

Segmentation of Skin Lesions and their Attributes in Dermatoscopic Images Based on Convolutional Neural Networks

Segmentação de Lesões de Pele e seus Atributos em Imagens Dermatoscópicas baseada em Redes Neurais Convolucionais

Antônio Vinicius de Moura Rodrigues¹, Roberta Barbosa Oliveira^{1*}

Abstract: Segmentation of skin lesions in dermoscopic images is a crucial step in diagnosing skin cancer, and Convolutional Neural Networks (CNNs) have emerged as powerful tools to address this challenge. This paper evaluated the effectiveness of two CNN models, TerausNet-16 and Mask R-CNN, in segmenting skin lesions and five of their attributes in the dermoscopic images from the ISIC 2018 Challenge dataset. Jaccard Similarity Index (JSI) and Dice Similarity Coefficient (DSC) have been used as evaluation metrics. The results revealed that Mask R-CNN significantly outperformed TerausNet-16. The best model achieved 82.57% JSI and 84.76% DSC for lesion segmentation, and 42.86% JSI and 51.60% DSC for attribute segmentation when addressing attribute imbalance. Despite the longer training time, the results highlighted the potential of Mask R-CNN for improving the effectiveness of melanoma segmentation.

Keywords: Deep Learning — Convolutional Neural Networks — TerausNet — MASK R-CNN — Image Segmentation — Skin Cancer

Resumo: A segmentação de lesões de pele em imagens dermatoscópicas é essencial para o diagnóstico de câncer de pele, e as Redes Neurais Convolucionais (CNNs) têm se destacado nessa tarefa. Este estudo comparou dois modelos de CNN, TerausNet-16 e Mask R-CNN, na segmentação de lesões de pele e cinco de seus atributos utilizando o conjunto de dados ISIC 2018 Challenge. O Índice de Similaridade Jaccard (JSI) e o Coeficiente de Similaridade Dice (DSC) foram usados como métricas de avaliação. Os resultados revelaram que o Mask R-CNN superou significativamente o TerausNet-16, com JSI de 82,57% e DSC de 84,76% para segmentação das lesões, e JSI de 42,86% e DSC de 51,60% para segmentação de atributos, ao mitigar o desbalanceamento dos atributos. Apesar do tempo de treinamento mais longo, os resultados ressaltaram o potencial do Mask R-CNN para melhorar a eficácia da segmentação do melanoma.

Palavras-Chave: Aprendizado Profundo — Redes Neurais Convolucionais — TerausNet — MASK R-CNN — Segmentação de Imagens — Câncer de Pele

¹ Department of Computer Science, University of Brasília (UnB), Brazil

*Corresponding author: roberta.oliveira@unb.br

DOI: <http://dx.doi.org/10.22456/2175-2745.143546> • Received: 24/10/2024 • Accepted: 02/12/2024

CC BY-NC-ND 4.0 - This work is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License.

1. Introduction

According to the International Agency for Research on Cancer (IARC), skin cancers are the most prevalent globally, with over 1.5 million new cases and 330,000 melanoma cases in 2022, leading to 60,000 fatalities [1]. Early detection of melanoma is crucial for improving survival rates [2], leading to the development of several innovative techniques to assist medical diagnosis. There is a global interest in developing Computer-Aided Diagnostic (CAD) systems, which aim to assist in the initial analysis of skin lesion indicators, as com-

puters can analyze complex data more quickly and accurately than humans. Deep Learning (DL), a subfield of machine learning that uses artificial neural networks to learn from data [3], is one promising approach for enhancing CAD systems. DL algorithms excel at dermoscopic image analysis tasks since such algorithms can learn complex patterns from the input data, including subtle changes in color and texture that characterize skin lesions [4].

Convolutional Neural Networks (CNNs) have significantly advanced the field of computer vision, leveraging the power of DL to autonomously extract intricate patterns and discern

subtle variations in color, texture, and structure within dermoscopic images [3]. While traditional image processing techniques remain valuable, CNNs enable automation of complex analyses, allowing CAD systems to identify skin lesion indicators with high effectiveness even in challenging cases with visual noise from artifacts like hair, reflections, or uneven lighting, thereby improving diagnostic effectiveness [5]. CNNs' ability to swiftly process vast datasets and identify anomalous features unlocks the potential to further advance skin lesion diagnosis and improve patient outcomes. While the CNN architecture is predominantly employed for classifying entire images into single categories, its basic structure can be adapted for other computer vision tasks, such as segmentation, which involves assigning each pixel in the image to a specific class or region [2]. This task is critically important because it allows the precise isolation of the lesion or its attributes from the rest of the image, enabling focused analysis by the CAD system.

