Vol. 34, No. 4, 2025
(this issue contains accepted papers online and is not closed yet)

Machine GRAPHICS & VISION

International Journal

Published by
The Institute of Information Technology
Warsaw University of Life Sciences – SGGW
Nowoursynowska 159, 02-776 Warsaw, Poland
in cooperation with
The Association for Image Processing, Poland – TPO

Skin Lesion Segmentation using SegNet with Spatial Attention

Maryam Arif[1], Almas Abbasi[1], Muhammad Arif[2] and Muhammad Rashid[3]
[1]*Department of Computer Science, Faculty of Computing and Information Technology, International Islamic University, Islamabad, Pakistan*
[2]*Department of Computer Science and Artificial Intelligence, College of Computing, Umm Al-Qura University, Makkah, Saudi Arabia*
[3]*Department of Computer and Network Engineering, College of Computing, Umm Al-Qura University, Makkah, Saudi Arabia*
[*]*Corresponding author: Muhammad Arif (mahamid@uqu.edu.sa)*

Submitted: Jan 2, 2025 **Accepted:** May 20, 2025 **Published:** Nov 5, 2025

Abstract Skin lesion segmentation identifies and outlines the boundaries of abnormal skin regions. Accurate segmentation may help in the early detection of skin cancer. Accurate Skin Lesion Segmentation is still challenging due to different skin color tones, variations in shape, and body hairs. Moreover, variability in the lesion appearance, quality of the images, and lack of clear skin boundaries make the problem even harder. This paper proposes a SegNet model with spatial attention mechanisms for skin lesion segmentation. Adding one component of spatial attention to SegNet allows the proposed model to focus more on specific parts across the image, eventually leading to a better segmentation of the lesion boundary. The proposed model was evaluated on the ISIC 2018 dataset. Our proposed model attained an average accuracy of 96.25%, and the average dice coefficient equals 0.9052. The model's performance indicates its possible application in automated skin disease diagnosis.

Keywords: skin lesion segmentation; deep learning; spatial attention; SegNet.

1. Introduction

Skin is the largest organ of the human body that is usually directly exposed to the air. In other words, it is the most vulnerable organ due to its exposure to ultraviolet rays from the Sun and other environmental toxins. It leads to various skin diseases, including skin cancer [28]. According to the International Agency for Research on Cancer (IARC), approximately 3 330 000 new cases of skin cancer were diagnosed worldwide in 2022 [14]. Moreover, almost 60 000 people died from the disease. Furthermore, the IARC has observed that there are 5.4 million new cases of skin cancer every year [31]. Therefore, the World Health Organisation ranks skin cancer as one of the most prevalent and fastest-growing cancers globally [12].

The cause of skin cancer is the proliferation or formation of skin cells unevenly or abnormally. Depending on their type and strength, this proliferation of skin cells can infiltrate or disseminate to other areas of the body. Based on different skin cells, the three important types of skin cancers are basal cell skin cancer, squamous cell skin cancer, and Melanoma. Physicians use these abnormal cells to determine the type of

skin cancer [2, 13, 21]. Basal cell skin cancer and squamous cell skin cancer are less dangerous as they hardly result in death. However, the most dangerous type of cancer is Melanoma, accounting for around 75% of deaths attributed to skin cancer. Its formation starts in melanin-producing cells that develop in melanocytes [11].

Even though Melanoma, a frequently occurring skin cancer, is lethal and the death rate of this disease is very high, it is easily curable if the detection is made in its early stages. According to [14], the in-time diagnosis of Melanoma decreases the mortality rate by 90%. Some other studies reveal that there is a 95% early diagnosis (stage I of the disease) survival rate and a 20% late discovery rate (stage IV of the disease) [19, 31]. It implies that early detection increases the chances of survival and improves therapy efficacy. For this reason, it is critical to diagnose and treat dermatoses as soon as possible.

One of the conventional methods for the diagnosis of melanoma and other skin cancer types is the biopsy. This procedure involves taking a sample from a suspected skin lesion to perform medical tests and determine if it is cancerous. However, undergoing a biopsy can be challenging as it involves extracting a sample of the lesion. It can be uncomfortable and requires time for the procedure and the subsequent analysis. The alternative to biopsy is the visual assessment of skin lesions. Since pigmented lesions are visible on the skin's surface, a skilled visual examination can often detect Melanoma at an early stage. It often involves ABCD Scale [13] that evaluates asymmetry, border irregularity, color variegation, and lesion diameter. The ABCDE Scale [6, 24] is an extension of the ABCD scale and adds evolving to account for changes in the lesion over time. Similarly, Glasgow 7-point Checklist [8] includes major criteria such as a change in size, shape, and color, along with minor criteria like inflammation, crusting or bleeding, sensory changes, and the diameter of the lesion. These algorithms provide a structured approach to assess skin lesions and help in the early detection of Melanoma by identifying key warning signs.

Dermatologists often use a dermatoscope to enhance the visibility of skin lesions by magnifying them with light. This enhanced visibility allows dermatologists to detect early Melanoma that might be invisible to the naked eye. While dermoscopy increases detection accuracy, the complexity of skin lesions and the sheer volume of dermoscopic images make visual inspection potentially non-reproducible, time-consuming, and subjective in medical practice. That is why advancements in computer-aided diagnostic systems have become so important, offering more consistent and efficient analysis of dermoscopic images [9].

Due to the above limitations of visual inspection and dermoscopy, there is a strong motivation to develop computer-assisted diagnosis (CAD) systems to support dermatologists in their examinations. A critical aspect of CAD systems for efficiently analyzing dermoscopic images is automatically segmenting skin lesions from dermoscopic images, enabling more focused and efficient automated analysis of those areas. This automatic segmentation significantly aids early skin cancer detection and diagnosis by improving

diagnostic accuracy. However, accurate automatic segmentation of skin lesions is challenging due to three main factors: (1) Melanomas can vary greatly in shape, size, color, texture, and skin type. Distortions and natural features like hair, air bubbles, and blood vessels complicate the segmentation process. (2) Skin lesions often have fuzzy or uneven edges, and the contrast between the lesion and surrounding skin can be minimal. (3) Early algorithms were trained on relatively small datasets. Gathering large-scale skin lesion annotations from medical experts is difficult.

