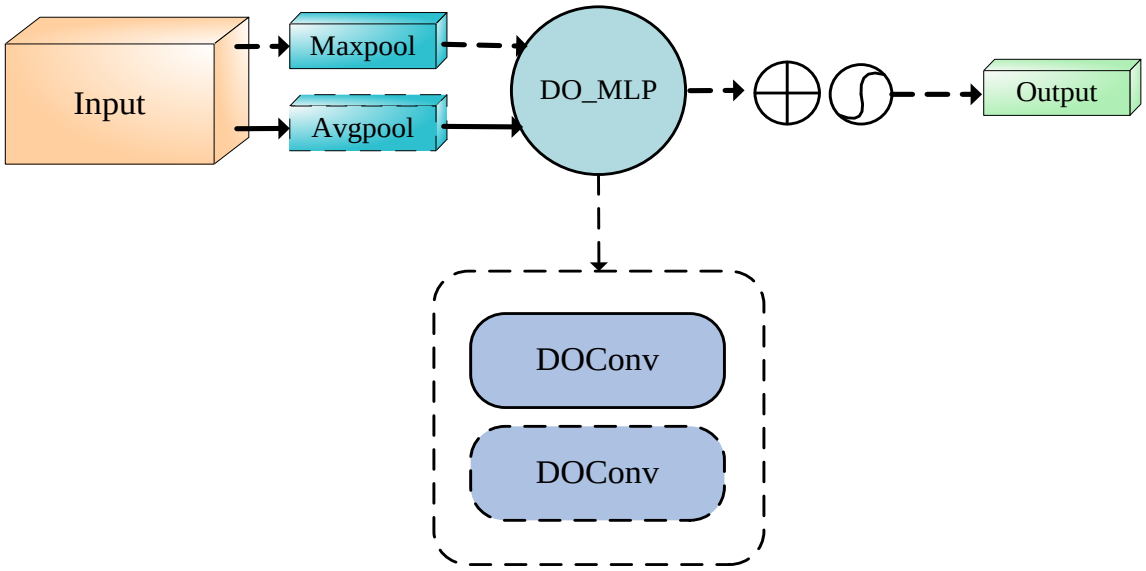


Fig. 4: The 'DOChannel att' submodule, utilized by DOCBAM
Source: created by the authors

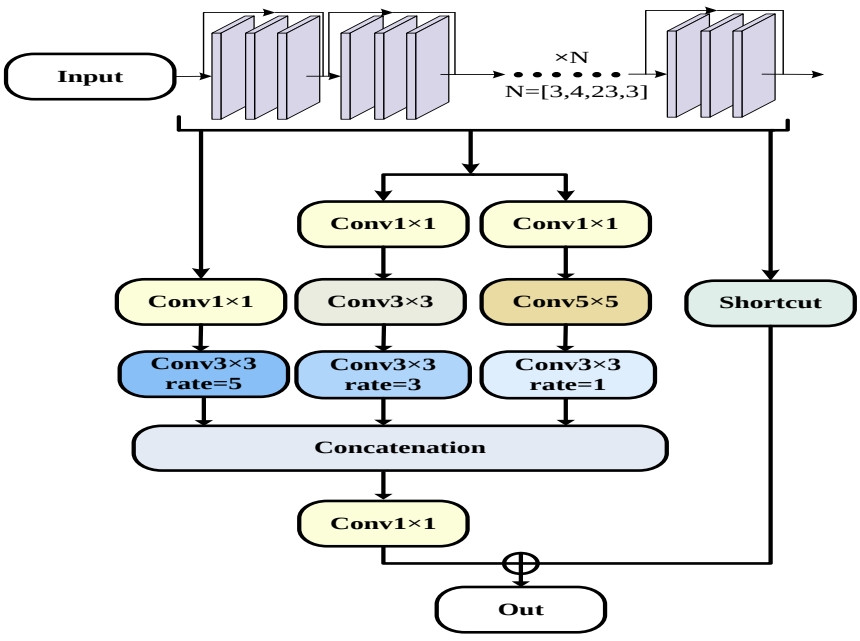


Fig. 5: The newly designed RRFB component, utilized by the proposed DFP-Net model
Source: created by the authors