A notable adaptation of the CNN for segmentation is the Fully Convolutional Neural Networks (FCNNs), which replace fully connected layers into convolutional layers, generating feature maps that enable the precise delineation of Regions of Interest (ROIs) in images. One well-established example of an FCNN is the U-Net model [6]. The evolution of CNNs has resulted in models such as TerausNet [7], which enhances the U-Net by integrating a pre-trained VGG11. The VGG architecture [8] consists of multiple convolutional layers followed by max-pooling layers, facilitating hierarchical feature extraction from input images. TerausNet initially used the VGG11 architecture, but adopting deeper architectures like VGG16, as in TerausNet-16, enhances feature extraction, segmentation performance, and training convergence [9].

Mask R-CNN [10] is another promising model for segmentation tasks. This model excels in recognizing objects within an image and precisely delineating the boundaries of each object instance. Mask R-CNN integrates a Region Proposal Network (RPN) for efficient object detection and a segmentation mask head for pixel-level instance segmentation, building upon the foundation laid by Faster R-CNN [11]. Mask R-CNN has demonstrated its instrumental role in tasks where a detailed understanding of object boundaries is crucial by simultaneously identifying objects and generating pixel-wise masks for each instance.

Here, a comparative analysis is carried out to evaluate the performance of the TerausNet-16 and Mask R-CNN models in segmenting skin lesions and five of their attributes: globules, milia-like cysts, negative networks, pigment networks, and streaks [12]. To the best of our knowledge, no previous studies in the literature have conducted a direct comparison between these models for these specific tasks [2, 4]. This paper provides novel insights into the effectiveness of these segmentation models for detailed skin lesion analysis.

This paper is organized as follows: Related works are described in Section 2, the adopted methodology is discussed in Section 3, experimental results are reported in Section 4, the

discussion is provided in Section 5, and finally, the conclusion and future work are presented in Section 6.

2. Related works

The application of deep learning to medical image segmentation has become a significant area of research in recent years, particularly in improving the segmentation of skin lesions and their attributes. Many studies utilize ISIC challenge datasets from 2016-2018, which provide thousands of annotated dermoscopic images for segmentation tasks. These studies explore a range of CNN architectures, each offering distinct advantages: U-Nets provide precise localization and sharp boundary delineation; ResNets mitigate vanishing gradients to enable deeper network training; DenseNets enhance information flow across layers; and attention mechanisms focus on critical image regions for improved effectiveness. The most commonly used evaluation metrics are the Jaccard Similarity Index (JSI) and the Dice Similarity Coefficient (DSC), or metrics derived from these [2].

2.1 Lesion Segmentation

Segmentation of skin lesions involves the precise delineation of the entire lesion area and it is a critical step in diagnosing skin cancer. In a comparative study, Alfaro et al. [13] investigate the performance of two well-established CNN models, U-Net and Mask R-CNN, using the ISIC 2017 dataset. Mask R-CNN performed better, achieving a top JSI score of 74.38% after 180 epochs, while U-Net reached 58.17%. On the other hand, Qian et al. [14] introduced a two-stage method for ISIC 2018 dataset, combining lesion detection with segmentation using advanced architectures like DeepLab and PSPNet. Their approach, incorporating innovative techniques like a modified ASPP block and ensemble post-processing, achieved a thresholded JSI of 80.2%.

2.2 Attribute Segmentation

Attribute segmentation is the precise delineation of specific structures within skin lesions, such as variations in color, texture, and shape, providing crucial information for accurate diagnosis and treatment. Chen et al. [15] proposed a multi-task U-Net model that achieved a 43.3% JSI on the ISIC 2018 dataset for segmenting all five attributes simultaneously, showing the potential of CNNs in enhancing diagnostic effectiveness. Likewise, Arshad et al. [12] introduced Dermo-Seg, a U-Net model with a ResNet-50 encoder and hybrid loss function, achieving mean JSI scores between 53% and 67% across various attributes on a modified ISIC 2018 dataset.

2.3 Lesion and Attribute Segmentation

Given the importance of detecting the regions discussed in the previous sections, some studies have explored models capable of segmenting the skin lesion and its internal attributes, which are essential for accurate dermatological diagnosis and treatment. Jahanifar et al. [4] introduce a novel skin lesion and

attribute segmentation framework, leveraging transfer learning and CNNs. Their approach utilizes pre-trained CNNs like ResNet152, DenseNet169, Xception, and Inception-ResNet v2 as encoders, with a pyramid pooling module in the decoder for multi-scale information integration. In the ISIC 2018 challenge, their framework achieved a thresholded JSI of 79.6% for lesion segmentation, along with a cumulative JSI of 47.3% for attribute segmentation.