Existing CAD methods to address the above limitations often give unsatisfactory outcomes because they struggle with artifacts such as corners and low-contrast regions. Moreover, hairs against the background remain critical challenges, making boundary definition a difficult task [25]. On the other hand, the spatial attention mechanism uses spatial relationships of features and creates a spatial attention map. It has been observed that spatial attention modules are helpful in many image processing applications [22, 30, 36]. Therefore, this paper proposes a transfer learning-based approach combined with the spatial attention technique utilizing SegNet to improve the accuracy of skin lesion segmentation.

The paper's main contributions are summarized below:

1. **The Spatial Attention** module is introduced in the feature extraction process of the encoder. This module effectively captures spatial dependencies. This module enables the network to selectively emphasize important regions in the feature maps, improving the understanding of fine details.
2. **The bottleneck layer structure** of SegNet is modified by integrating a spatial attention module. This design increases the receptive field and allows the network to capture contextual information, resulting in more precise segmentation.

We have evaluated the performance of our proposed method on the ISIC 2018 dataset [7, 32]. This dataset contains dermoscopic images for skin lesion analysis. It was selected because it contains diverse skin lesion types collected from many patients. The proposed model showed the highest segmentation accuracy on this dataset compared to the published results on the same dataset. Hence, it proves the efficacy of the proposed model in accurate skin lesion segmentation.

2. Background and related works

This section first describes the essential background of skin lesion segmentation techniques. Subsequently, state-of-the-art skin lesion segmentation techniques are presented.

2.1. Background on skin lesion segmentation techniques

Skin lesion segmentation is pivotal in the fight against skin cancer, particularly Melanoma. Automated image analysis technique separates suspicious moles or lesions from the surrounding healthy skin in digital images, offering several benefits for dermatologists:

- **Enhanced Diagnostic Accuracy:** It provides a clear picture of the lesion's boundaries, allowing for a detailed analysis of its characteristics, such as color, texture, and borders. Moreover, it helps to distinguish between benign and malignant moles, detecting subtle variations that might be missed otherwise.
- **Earlier Detection:** By clearly highlighting suspicious areas, segmentation aids in identifying melanomas at an early stage, when treatment is most effective.
- **Improved Workflow Efficiency:** Automating the isolation of lesions saves time for dermatologists, allowing them to focus on interpreting the segmented data and making diagnoses, especially for complex cases.

However, achieving accurate segmentation is challenging due to:

- **Variability in Lesions:** Skin lesions can vary significantly in color, shape, texture, and size, making a one-size-fits-all approach difficult.
- **Artifacts:** Features like hair, blood vessels, or wrinkles can mimic lesion features and complicate the segmentation process.
- **Image Quality:** Variations in illumination, camera focus, and resolution can hinder accurate delineation.

These challenges underscore the importance of advanced techniques and tools in improving the precision and reliability of skin lesion segmentation.

2.2. Related work on skin lesion segmentation techniques

Researchers are actively working to overcome the hurdles above. The field of skin lesion segmentation primarily relies on two approaches:

- **Traditional Image Processing Techniques:** These methods use algorithms to analyze various image properties like color intensity and texture. While they can be effective, they often struggle with the high variability in skin lesions, limiting their accuracy and reliability.
- **Deep Learning-based Techniques:** These have revolutionized the field by leveraging Convolutional Neural Networks (CNNs). CNNs are trained on extensive datasets of labeled images, allowing them to learn complex patterns and identify subtle features that distinguish lesions from healthy skin. Due to their ability to handle the variability in lesions and their superior accuracy, deep learning approaches are considered state-of-the-art.

The initial research on combining transfer learning and fine-tuning techniques with a melanoma segmentation strategy based on U-net and LinkNet deep learning networks is found in [3]. The experiments were performed on PH2, ISIC 2018, and DermIS datasets.

The method faced limitations due to the image capture device, which affected the model's learning of disease characteristics like resolution, color, sharpness, and lighting. The authors claimed that the reproducibility of results is also limited by the diversity of skin tones across different populations.

A pyramid module incorporating lateral connections and top-down paths was used to compensate for the loss of spatial feature information [1].

This method integrated RetinaNet and MaskRCNN, with the Melanoma ISIC 2018 and PH2 datasets serving as training and validation grounds. The method's limitations included segmentation accuracy being affected by high occlusions near lesions and data class imbalance due to the absence of additional lesion data.

In [16], a fully automated multi-class skin lesion segmentation and classification approach was proposed using the most discriminant deep Learning Features and Improved Moth Flame Optimization. The proposed methodology's segmentation performance was evaluated on the ISBI 2016, ISBI 2017, ISIC 2018, and PH2 datasets. However, the computational time was one of the work's limitations.

In response to skin lesion segmentation, a novel EIU-Net method was proposed to tackle the challenging task [35]. Inverted residual blocks and an efficient pyramid squeeze attention (EPSA) block were proposed as the main encoders at different stages to capture the local and global contextual information. In contrast, atrous spatial pyramid pooling (ASPP) was utilized after the last encoder, and the soft-pool method was introduced for downsampling. Also, they proposed a novel method named multi-layer fusion (MLF) module to effectively fuse the feature distributions and capture significant boundary information of skin lesions in different encoders to improve the network's performance. Furthermore, a reshaped decoder fusion module was used to obtain multi-scale information by fusing feature maps of different decoders to improve the final results of skin lesion segmentation. To validate the performance of this network, it was compared with other methods on four public datasets, including the ISIC 2016, ISIC 2017, ISIC 2018, and PH2 datasets. Moreover, the main metric Dice Score achieved by the proposed EIU-Net are 0.919, 0.855, 0.902, and 0.916 on the four datasets; our EIU-Net can improve the accuracy of skin lesion segmentation [35].

In [23], the authors introduce a novel end-to-end trainable network for skin lesion segmentation. The proposed methodology comprises an encoder-decoder, a region-aware attention approach, and a guided loss function. The trainable parameters are reduced using depth-wise separable convolution, and the attention features are refined using a guided loss, resulting in a high Jaccard index. We assessed the effectiveness of our proposed RA-Net on four frequently utilized benchmark datasets for skin lesion segmentation: ISIC 2016, ISIC 2017, ISIC 2018, and PH2.

Integrating conventional treatment methods with deep learning frameworks to enhance skin lesion identification is proposed in [29]. The study used image data, handcrafted lesion features, and patient-centric metadata for effective skin cancer diagnosis. It combines image features transferred from Efficient Nets, color and texture information extracted from images, and pre-processed patient metadata to build a hybrid model. Each model underwent training and evaluation using the ISIC 2018 and ISIC 2019 datasets widely used for skin cancer analysis. However, a notable limitation of this

approach is the extreme imbalance in the datasets, which requires careful consideration of appropriate evaluation metrics. Despite achieving high sensitivity (90.49%) and specificity (97.76%) on the ISIC 2018 dataset, the model's performance may vary when applied to different datasets or real-world scenarios with varying data imbalance or complexity levels.