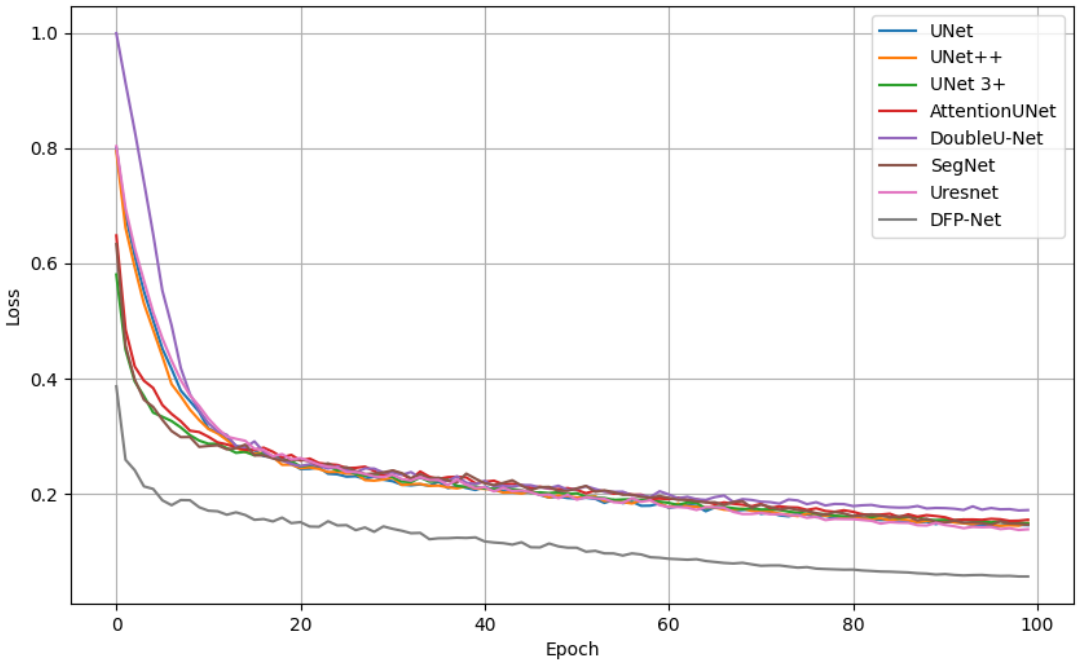


Fig. 6: The loss curves of models, compared in Table 1
Source: created by the authors

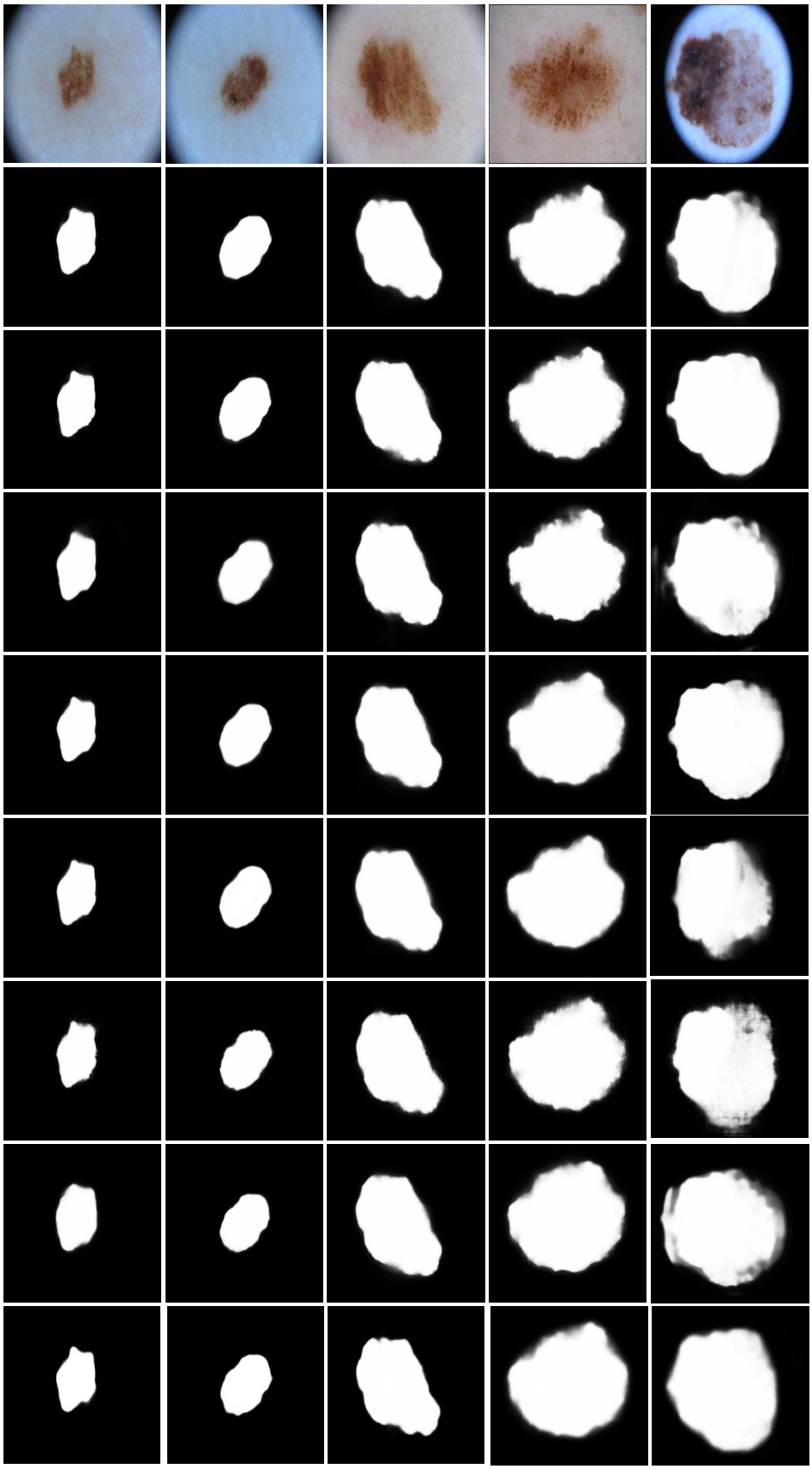


Fig. 7: Sample segmentation results of models, compared on the ISIC2018 dataset (from top to bottom: U-Net, UNet++, UNet 3+, AttentionUNet, DoubleU-Net, SegNet, UResnet, DFP-Net)