3. Methodology

This paper compared two well-known segmentation models, TerausNet-16 [9] and Mask R-CNN [10], to segment skin lesion and their attributes in dermoscopic images. TerausNet-16, an advanced segmentation model, integrates components from both VGG16 and U-Net presented in Section 1. Its encoder leverages VGG16's convolutional layers to extract intricate features from input images effectively. Meanwhile, inspired by U-Net, the decoder utilizes deconvolutional layers to reconstruct segmented images while preserving spatial information. TerausNet-16 also integrates skip connections between the encoder and decoder, further enhancing segmentation precision. In addition, it benefits from transfer learning with pre-trained weights, enabling expedited training and improved generalization segmentation [9]. Building upon the foundation laid by Faster R-CNN cited in Section 1, the Mask R-CNN takes object detection a step further. While Faster R-CNN identifies ROIs and refines bounding boxes, Mask R-CNN introduces a third step, which generates high-precision binary masks for each detected object, indicating which pixels in the image belong to the ROI and which of them do not [10]. The Mask R-CNN architecture includes a CNN and a segmentation header, composed of a pixels classification network and a mask regression network. The pixels classification network assigns to each pixel in the image a probability of belonging to an object or a background. The mask regression network estimates the probability of each pixel belonging to a specific object, regardless of the previous classification. Both networks are trained along with a combined loss function to ensure the precise classification of the pixels and correct mask generation.

3.1 Datasets

This study utilized two distinct datasets from the ISIC 2018 Challenge [16, 17]: one dedicated to skin lesion segmentation and the other to attribute segmentation. The ISIC 2018 dataset was selected for this study as it is a widely used benchmark in skin lesion analysis, enabling comparison with previous research. Its unique inclusion of annotations for both lesion and attribute segmentation within the same images ensures a consistent and comprehensive analysis. Each dataset includes 2,594 dermoscopic images for training, 100 for validation, and 1,000 for testing. The images are in JPEG format, using the RGB color space, with varying resolutions. The ground truth (GT) masks were provided in PNG format, representing ROIs (lesions or attributes) as white pixels on a black back-

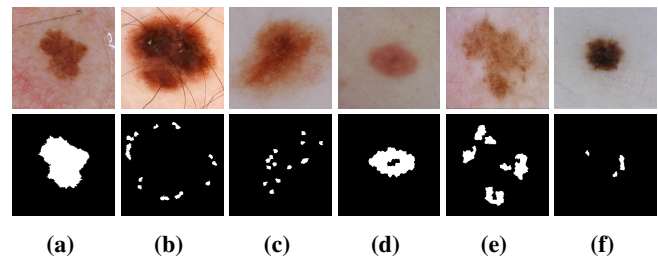


Figure 1. Examples of skin lesions (top), with corresponding GTs to ROI from the entire lesion area or a given attribute area (bottom): (a) Skin lesion, (b) Globules, (c) Milia-like cysts, (d) Negative network, (e) Pigment network, (f) Streaks. Source: [16, 17].

ground. Figure 1 contains examples of skin lesions according to given ROIs.

3.1.1 Skin Lesion Dataset

Each image is paired with a single GT mask that precisely outlines the primary skin lesion. These masks represent the lesion as a single, connected region, excluding any disconnected components or holes. The size and position of the lesion within the image are variable, with lesions potentially occupying the entire image or extending to its borders.

3.1.2 Attribute Dataset

Each image is accompanied by five distinct GT masks. The presence of these attributes is not uniform across all images, leading to an imbalance in the dataset. These masks individually delineate the following specific attributes of the skin lesion [12]: globules, milia-like cyst, negative network, pigment network, and streaks.

3.2 Segmentation Challenges

This paper concentrated on two different segmentation challenges in analyzing skin lesions in dermoscopic images. The first challenge involved the segmentation of entire skin lesions, where the goal was to accurately delineate the outer boundaries of the lesion within the image. This was achieved by employing the Skin Lesion Dataset. The second challenge focused on the segmentation of specific attributes within the lesion, using the Attribute Dataset. The lesion and attribute segmentation methodologies are visually summarized in Figure 2.

3.2.1 Lesion segmentation

Image preprocessing The main adjustment involved resizing the images to a uniform dimension of 256×256 pixels using the bilinear interpolation method [18]. This resizing ensured consistent input dimensions for the CNNs models, facilitating both the training process and the evaluation of their performance.

CNN model training Each model was trained for 50 epochs using the hyperparameter configurations specified in their original papers. For TerausNet-16 model [9], the configuration

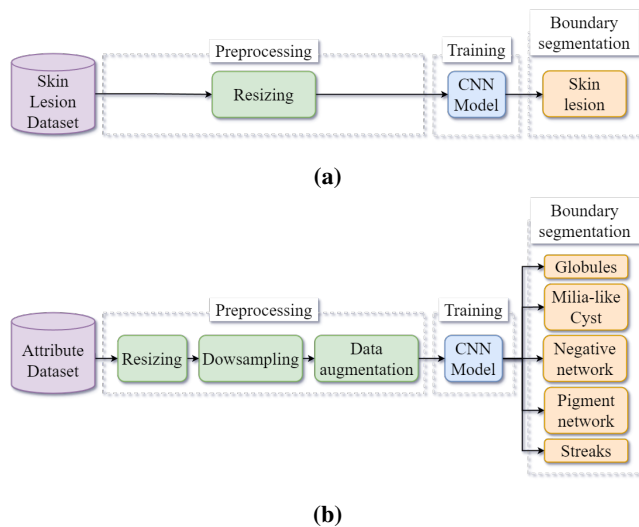


Figure 2. Overview of the proposed methodology: (a) Lesion segmentation and (b) Attribute segmentation. Source: Author's own production.