A novel deep-learning method named ChimeraNet was proposed for detecting hair and ruler marks in skin lesion images [17]. ChimeraNet employs an encoder-decoder architecture, incorporating a pre-trained EfficientNet and the decoder's squeeze-and-excitation residual (SERes) structures. However, this technique demands significant computational resources and training time due to the complexity of the encoder-decoder architecture and pre-trained models.

For accurate detection and delineation of hair in skin images, a researcher in [4] proposed a deep learning strategy based on a hybrid network of convolutional and recurrent layers for hair segmentation using weakly labeled data and deep encoded features. The spatial dependencies between disjointed patches were encoded by feeding the encoded features into recurrent neural network layers. The proposed method achieved segmentation accuracy with a Jaccard Index of 77.8 percent.

In [27], the researcher presented a machine learning-based methodology for segmenting skin lesions with novel borders and hair removal. The suggested approach removes any corner boundaries from an RGB skin picture as input. The skin hairs covering the image are then found and eliminated. The generated picture is then improved, and the GrabCut method is used to segment lesions. The research showed that the skin lesion segmentation method proposed in this paper had Jaccard indices of 0.77 and 0.80 on PH2 and ISIC 2018 datasets, respectively, and Dice indices of 0.87 and 0.82, respectively. The method failed to perform well on images with tiny affected areas. It automatically draws a rectangle around the region using the GrabCut method. However, when we deal with dermoscopic pictures with tiny lesions, initiating too big or too small rectangles for over-segmentation will occur during this method's selection process.

Segmentation accuracy degradation and occlusions in dermoscopic images constitute the significant problems identified here. High resolution and elaborate surface structures make conventional segmentation algorithms struggle with dermoscopic pictures. A mistake in segmentation accuracy may give wrong interpretations or cause detection failures, which affect the reliability of the diagnosis results. Moreover, occlusions within these images, such as those brought about by artifacts, hair follicles, and other foreign matter, block important details required during skin lesion boundary determination, leading to distortion. Therefore, it is essential to address these challenges if automated methods of segmenting medical pictures are to be effective in the field of dermatology.

Our proposed method is the Spatial SegNet model, designed based on attention mechanisms, which work well for increasing precision levels and handling occluded areas in dermoscopic images.

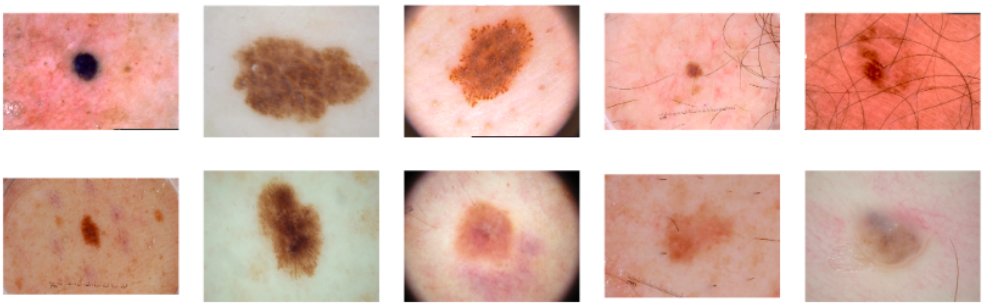

Fig. 1. Few examples of skin lesion samples in ISIC 2018.

3. Materials and methods

3.1. Data acquisition

In this research work, the ISIC 2018 dataset [7,32] was used to evaluate the results of our proposed method. The ISIC 2018 dataset is a comprehensive collection of dermoscopic images curated for skin lesion analysis. It contains 2 594 images in JPG format, each accompanied by ground truth segmentation masks. This dataset was selected for its diversity and the high-quality annotations it provides, which are essential for an accurate evaluation of segmentation methods. The images in the ISIC 2018 dataset vary significantly in lesion type and appearance, offering a robust challenge for our segmentation model. The ground truth masks serve as a benchmark for assessing the performance of our method, allowing us to measure the accuracy and effectiveness of the segmentation results quantitatively. Using this well-established dataset, we ensure our rigorous evaluation is relevant to real-world clinical scenarios. Figure 1 presents some samples of complex skin lesions in the dataset of ISIC 2018.

3.2. Proposed model

This section presents the details of our proposed method. SegNet has an encoder-decoder structure followed by a pixel-wise classification layer. The encoder architecture of SegNet is identical to VGG16's convolutional layers in topology. The decoder network maps the encoder feature maps to input resolution-sized feature maps for pixel-wise classification. In the proposed model, spatial attention layers are added to the encoder network of the SegNet architecture. In skin lesion images, spatial attention will help the model to focus on essential parts or regions of interest, thereby improving accuracy in segmentation by concentrating on relevant areas while ignoring irrelevant or noisy parts. The SegNet decoder network has several decoders organized in a hierarchy, each corresponding to

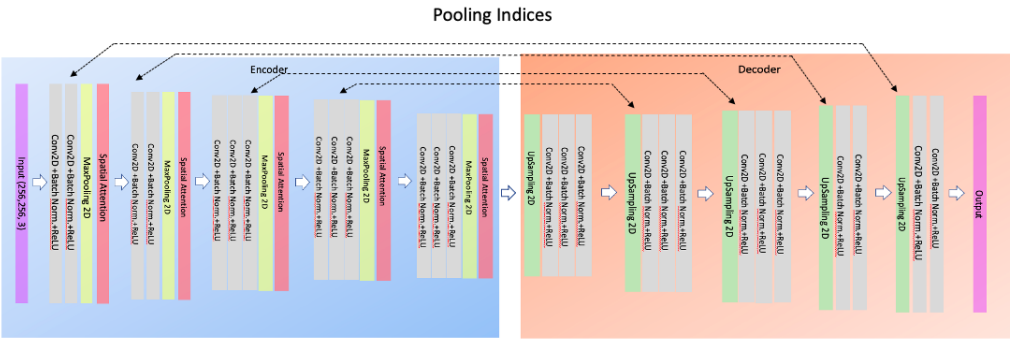

Fig. 2. SEGNET with Spatial Attention Architecture.