Source: [47] *(top row); created by the authors (remaining rows)*

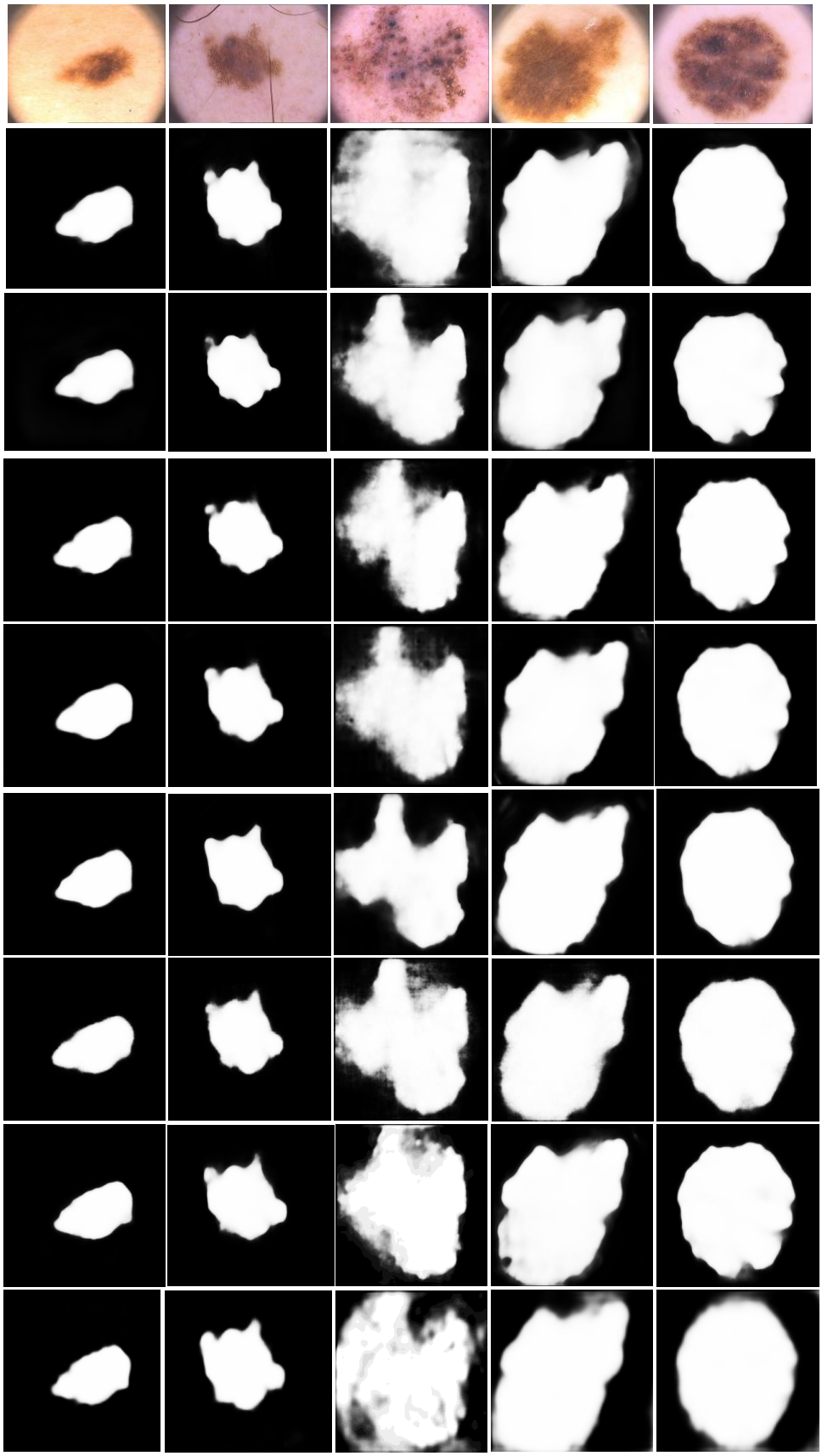


Fig. 8: Sample segmentation results of models, compared on the combined public–private dataset:
(from top to bottom: U-Net, UNet++, UNet 3+, AttentionUNet, DoubleU-Net, SegNet, UResnet, and DFP-Net)
Source: [47] *(top row); created by the authors (remaining rows)*