included 32 initial filters and a kernel size of 3×3 pixels for convolutions. The Adam optimizer was used with an initial learning rate of 0.001. If the validation loss did not improve for 5 consecutive epochs, the learning rate was reduced by a factor of 0.1, with a minimum value of 0.00001. A batch size of 16 was used, and Binary Cross-Entropy served as the loss function to optimize segmentation performance. Additionally, TerausNet-16 used transfer learning by initializing the model with pre-trained weights from the ImageNet dataset [19], enhancing feature extraction and improving generalization [20]. Similarly, Mask R-CNN employed transfer learning, initializing with pre-trained weights from the COCO dataset [21]. Key settings for Mask R-CNN [10] included a ResNet-101 [22] backbone, a batch size of 2, a minimum detection confidence threshold of 0.7, and a non-maximum suppression threshold of 0.3. The network incorporated 256 top-down layers in the feature pyramid, utilized a learning rate of 0.001, and applied weight decay regularization set to 0.0001. The loss function used for Mask R-CNN was a composite loss that combines classification loss, bounding box regression loss, and mask segmentation loss.

3.2.2 Attribute segmentation

Image preprocessing The bilinear interpolation method was applied to resize the images to a uniform dimension of 256×256 pixels for the same reasons presented earlier for the lesion segmentation. However, unlike the lesion segmentation dataset where each image corresponded to a single lesion, not all skin lesion images in the attribute dataset contained all five attributes, which resulted in some segmentation masks being empty. The number of non-empty masks is detailed in Table 1, considering each attribute within Train, Validation (Val.) and Test subsets. Here, a strategic combination of down-sampling and data augmentation was implemented to address the class imbalance in the distribution of non-empty attribute

masks across the training and validation subsets, resulting in a modified dataset. The purpose of these operations was to create a more balanced representation of attributes, eliminating bias towards a specific attribute and diversifying the data [23]. Initially, all images containing only the “Pigment Network” attribute were removed (817 images of the training set and 52 of the validation set). This target downsampling was motivated because the number of non-empty masks for this attribute is nearly equal to the sum of all the others. Subsequently, data augmentation techniques [24] were randomly applied to the remaining attributes that had fewer non-empty masks than the initial average for both the training set (619.60), i.e., “Negative network” and “Streaks”, and the validation set (22.60), which included “Milia-like cysts”, “Negative network”, “Streaks”, and “Globules”. These techniques were applied with specific probabilities, randomly chosen from the following: random cropping (75%), contrast adjustment (90%), brightness changes (50%), color changes (50%), affine transformations (50%), geometric transformations (50%), reflection (50%), and elastic transformations (50%). The effects of these operations on segmentation masks are demonstrated in Table 1 (modified dataset). It is important to highlight that no modification was applied to the test set.

Table 1. Number of non-empty attribute masks in the original and modified dataset.

	Original Dataset			Modified dataset		
	Train	Val.	Test	Train	Val.	Test
Globules	603	19	206	603	21	206
Milia-like cysts	682	6	85	682	20	85
Negative network	190	9	76	622	20	76
Pigment network	1523	73	648	706	21	648
Streaks	100	6	77	604	20	77

CNN model training Two distinct approaches to evaluate model performance under varying conditions were conducted. In the first approach, the CNN models were trained on the entire modified attribute dataset, aiming to segment simultaneously all five attributes within skin lesions. This multi-label segmentation approach enables the models to exploit potential correlations and dependencies between different attributes, fostering a more comprehensive understanding of the lesion's features. In the second approach, a single-label segmentation strategy was employed. The attribute dataset was partitioned into five subsets, each dedicated to a single attribute type. Each model was then trained independently on these subsets, facilitating focused learning on individual attributes. This strategy is particularly advantageous for addressing the class imbalance issue inherent in the original dataset, as it ensures adequate representation of each attribute during training. For both approaches, the hyperparameters and training configuration were consistent with those detailed in Section 3.2.1.

3.3 Comparative analysis

In this paper, widely recognized metrics in the literature were used to evaluate the performance of segmentation models in medical images, which also served as a basis for the ISIC 2018 [16, 17]. The main used metric was the JSI, which measures the overlap between the predicted mask and the GT mask by calculating the ratio of the intersection to the union of these images. This metric provides insight into how well the predicted segmentation matches the reference by considering both false positives and false negatives. To emphasize high-quality segmentations, a thresholded JSI (TJI) is also introduced, considering only segmentations with a JSI score of 65% or higher. Additionally, the Dice Similarity Coefficient (DSC) was used to further assess the similarity between the predicted and GT masks. The DSC is calculated as twice the intersection of the predicted mask and the GT mask divided by the sum of their areas. The evaluation process was performed on a test set of 1,000 images using the JSI and DSC to assess the performance of the lesion and attribute segmentation models. Moreover, both training and test time was also computed as a metric, providing insight into the computational efficiency of the models.