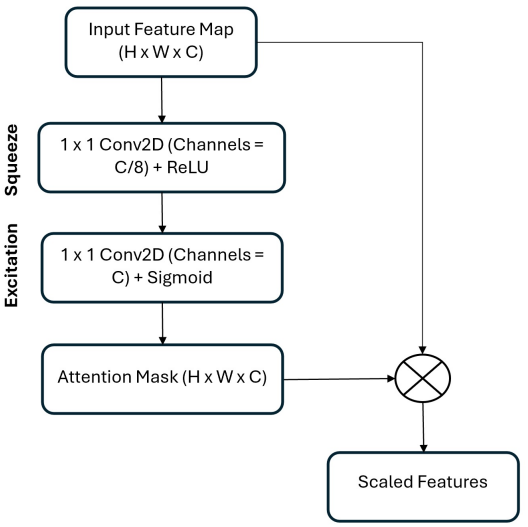

Fig. 3. Spatial Attention Module

one encoder. The correct decoders take their input feature maps and perform non-linear upsampling using max pooling indices that they receive from their respective encoders. It was derived from an architecture used for unsupervised feature learning [26]. Here are many practical advantages of reusing max-pooling indices during decoding. The architecture of SegNet with Spatial Attention is shown in Figure 2.

This model combines pre-trained VGG16 layers with spatial attention mechanisms in the encoder network. In the following subsections, every layer is explained in detail.

Input layer

An RGB image of a fixed size is an input to this layer. It is usually prepared after normalization, subtracting the mean from the image or scaling all values within a certain range. It accepts input images of size 256×256×3 (height, width, color channels).

Encoder blocks

Pre-trained VGG16 layers are used for feature extraction in the SegNet encoder network. The following layers from the pre-trained VGG16 model are used in the proposed architecture. Each convolutional block typically includes 3×3 filters with a stride of 1 and padding of 1 to extract features such as edges, textures, and color patterns. These layers apply learnable filters to the input feature maps. The filters essentially slide across the input, extracting features such as edges, textures, and color patterns. The number of filters used determines the complexity and richness of the extracted features. The activation function ReLU is applied after the convolutional layers (Conv) to introduce nonlinearity and allow the model to capture more complex relationships in the data. The batch normalization layers normalize the activations of the previous convolution layer. It facilitates faster convergence during training and enhances the stability of the learning process. The batch norm essentially standardizes the activations across different mini-batches, mitigating the issue of internal covariate shift. The pooling layers downsample the feature maps spatially using max pooling. It also makes the model robust to translations that occur very close together, while simultaneously making it less sensitive to noise because features become more generalized.

A Spatial Attention Block is a custom module added to the model to improve the segmentation of skin lesions (Figure 3). It acts on feature maps encoded in the previous convolutional block borrowed from the pre-trained VGG16 model. Specifically, it is inserted after every pooling layer within the VGG16 encoder. Its main objective is to enhance the encoded features and focus on them. Details of each spatial attention block are explained as follows:

- **Squeeze Operation:** First, feature maps are made smaller through a 1x1 convolution. This step aims to reduce the dimensionality of the space and the computational cost incurred by modulating units to learn their interaction.
- **Excitation Operation:** The method spins a spatial attention map to identify vital skin detection areas. It is achieved by applying another 1x1 convolution operation and performing a sigmoid activation function. The resultant map assigns different values between 0 and 1 for each part of an image, where a value of zero means least significant and a value of one corresponds to the most significant pixels, thereby highlighting those regions necessary for segmentation through information obtained from prior layers.
- **Element-wise Multiplication:** The last part includes element-wise multiplication of the initial feature maps with the produced spatial attention map. By doing this,

characteristics identified as important by our attention mechanism are emphasized, thus enabling the model to concentrate on particular parts during its segmenting duties. Its intensified focus strengthens the model's ability to distinguish between unhealthy cells and their surrounding healthy tissues.

Further elaborating on the workings of the Spatial Attention Block, the first Conv2D layer squeezes the feature maps by decreasing the channels or properties in this module. For example, if the entry feature maps have 512 channels, the squeezing layer might bring this down to a smaller amount, like 64 channels. Then, using the sigmoid activation function, the second Conv2D generates attention weights representing a probability distribution that shows the importance of different spatial locations within the given feature maps. After that, these produced attention weights are multiplied with original features so that some features can be amplified or suppressed selectively based on their importance towards achieving the segmentation goal. Thus, resulting scaled feature maps will center around crucial areas, helping the model capture fine details and semantics necessary for accurate segmentation.

Decoder layers (segmentation mask reconstruction)

The decoder part takes the encoded feature maps obtained from the encoder along with the spatial attention. It has a symmetric structure relative to the encoder and progressively increases the resolution of feature maps through transpose convolution operations. Convolution layers are attached after these upsampling operations to refine features and learn spatial relationships between pixels. Unlike its counterpart, which extracts them, this one aims to recover spatial information while predicting probabilities for individual pixels to be part of skin lesions. For example, (background versus lesion) background versus lesion class probability maps may be obtained by applying the softmax activation function to the final output layer on a class basis. These layers receive processed features, including effects caused by spatial attention blocks within the encoder, and then gradually reconstruct an image that focuses on the segmentation task. They function oppositely from encoders, i.e., starting with a high-level understanding of the picture and adding more detailed spatial information stepwise downwards towards the lowest level segmentation features being dealt with at every decoder block stage. Each block typically consists of :

- **Upsampling 2D:** It Increases feature maps' spatial resolution by this layer. Unlike the traditional Conv layer, this layer learns upsampled filters that expand the feature maps while introducing new spatial information. It allows the model to recover spatial details lost during pooling in the encoder.
- **Convolutional layers:** After Upsampling layers, Conv layers are applied to refine the features and learn the relationships between pixels similar to the encoder. They help to combine upsampled information with high-level features from the encoder.

- **Batch Normalization layers (Batch Norm):** These are used for normalizing activations after upsampling, similar to the encoder, for better training stability.

Output layer

The final decoder output is fed to a softmax classifier layer to produce the class probabilities. The softmax function produces a probability map for every pixel in the image. This map shows how likely each pixel is to belong to a specific class (e.g., background or skin lesion). The class with the highest probability for each pixel becomes the predicted segmentation label.

Spatial attention mechanisms are incorporated into models designed for skin lesion segmentation, improving the overall performance and reliability of the proposed model. This technique combines pre-trained features, spatial attention mechanisms, and SegNet's decoder architecture to achieve accurate skin lesion segmentation. Pre-trained VGG16 weights extract essential image features more effectively, reducing training time and enhancing its generalization ability over new data. Considering different skin lesion sizes and appearances, introducing a spatial attention block narrows down the essential parts of an image, thus leading to precise skin lesion segmentation.

3.3. Evaluation metrics

To assess the performance of our proposed skin lesion segmentation method, we employed a variety of evaluation metrics that provide a comprehensive analysis of segmentation accuracy and quality. The metrics used in this study include the Dice Coefficient and Binary Accuracy.

TP and FP refer to lesion pixels extracted as lesion pixels and non-lesion pixels extracted as lesion pixels, respectively. At the same time, FN and TN represent lesion pixels extracted as non-lesion pixels and non-lesion pixels extracted as non-lesion pixels, respectively.

Dice Coefficient

The Dice Coefficient is an essential metric for evaluating segmentation quality. It is calculated as the ratio of twice the area of overlap between the predicted and ground truth masks to the sum of the areas of both masks. The Dice Coefficient ranges from 0 to 1, with a value closer to 1 indicating better segmentation accuracy. This metric is beneficial for handling class imbalance, as it emphasizes the correct prediction of positive samples. The dice similarity coefficient is a spatial overlap index and a reproducibility validation metric, and it computes the similarity index between the given images.

$$\text{Dice Coefficient} = \frac{2\text{TP}}{(\text{FP} + \text{TP}) + (\text{TP} + \text{FN})}\,. \tag{1}$$

Accuracy

Accuracy refers to the proportion of correctly predicted pixels (lesion and non-lesion) out of the total number of pixels. It is calculated as follows:

$$\text{Accuracy} = \frac{\text{TP} + \text{TN}}{\text{TP} + \text{TN} + \text{FN} + \text{FP}} . \tag{2}$$

Precision

Precision refers to the proportion of true positive predictions among all the pixels predicted as lesions. It indicates the model's accuracy in identifying the lesion pixels out of all the pixels it labeled as lesions. Precision is calculated as follows:

$$\text{Precision} = \frac{\text{TP}}{\text{TP} + \text{FP}} . \tag{3}$$

Sensitivity

Sensitivity, also called Recall, measures the proportion of actual positives (lesions) the model correctly identifies. It indicates the model's ability to detect the lesion pixels

$$\text{Sensitivity} = \frac{\text{TP}}{\text{TP} + \text{FN}} . \tag{4}$$

Specificity

Specificity measures the proportion of actual negatives (non-lesions) the model correctly identifies. It indicates the model's ability to avoid false positives.

$$\text{Specificity} = \frac{\text{TN}}{\text{TN} + \text{FP}} . \tag{5}$$

F1 Score

F_1 Score is the harmonic mean of Precision and Recall (Sensitivity). It is a balanced measure that considers both false positives and false negatives.

$$F_1 = \frac{2(\text{Precision} \times \text{Recall})}{\text{Precision} + \text{Recall}} . \tag{6}$$

IOU

IOU is used to measure the overlap between two images.

$$\text{IOU} = \frac{\text{TP}}{\text{TP} + \text{FP} + \text{FN}} . \tag{7}$$

4. Experimental results

An open-source machine learning framework, TensorFlow, implements the methodology. It is a user-friendly interface for working with deep neural networks, designed for ease of use rather than machine-level interactions. It is a library mainly used for developing real-time computer vision applications.

- **CPU Resources**:
  - Environment 1: 256MB memory limit on device `/device:CPU:0`.
  - Environment 2: XLA CPU with 16GB memory limit on device `/device:XLA_CPU:0`.
- **GPU Resources**:
  - Tesla T4 GPUs: Two GPUs with 14.8GB memory each, identified as `/device:GPU:0` and `/device:GPU:1`, with PCI Bus IDs `0000:00:04.0` and `0000:00:05.0`, respectively. Both GPUs have Compute Capability 7.5.
  - XLA GPUs: Two GPUs with 16GB memory each, denoted as `/device:XLA_GPU:0` and `/device:XLA_GPU:1`.

This combination of hardware configurations provided the computational capacity necessary for efficient training and testing of deep learning models, enabling the handling of large-scale data processing and complex model architectures.

4.1. Hyperparameters

To achieve a skin lesion segmentation model, hyperparameters shown in Table 1 are chosen to optimize performance and manage computational resources effectively. A learning rate of 5×10^{-6} is important as it allows for small steps to be taken by the optimizer while minimizing the loss function. It ensures the model converges slowly and steadily without overshooting the minimum loss function. Batch size 8 strikes a balance between memory efficiency and accurate gradient estimation.

To validate our skin lesion segmentation method, we carried out a 5-fold cross-validation. The steps involved partitioning a dataset into five equal sets, training on four, and validating against the fifth set. The results are shown in Table 2. The mean of the five-fold validation results is given in the last row of the table. The results show

Tab. 1. Hyperparameters for the proposed model.

Parameter Name	Parameter Value
Learning Rate	5×10^{-6}
Batch Size	8
Input Size	256, 256, 3
Optimizer	Adam Optimizer
Epoch	60

Tab. 2. Results of the proposed model with 5-fold cross validation (STD: standard deviation).

Folds	IoU	Dice Coefficient	Precision	Sensitivity	Specificity	Accuracy
1	0.8026	0.8980	0.8969	0.9086	0.9718	0.9657
2	0.8240	0.9242	0.9163	0.8969	0.9744	0.9616
3	0.8026	0.9051	0.9272	0.8890	0.9820	0.9611
4	0.8240	0.9022	0.9086	0.8725	0.9718	0.9631
5	0.8026	0.8965	0.9310	0.8619	0.9820	0.9611
Mean	**0.8111**	**0.9052**	**0.9160**	**0.8857**	**0.9764**	**0.96252**
STD	**0.0092**	**0.0045**	**0.0155**	**0.0172**	**0.0053**	**0.0026**

a high value of the Dice coefficient (0.9052) and segmentation accuracy (96%). The sensitivity of the proposed model is 0.8857.

Table 3 compares the results of our proposed model against state-of-the-art published results using the ISIC 2018 dataset. Our proposed model of skin lesion segmentation, tested on the ISIC 2018 dataset, shows significant improvements in segmentation. The primary comparison tools used to judge the outcomes are the Dice Coefficient and Binary Accuracy. The high value of the Dice Coefficient shows more similarity of the predicted results with the ground truth mask.