4. Results

The experiments were conducted in Python 3.10 using the Keras library from TensorFlow 2.11, with Numpy 1.23 and scikit-learn 1.20 for data processing and analysis. The models were trained on Google Colaboratory¹, utilizing an Ubuntu 22 system with an Intel Xeon processor, 52 GB RAM, 160 GB storage, and an NVIDIA Tesla T4 GPU with 16 GB of memory. All source code is available in a public repository on GitHub².

4.1 Lesion segmentation

The performance of the TernaNet-16 and Mask R-CNN models was evaluated for the lesion segmentation challenge using the testing subset of the skin lesion dataset. The results, detailed in Table 2, indicate that Mask R-CNN outperformed TernaNet-16 in both JSI and DSC metrics. This exceptional performance is demonstrated in Figure 3, which presents a side-by-side comparison of lesion segmentations produced by both models.

Table 2. Comparative performance of TernaNet-16 and Mask R-CNN in skin lesion segmentation (Best results in bold).

	TJI (%)	JSI (%)	DSC (%)
TernaNet-16	71.59	75.76	78.16
Mask R-CNN	78.64	82.57	84.76

¹<https://colab.research.google.com/>

²https://github.com/antonio-vmoura/skin_cancer

4.2 Attribute segmentation

The TernaNet-16 and Mask R-CNN models were evaluated using the testing subset of the attribute dataset to address the attribute segmentation challenge. The findings, presented in Tables 3 and 4, consistently demonstrated that Mask R-CNN outperformed TernaNet-16. This performance advantage was evident in both multi-label and single-label segmentation approaches, as reflected by higher JSI and DSC scores. Figure 4 provides a visual comparison of the attribute segmentations produced by CNN models.

Table 3. Comparative performance of TernaNet-16 and Mask R-CNN in all attributes segmentation (Best results in bold).

	Original Dataset		Modified Dataset	
	JSI (%)	DSC (%)	JSI (%)	DSC (%)
TernaNet-16	24.36	29.41	33.79	40.35
Mask R-CNN	27.54	34.74	42.86	51.60

5. Discussion

Mask R-CNN consistently demonstrates superior performance across all experiments and evaluation metrics, achieving higher JSI and DSC scores in both lesion and attribute segmentation. In the attribute segmentation challenge, preprocessing techniques play a crucial role: data augmentation enhances generalization for underrepresented attributes like 'Negative Network' and 'Streaks,' leading to a substantial increase in JSI and DSC scores, as shown in Table 3. Meanwhile, downsampling the 'Pigment Network' attribute helps mitigate potential model bias, contributing to a more balanced dataset, although it results in a slight performance decrease for this attribute, as observed in Table 4. This strategic balancing underscores the importance of tailoring preprocessing techniques to optimize segmentation performance and address class imbalances.

The comparison of training times, average testing times per image, and GPU memory (VRAM) usage reveals a substantial difference in the computational demands of the two models. TernaNet-16 completes lesion segmentation in 2 hours and 54 minutes, utilizing an average of 9.74 GB of VRAM, and handles attribute segmentation in 4 hours and 38 minutes, requiring 11.1 GB of VRAM, with average test times of 14 seconds per image for lesion segmentation and 21 seconds per image for attribute segmentation. In contrast, Mask R-CNN, due to its more complex architecture, demands significantly longer training times, taking 4 hours and 29 minutes for lesion segmentation with an average consumption of 12.87 GB of VRAM, and 7 hours and 15 minutes for attribute segmentation using 15.34 GB of VRAM, with average test times of 19 seconds per image for lesion segmentation and 30 seconds per image for attribute segmentation. This extended training and testing duration, along with the higher memory requirements for Mask R-CNN, could limit its practicality in