Adding spatial attention to the SegNet architecture achieves better skin lesion segmentation results. Spatial attention assigns weights to each pixel, highlighting the areas of interest and allowing the model to distinguish between lesion and non-lesion regions. This improves the segmentation accuracy as the model can focus on the spatial locations with features relevant to skin lesions, such as irregular shapes and varying pigmentation over background noise.

The SegNet with spatial attention model quantitatively improved the Dice similarity coefficient, IoU, and accuracy scores. These improvements are significant compared to the other segmentation models (Table 3). Figure 4 shows skin lesion segmentation results. The segmentation output looks better in the Figure 4, and lesion boundaries are more precise and consistent.

Tab. 3. Comparative analysis with state-of-the-art techniques.

Model	Dataset Split	Parameters [10^6]	Accuracy	Dice Coefficient
TMU Net [5]	70% training, 10% validation, and 20% testing	–	0.9603	0.905
UNeXt [33]	80% training, 20% testing	1.47	0.9586	0.8873
FAT-Net [34]	70% training, 10% validation, and 20% testing	30	0.9578	0.8903
CPFNET [10]	5-fold cross-validation	43	0.9496	0.8769
DAGAN [18]	2296 images for training, 300 images for testing.	54	0.9324	0.87707
CKDNet [15]	–	51	0.9492	0.8779
REDAUNet [20]	70% training, 10% validation, and 20% testing	47.77	0.9444	0.902
SA SegNet (Ours)	Five-Folds Cross-Validation.	29.6	**0.9625**	**0.9052**

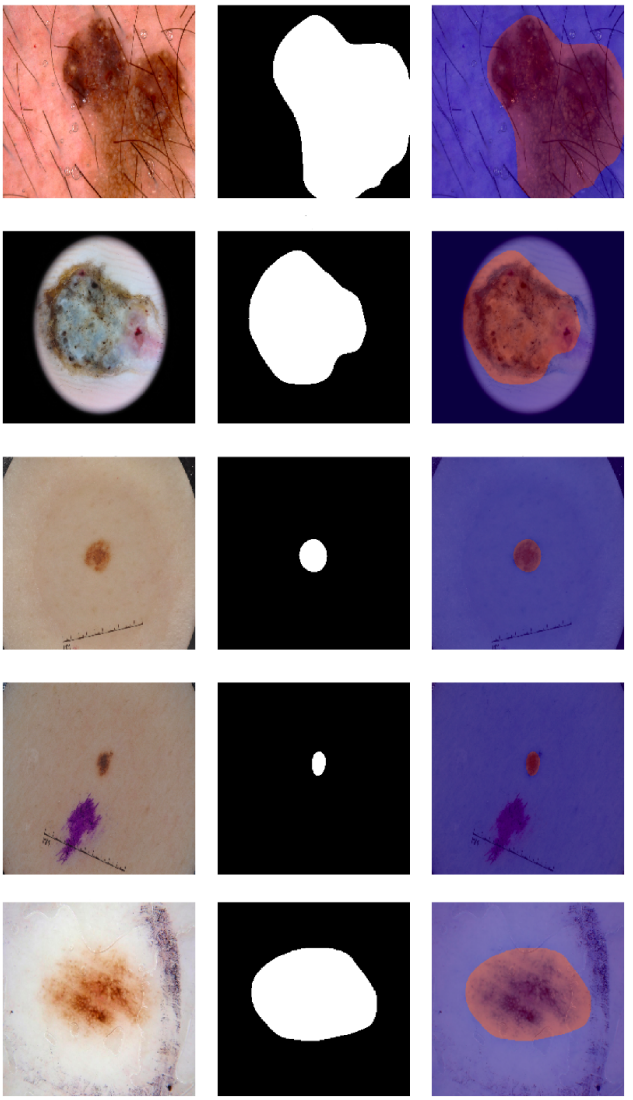

Fig. 4. Some of the segmented images. Vertically: five cases. Horizontally: (**a**) original image; (**b**) predicted mask, Dice = 0.85; (**c**) overlay image.

4.2. Impact of batch size and learning rate

In this experiment, we analyzed the influence of hyperparameters on the model performance. This analysis aims to understand the effect of batch size and learning rate on the Dice Coefficient and the Accuracy metrics. In these experiments, the whole dataset is divided in the ratio of 0.7:0.1:0.2 for training, validation, and testing, respectively.

The Table 4 presents results for the model at various initial learning rates and their effect on the Dice Coefficient and Accuracy. The learning rate of 1×10^{-6} is too low for the model to learn efficiently, as it yields the lowest performance, with a Dice coefficient of 0.8611 and an accuracy of 0.9395. A learning rate of 1×10^{-5} performs the best with the highest value of the Dice coefficient of 0.9053, an accuracy of 0.9626. Thus, it infers that this is the ideal rate of learning and generalization. If the learning rate is increased to 1×10^{-4}, the performance decreases slightly, as the Dice coefficient goes to 0.8794, and an accuracy of 0.9527 is achieved. This shows that although the model performs well, the learning rate is too large for optimal training. As the learning rate is increased to 1×10^{-3}, the model training is the worst, with a Dice coefficient of 0.8761 and an accuracy of 0.9436 on the testing dataset. It may indicate that the model is converging too fast and missing some finer details in the data. The learning rate of 1×10^{-5} is the most effective, being the best in segmentation and classification tasks, while increasing or decreasing the learning rate worsens the performance.

Batch size determines the number of samples in the training dataset to update the parameters. Increasing the batch size means fewer weight updates in an epoch. Hence, memory and computational requirements are lower for smaller batch sizes due to the smaller number of samples per update. However, for smaller batch sizes, the effect of noise and variance of the loss gradient will be more on the weight updates of the model. The Table 5 compares the model's performance across different batch sizes in

Tab. 4. Comparison of Dice Coefficient and Accuracy for different initial learning rates

Initial Learning Rate	Dice Coefficient	Accuracy
1×10^{-6}	0.8611	0.9395
1×10^{-5} **(Ours)**	0.9053	0.9626
1×10^{-4}	0.8794	0.9527
1×10^{-3}	0.8761	0.9436

Tab. 5. Comparison of model performance for different batch sizes.

Batch Size	Test Dice Coefficient	Test Accuracy
4	0.9005	0.9578
8 (Ours)	0.9053	0.9626
12	0.8967	0.9570
16	0.8733	0.9509

Tab. 6. Comparison of models based on model's variations.