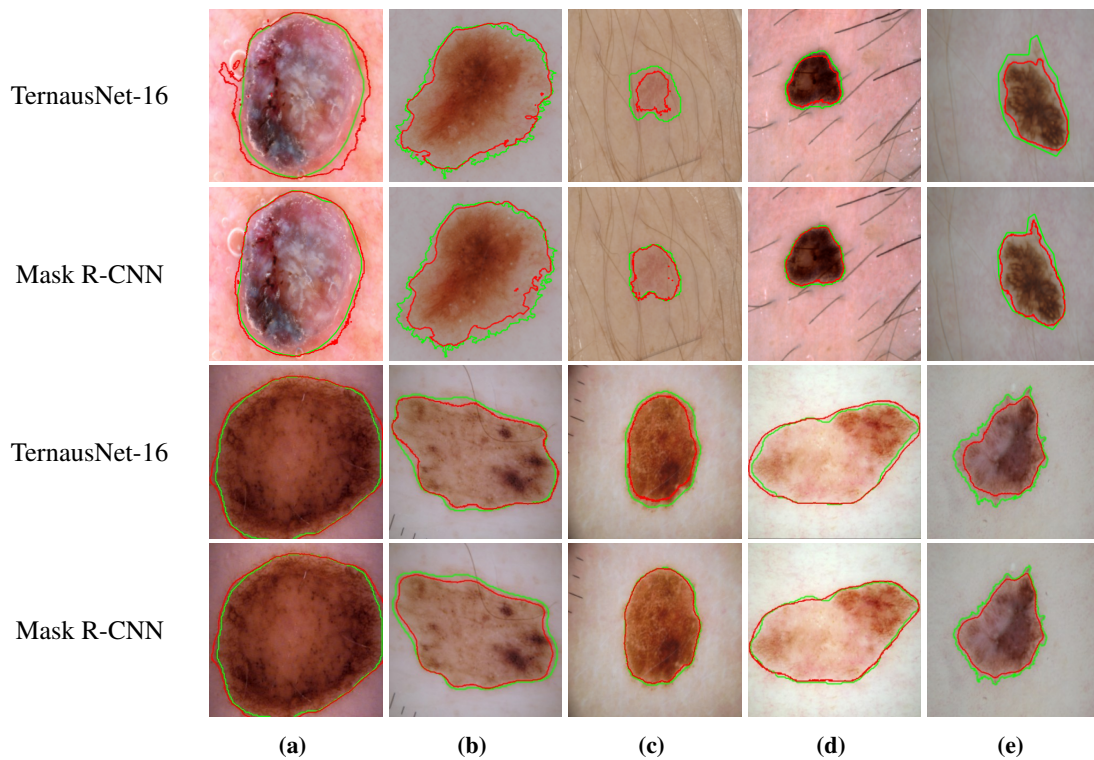


Figure 3. Lesions segmentation results generated by TernaNet-16 and Mask R-CNN: (a-e) Various skin lesions. (Green line = Ground truth; Red line = Model segmentation). Source: Author's own production.

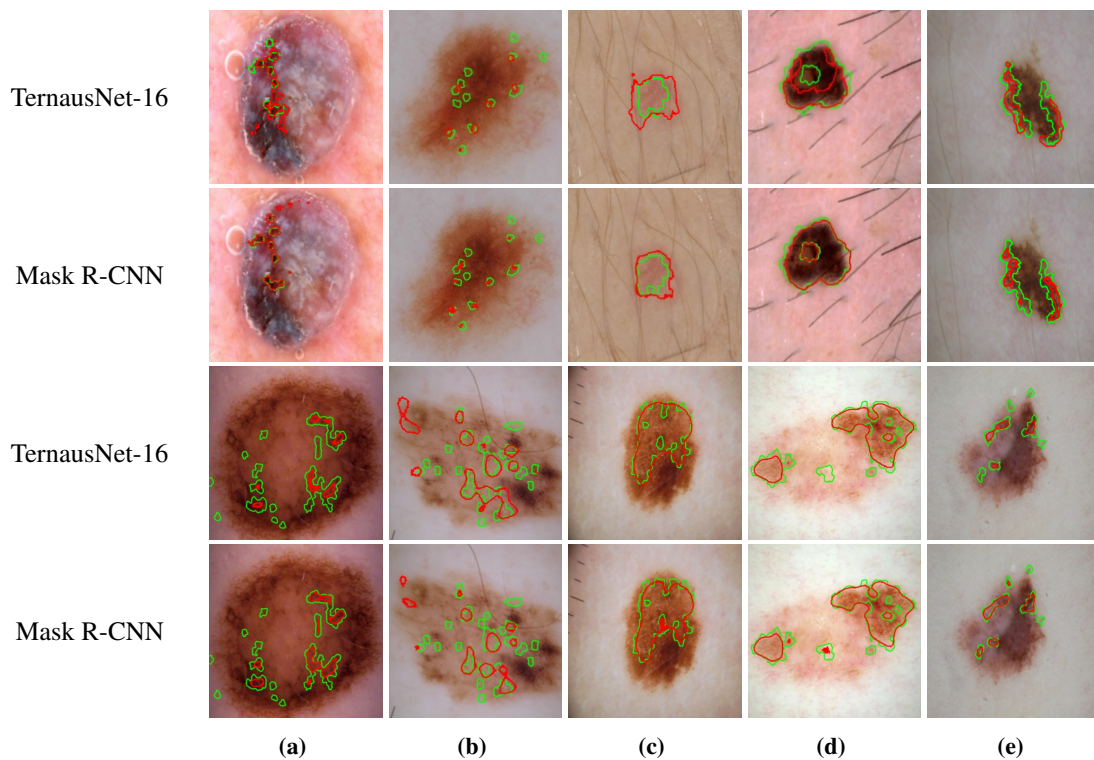


Figure 4. Attributes segmentations results generated by TernaNet-16 and Mask R-CNN: (a) Globules, (b) Milia-like cysts, (c) Negative network (d) Pigment network, (e) Streaks. (Green line = Ground truth; Red line = Model segmentation). Source: Author's own production.

Table 4. Comparative performance of TernaNet-16 and Mask R-CNN in segmenting individual skin lesion attributes (Best results in bold).

	Original dataset				Modified dataset			
	TernaNet-16		Mask R-CNN		TernaNet-16		Mask R-CNN	
	JSI (%)	DSC (%)	JSI (%)	DSC (%)	JSI (%)	DSC (%)	JSI (%)	DSC (%)
Globules	28.24	39.12	40.57	47.15	29.56	39.63	41.07	48.56
Milia-like cysts	32.77	41.70	42.07	50.46	32.06	40.89	41.11	48.81
Negative network	16.07	23.64	21.79	30.82	24.98	31.13	27.22	35.46
Pigment network	46.18	54.21	51.74	62.21	41.37	50.03	45.22	52.48
Streaks	12.24	20.81	18.77	27.13	19.92	28.98	24.12	32.34

scenarios where rapid model development and deployment are crucial.

When comparing Mask R-CNN's performance with other approaches in the literature, as discussed in Section 2, it becomes evident that the model delivers competitive results. For lesion segmentation, Mask R-CNN achieved a TJI of 78.64%, slightly below the 80.2% reported by Qian et al. [14]. Notably, their two-stage method relied on advanced post-processing techniques that were not employed in this approach, highlighting Mask R-CNN's ability to achieve comparable outcomes with a more straightforward methodology. For attribute segmentation, Mask R-CNN achieved a JSI of 42.86% when segmenting all five attributes simultaneously, after addressing class imbalance through downsampling and data augmentation. This result is comparable to the 43.3% JSI achieved by Chen et al. [15] using a multi-task U-Net with classification heads designed to handle sparse labels. These comparisons emphasize the effectiveness of Mask R-CNN in achieving strong segmentation performance for both lesions and attributes, without the need for specialized architectural modifications or complex post-processing strategies.

While this paper provides valuable insights, it is essential to acknowledge certain limitations. The reliance on a relatively small and homogeneous dataset may hinder the generalizability of the findings to broader populations and a wider range of lesion types. Additionally, the superior performance of Mask R-CNN comes at the cost of increased computational demands, including significantly longer training times, which could pose challenges for practical implementation in resource-constrained environments.

6. Conclusion and Future Work

This paper presents a rigorous comparative analysis of TernaNet-16 and Mask R-CNN models for segmenting ROIs in dermoscopic images, specifically focusing on the delineation of skin lesions and their attributes. The results reveal Mask R-CNN consistently outperformed TernaNet-16 across both challenges and all evaluation metrics, demonstrating its superior ability to accurately delineate lesion and attribute boundaries. However, it is important to acknowledge the trade-off between performance and computational efficiency, as Mask

R-CNN requires longer training times due to its complex architecture.

To further advance this research and enhance segmentation models, future work will focus on comparing more recent and diverse models, particularly those incorporating attention mechanisms, such as the Attention U-Net [25]. These mechanisms allow models to prioritize the most relevant regions in an image, which could significantly improve the segmentation of complex lesions [25]. Additionally, exploring ensemble methods that combine the strengths of different architectures may strike a better balance between segmentation effectiveness and computational efficiency. Future efforts should investigate advanced image preprocessing techniques to mitigate artifacts like hairs while highlighting melanoma-specific features, such as irregular borders and color variations, to further improve segmentation quality and overall model performance [4].

Acknowledgment

The first author acknowledges the “Conselho Nacional de Desenvolvimento Científico e Tecnológico” (CNPQ) for the research scholarship.

Author contributions

Antônio Vinicius de Moura Rodrigues: Conducted the literature review, implemented the models, performed the experiments, analyzed the results, and wrote the manuscript. Roberta Barbosa Oliveira: Supervised the research, provided guidance on methodology and experimental design, contributed to the analysis and interpretation of results, and revised the manuscript.

References

- [1] IARC, I. A. for Research on C. *Skin cancer*. 2023. Accessed on: July 14, 2024. Available at: <https://www.iarc.who.int/cancer-type/skin-cancer/#:~:text=Skin%20cancers%20are%20the%20most,people%20died%20from%20the%20disease>.