Model	Test Dice Coefficient	Test Accuracy
SegNet only	0.8977	0.9581
Partial Removal of Spatial Attention	0.8986	0.9561
Removal of Batch Normalization	0.8827	0.9505
Spatial Attention SegNet (Ours)	0.9052	0.9625

terms of Dice Coefficient and Accuracy. Comparing the dice coefficient and accuracy for various batch sizes, a batch size of eight is optimal. The model performs best with a Dice coefficient of 0.9053 and an accuracy of 0.9626. The performance degrades as the batch size increases from eight, and the dice coefficient and accuracy decrease. Finally, with the batch size of 16, the performance significantly drops (Dice coefficient of 0.8733 and accuracy of 0.9509), which implies that the bigger batch sizes may preclude the model's ability to converge effectively and even generalize well. The overall results, however, indicate that batch size eight is more likely to bring equilibrium to the model's computing efficiency and practical utility. Therefore, it is the most suitable option for this experiment.

4.3. Ablation experiments

In this section, we perform an ablation study on the proposed model. We have studied the effect of the spatial attention layer and batch normalization layer.

The Table 6 showcases the comparison of the performances of the different versions of the models. The core of the system is the SegNet architecture, and performance can be greatly enhanced by the introduction of some components, like batch normalization and spatial attention. The spatial attention technique is a method for improving segmentation accuracy, which allows the model to focus on relevant areas of the input.In our model we have included another highly significant layer, which is called batch normalization (BN). By doing BN the input to each layer, the result is the stabilization of the process of learning and reduction of internal co-variate shifts. An ablation study shows that when batch normalization is taken away, both the Dice coefficient and the accuracy fall drastically. The removal of batch normalization from the model greatly decreases the model's accuracy, thus proving its importance in ensuring a successful learning period.

Our model, proposed in this paper, combines spatial attention and batch normalization layers. Both layers in the model provides best Dice coefficient of 0.9052 and the highest accuracy of 0.9625. Hence, it is clear that spatial attention helps the model pick the salient parts of the image, while batch normalization ensures the model's training

runs smoothly and can generalize well, thereby enhancing both segmentation and classification performance. The implementing these layers is essential for the robustness, accuracy, and capacity to deal with the complexity of the patterns in the data.

5. Conclusion and future work

This paper provides a detailed framework for the proposed skin lesion Segmentation model. Our proposed approach uses SegNet architecture combined with spatial attention layers. Encoder layers are taken from the pre-trained model of VGG16. Our model showed better segmentation accuracy and improved lesion boundary delineation precision. It is evident from Table 3 that the proposed model performed better on the ISIC 2018 dataset than other published state-of-the-art models. We have achieved a dice coefficient of **0.9052** and a segmentation accuracy of **0.9625**.

An important aspect of future work is the incorporation of multimodal data. Combining dermoscopic images with clinical information provides a more holistic approach to analyzing skin lesions; thus, this approach may increase diagnostic performance by improving segmentation accuracy. This kind of approach uses different data strengths to give a more precise and reliable diagnosis. It is also essential to build real-time segmentation systems for clinical purposes. These systems need to work efficiently on edge devices or mobile platforms to be accessible for use in different clinical environments.

References

[1] N. Ahmed, X. Tan, and L. Ma. A new method proposed to melanoma-skin cancer lesion detection and segmentation based on hybrid convolutional neural network. *Multimedia Tools and Applications* 82(8):11873–11896, 2023. doi:10.1007/s11042-022-13618-0.

[2] Z. Apalla, A. Lallas, E. Sotiriou, E. Lazaridou, and D. Ioannides. Epidemiological trends in skin cancer. *Dermatology Practical & Conceptual* 7(2):1, 2017. doi:10.5826/dpc.0702a01.

[3] R. L. Araújo, F. H. D. de Araújo, and R. R. V. e. Silva. Automatic segmentation of melanoma skin cancer using transfer learning and fine-tuning. *Multimedia Systems* 28(4):1239–1250, 2022. doi:10.1007/s00530-021-00840-3.

[4] M. Attia, M. Hossny, H. Zhou, S. Nahavandi, H. Asadi, et al. Digital hair segmentation using hybrid convolutional and recurrent neural networks architecture. *Computer methods and programs in biomedicine* 177:17–30, 2019. doi:10.1016/j.cmpb.2019.05.010.

[5] R. Azad, M. Heidari, Y. Wu, and D. Merhof. Contextual Attention Network: Transformer meets U-Net. In: *International Workshop on Machine Learning in Medical Imaging (MLMI 2022)*, vol. 13583 of *Lecture Notes in Computer Science*, pp. 377–386. Springer, 2022. doi:10.1007/978-3-031-21014-3_39.

[6] Canadian Skin Cance Foundation. Skin cancer early detection. In: Skin Cancer, 2024. `https://www.canadianskincancerfoundation.com/early-detection/`. [Accessed: 2024-11-05].

[7] N. Codella, V. Rotemberg, P. Tschandl, M. E. Celebi, S. Dusza, et al. Skin lesion analysis toward

melanoma detection 2018: A challenge hosted by the International Skin Imaging Collaboration (ISIC). arXiv, arXiv:1902.03368, 2019. doi:10.48550/arXiv.1902.03368.

[8] J. Dinnes, J. J. Deeks, M. J. Grainge, N. Chuchu, L. Ferrante di Ruffano, et al. Visual inspection for diagnosing cutaneous melanoma in adults. *Cochrane Database of Systematic Reviews* 2018(12), 1996. doi:10.1002/14651858.CD013194.

[9] H. C. Engasser and E. M. Warshaw. Dermatoscopy use by US dermatologists: A cross-sectional survey. *Journal of the American Academy of Dermatology* 63(3):412–419, 2010. doi:10.1016/j.jaad.2009.09.050.

[10] S. Feng, H. Zhao, F. Shi, X. Cheng, M. Wang, et al. CPFNet: Context pyramid fusion network for medical image segmentation. *IEEE Transactions on Medical Imaging* 39(10):3008–3018, 2020. doi:10.1109/TMI.2020.2983721.

[11] Z. Ge, S. Demyanov, R. Chakravorty, A. Bowling, and R. Garnavi. Skin disease recognition using deep saliency features and multimodal learning of dermoscopy and clinical images. In: *Proc. 20th Int. Conf. Medical Image Computing and Computer Assisted Intervention (MICCAI 2017)*, vol. 10435 part III of *Lecture Notes in Computer Science*, pp. 250–258. Springer, 2017. doi:10.1007/978-3-319-66179-7_29.