- [2] MIRIKHARAJI, Z. et al. A survey on deep learning for skin lesion segmentation. *Medical Image Analysis*, Elsevier BV, v. 88, p. 102863, Aug. 2023. ISSN 1361-8415. Available at: <http://dx.doi.org/10.1016/j.media.2023.102863>.
- [3] HESAMIAN, M. H. et al. Deep Learning Techniques for Medical Image Segmentation: Achievements and Challenges. *Journal of Digital Imaging*, v. 32, p. 582–596, 2019. ISSN 1618-727X. Available at: <https://doi.org/10.1007/s10278-019-00227-x>.
- [4] JAHANIFAR, M. et al. Segmentation of Skin Lesions and their Attributes Using Multi-Scale Convolutional Neural Networks and Domain Specific Augmentations. *CoRR*, 2019. Available at: <https://arxiv.org/abs/1809.10243>.
- [5] LÓPEZ-LABRACA, J. et al. An interpretable CNN-based CAD system for skin lesion diagnosis. *Artificial Intelligence in Medicine*, v. 132, p. 102370, 2022. ISSN 0933-3657. Available at: <https://www.sciencedirect.com/science/article/pii/S0933365722001300>.
- [6] RONNEBERGER, O.; FISCHER, P.; BROX, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. *CoRR*, abs/1505.04597, 2015. Available at: <http://arxiv.org/abs/1505.04597>.
- [7] IGLOVIKOV, V.; SHVETS, A. TeraNet: U-Net with VGG11 Encoder Pre-Trained on ImageNet for Image Segmentation. *CoRR*, abs/1801.05746, 2018. Available at: <http://arxiv.org/abs/1801.05746>.
- [8] SIMONYAN, K.; ZISSERMAN, A. Very Deep Convolutional Networks for Large-Scale Image Recognition. *CoRR*, abs/1409.1556, 2014. Available at: <https://api.semanticscholar.org/CorpusID:14124313>.
- [9] SHVETS, A. A. et al. Automatic Instrument Segmentation in Robot-Assisted Surgery using Deep Learning. In: *2018 17th IEEE International Conference on Machine Learning and Applications (ICMLA)*. IEEE, 2018. Available at: <http://dx.doi.org/10.1109/ICMLA.2018.00100>.
- [10] HE, K. et al. Mask R-CNN. In: *2017 IEEE International Conference on Computer Vision (ICCV)*. [S.l.: s.n.], 2017. p. 2980–2988.
- [11] REN, S. et al. Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks. *CoRR*, abs/1506.01497, 2015. Available at: <http://arxiv.org/abs/1506.01497>.
- [12] ARSHAD, S. et al. Dermo-Seg: ResNet-UNet Architecture and Hybrid Loss Function for Detection of Differential Patterns to Diagnose Pigmented Skin Lesions. *Diagnostics*, v. 13, n. 18, 2023. ISSN 2075-4418. Available at: <https://www.mdpi.com/2075-4418/13/18/2924>.
- [13] ALFARO, E. et al. A Brief Analysis of U-Net and Mask R-CNN for Skin Lesion Segmentation. In: . [S.l.: s.n.], 2019. p. 000123–000126.
- [14] QIAN, C. et al. A Detection and Segmentation Architecture for Skin Lesion Segmentation on Dermoscopy Images. *CoRR*, abs/1809.03917, 2018. Available at: <http://arxiv.org/abs/1809.03917>.
- [15] CHEN, E. Z. et al. Lesion Attributes Segmentation for Melanoma Detection with Multi-Task U-Net. In: *2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*. [S.l.: s.n.], 2019. p. 485–488.
- [16] TSCHANDL, P.; ROSENDAHL, C.; KITTLER, H. The HAM10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. *Scientific Data*, v. 5, n. 1, p. 180161, 2018. ISSN 2052-4463. Available at: <https://doi.org/10.1038/sdata.2018.161>.
- [17] CODELLA, N. C. F. et al. Skin Lesion Analysis Toward Melanoma Detection 2018: A Challenge Hosted by the International Skin Imaging Collaboration (ISIC). *CoRR*, abs/1902.03368, 2019. Available at: <http://arxiv.org/abs/1902.03368>.
- [18] PARSANIA, P.; VIRPARIA, P. A review: Image interpolation techniques for image scaling. *International Journal of Innovative Research in Computer and Communication Engineering*, v. 02, p. 7409–7414, 01 2015.
- [19] DENG, J. et al. ImageNet: A large-scale hierarchical image database. p. 248–255, June 2009. ISSN 1063-6919.
- [20] LI, L.; SEO, W. Deep Learning and Transfer Learning for Skin Cancer Segmentation and Classification. In: *2021 IEEE 21st International Conference on Bioinformatics and Bioengineering (BIBE)*. [S.l.: s.n.], 2021. p. 1–5.
- [21] LIN, T. et al. Microsoft COCO: Common Objects in Context. *CoRR*, abs/1405.0312, 2014. Available at: <http://arxiv.org/abs/1405.0312>.
- [22] HE, K. et al. Deep Residual Learning for Image Recognition. *CoRR*, abs/1512.03385, 2015. Available at: <http://arxiv.org/abs/1512.03385>.
- [23] PEREZ, F. et al. Data Augmentation for Skin Lesion Analysis. In: *Lecture Notes in Computer Science*. Springer International Publishing, 2018. p. 303–311. Available at: https://doi.org/10.1007/978-3-030-01201-4_33.
- [24] ALOMAR, K.; AYSEL, H.; CAI, X. Data augmentation in classification and segmentation: A survey and new strategies. *Journal of Imaging*, v. 9, p. 46, 02 2023.
- [25] OKTAY, O. et al. *Attention U-Net: Learning Where to Look for the Pancreas*. 2018. Available at: <https://arxiv.org/abs/1804.03999>.