[12] K. Hauser, A. Kurz, S. Haggenmüller, R. C. Maron, C. von Kalle, et al. Explainable artificial intelligence in skin cancer recognition: A systematic review. *European Journal of Cancer* 167:54–69, 2022. doi:10.1016/j.ejca.2022.02.025.

[13] W. Hu, L. Fang, R. Ni, H. Zhang, and G. Pan. Changing trends in the disease burden of non-melanoma skin cancer globally from 1990 to 2019 and its predicted level in 25 years. *BMC cancer* 22(1):836, 2022. doi:10.1186/s12885-022-09940-3.

[14] IARC. Skin cancer. In: International Agency for Research on Cancer (IARC), WHO, 2024. `https://www.iarc.who.int/cancer-type/skin-cancer/`. [Accessed: 05-11-2024].

[15] Q. Jin, H. Cui, C. Sun, Z. Meng, and R. Su. Cascade knowledge diffusion network for skin lesion diagnosis and segmentation. *Applied Soft Computing* 99:106881, 2021. doi:10.1016/j.asoc.2020.106881.

[16] M. A. Khan, M. Sharif, T. Akram, R. Damaševičius, and R. Maskeliūnas. Skin lesion segmentation and multiclass classification using deep learning features and improved moth flame optimization. *Diagnostics* 11(5):811, 2021. doi:10.3390/diagnostics11050811.

[17] N. Lama, R. Kasmi, J. R. Hagerty, R. J. Stanley, R. Young, et al. ChimeraNet: U-Net for hair detection in dermoscopic skin lesion images. *Journal of Digital Imaging* 36(2):526–535, 2023. doi:10.1007/s10278-022-00740-6.

[18] B. Lei, Z. Xia, F. Jiang, X. Jiang, Z. Ge, et al. Skin lesion segmentation via generative adversarial networks with dual discriminators. *Medical Image Analysis* 64:101716, 2020. doi:10.1016/j.media.2020.101716.

[19] W. Li, A. N. J. Raj, T. Tjahjadi, and Z. Zhuang. Digital hair removal by deep learning for skin lesion segmentation. *Pattern Recognition* 117:107994, 2021. doi:10.1016/j.patcog.2021.107994.

[20] L. Liu, X. Zhang, Y. Li, and Z. Xu. An improved multi-scale feature fusion for skin lesion segmentation. *Applied Sciences* 13(14), 2023. doi:10.3390/app13148512.

[21] P. A. Lyakhov, U. A. Lyakhova, and D. I. Kalita. Multimodal analysis of unbalanced dermatological data for skin cancer recognition. *IEEE Access* 11:131487–131507, 2023. doi:10.1109/ACCESS.2023.3336289.

[22] R. Maurya, N. N. Pandey, M. K. Dutta, and M. Karnati. FCCS-Net: Breast cancer classification using Multi-Level fully Convolutional-Channel and spatial attention-based transfer learning approach. *Biomedical Signal Processing and Control* 94:106258, 2024. doi:10.1016/j.bspc.2024.106258.

[23] A. Naveed, S. S. Naqvi, S. Iqbal, I. Razzak, H. A. Khan, et al. RA-Net: Region-Aware attention Network for skin lesion segmentation. *Cognitive Computation* 16:2279–2296, 2024. doi:10.1007/s12559-024-10304-1.

[24] S. Rajpar and J. Marsden. *ABC of Skin Cancer*. John Wiley & Sons, 2009.

[25] G. Ramella. Hair removal combining saliency, shape and color. *Applied Sciences* 11(1):447, 2021. doi:10.3390/app11010447.

[26] M. Ranzato, F. J. Huang, Y.-L. Boureau, and Y. LeCun. Unsupervised learning of invariant feature hierarchies with applications to object recognition. In: *2007 IEEE Conference on Computer Vision and Pattern Recognition*, pp. 1–8. IEEE, 2007. doi:10.1109/CVPR.2007.383157.

[27] M. Rehman, M. Ali, M. Obayya, J. Asghar, L. Hussain, et al. Machine learning based skin lesion segmentation method with novel borders and hair removal techniques. *Plos one* 17(11):e0275781, 2022. doi:10.1371/journal.pone.0275781.

[28] M. A. Richard, C. Paul, T. Nijsten, P. Gisondi, C. Salavastru, et al. Prevalence of most common skin diseases in Europe: a population-based study. *Journal of the European Academy of Dermatology and Venereology* 36(7):1088–1096, 2022. doi:10.1111/jdv.18050.

[29] M. Sharafudeen and S. S. V. Chandra. Detecting skin lesions fusing handcrafted features in image network ensembles. *Multimedia Tools and Applications* 82:3155–3175, 2022. doi:0.1007/s11042-022-13046-0.

[30] S. Tehsin, I. M. Nasir, R. Damaševičius, and R. Maskeliūnas. DaSAM: Disease and spatial attention module-based explainable model for brain tumor detection. *Big Data and Cognitive Computing* 8(9), 2024. doi:10.3390/bdcc8090097.

[31] The American Cancer Society medical and editorial content team. Key Statistics for Melanoma Skin Cancer. In: Melanoma Skin Cancer, 2025. `https://www.cancer.net/cancer-types/melanoma/statistics`. [Accessed: 2024-11-05].

[32] The International Skin Imaging Collaboration. ISIC CHALLENGE, 2018-2024. `https://challenge.isic-archive.com/data/`.

[33] J. M. J. Valanarasu and V. M. Patel. UNeXt: MLP-based rapid medical image segmentation network. In: *Medical Image Computing and Computer Assisted Intervention (MICCAI 2022)*, vol. 13435 of *Lecture Notes in Computer Science*, pp. 23–33. Springer Nature Switzerland, 2022. doi:10.1007/978-3-031-16443-9_3.

[34] H. Wu, S. Chen, G. Chen, W. Wang, B. Lei, et al. FAT-Net: Feature adaptive transformers for automated skin lesion segmentation. *Medical Image Analysis* 76:102327, 2022. doi:https://doi.org/10.1016/j.media.2021.102327.

[35] Z. Yu, L. Yu, W. Zheng, and S. Wang. EIU-Net: Enhanced feature extraction and improved skip connections in U-Net for skin lesion segmentation. *Computers in Biology and Medicine* 162:107081, 2023. doi:10.1016/j.compbiomed.2023.107081.

[36] Y. Zhong, Z. Shi, Y. Zhang, Y. Zhang, and H. Li. CSAN-UNet: Channel spatial attention nested UNet for infrared small target detection. *Remote Sensing* 16(11):1894, 2024. doi:10.3390/rs16111894.