

Anti-Aliasing Attention U-net Model for Skin Lesion Segmentation

Subjects: [Computer Science](#), [Artificial Intelligence](#)

Contributor: [Phuong Thi Le](#) , [Bach-Tung Pham](#) , [Ching-Chun Chang](#) , [Yi-Chiung Hsu](#) , [Tzu-Chiang Tai](#) , [Yung-Hui Li](#) , [Jia-Ching Wang](#)

The need for a lightweight and reliable segmentation algorithm is critical in various biomedical image-prediction applications. However, the limited quantity of data presents a significant challenge for image segmentation. Additionally, low image quality negatively impacts the efficiency of segmentation, and previous deep learning models for image segmentation require large parameters with hundreds of millions of computations, resulting in high costs and processing times.

computer-aided diagnosis

skin lesion segmentation

deep learning

1. Introduction

Melanoma is one of the most prevalent types of cancer. It is the primary factor in most skin cancer deaths. In the United States in 2022, 99,780 cases were diagnosed as new melanoma, 57,180 in men and 42,600 in women. There were 7560 deaths by melanoma, with 5080 in men and 2570 in women. Compared with the report in 2020 by Worldwide, the number of patients diagnosed with melanoma was 324,635 people. Melanoma is considered as one of the deadliest types of skin cancer; early detection can extend survival by five years and increase the overall rate of survival. Thus, the demand for the diagnosis of melanoma in its early stages is increasing significantly. Furthermore, automated skin segmentation can be widely applied in many sensitive applications such as face tracking [\[1\]](#), face detection [\[2\]](#), and gesture recognition [\[3\]](#). Other samples can be listed as content-based retrieval [\[4\]](#), robotics [\[5\]](#), virtual reality [\[6\]](#), face recognition [\[7\]](#), and human-computer interaction [\[8\]](#).

One solution for predicting melanoma is based on visual examination by dermatologists. However, this approach has become less popular due to its low accuracy, time-consuming nature, and reliance on human factors and trained experts. With the advancement of technology, machine learning and deep learning have emerged as promising techniques for predicting melanoma and other applications. The hope is that these methods can make the diagnosis of melanoma simpler, better, and more convenient. Deep learning has been applied to the prediction of skin segmentation, which supports melanoma diagnosis and significantly increases the efficiency of skin segmentation. The results also indicated that the effectiveness of previous approaches were dependent on equipment, color information of input datasets, personal factors, nonlinear illumination, and the availability of datasets. There are several factors that can affect skin lesion images, including the position of the lesion on the body, the type of equipment used to capture the image, and the presence of noise in the image [\[9\]](#)[\[10\]](#). Noise can manifest in various forms, such as ink spots, fuzzy borders, black hair, undesirable lighting, lesion artifacts, brown

lesions, black hair with brown lesions, markers, white residues on the lesion, and hard scenes for the lesion. These factors can all pose significant challenges for accurate skin lesion segmentation. In order to improve segmentation performance, it is important to address these challenges and develop methods that can effectively handle the various types of noise and variability present in skin lesion images. Some factors affecting lesion skin images are as follows:

- Illumination: the skin color is affected by the level of illumination (shadows, nonwhite lights, and whether the image was captured indoors or outdoors).
- Camera equipment: the sensor quality of the camera directly impacts the color image.
- Ethnicity: people across different regions, as well as ethnic groups, have different skin colors.
- Individual characteristics: age and working environment also influence skin color.
- Other factors: background, makeup, or glasses also affect skin color.

Figure 1 illustrates some examples from skin lesion datasets with poor quality input images containing hair, noise from light, and resistance images. Moreover, the limited quantity of available images decreases the accuracy of skin segmentation. Prior studies have attempted to address these challenges by applying preprocessing steps to improve the quality of the input datasets. These preprocessing algorithms include resizing the input images, removing black hair, eliminating markers, and processing nondominant lesions and ink spots to enhance segmentation accuracy. However, these preprocessing steps can also increase the computational load on the system and prolong the processing time. In addition, if not carefully implemented, preprocessing algorithms can result in a loss of critical features from the input data.



Figure 1. Some skin lesion images with noise.

Another difficulty of skin lesion segmentation is uncertainty at the boundaries of the lesion, as described in **Figure 2**.



Figure 2. Uncertainty at the boundaries of the lesion.

Dermoscopy is a technique used to enhance the information obtained from skin lesions. By using a dermatoscope, information about the lesion can be seen more clearly, such as the presence of structures, colors, and patterns. However, even with dermoscopy, accurately delineating the boundary of the lesion remains a challenge, as well as determining the difference between healthy and lesion skin. These challenges highlight the need for further improvement in skin lesion segmentation techniques. Some prior studies tried to handle this problem. They used traditional segmentation methods, namely thresholding, colour-based segmentation algorithms, discontinuity-based segmentation, and region-based segmentation. Thresholding determines the threshold and divides the pixels into groups [11]. Color-based segmentation algorithms are based on color-discrimination segment images according to principle components and spherical coordinate transforms [12]. Discontinuity-based segmentation uses radial searching techniques or Laplacian or Gaussian zero crossing for segmentation [13]. Region-based segmentation [14] splits an image into small parts and segments by statistical region merging and multiregion growing. Recently, deep learning models have been published to perform skin segmentation. U-net is an example that is constructed with decoder and encoder paths [15]. Due to its simplicity and effectiveness, the u-net architecture was proposed in 2015 for biomedical image segmentation and has been considered state-of-the-art in this field ever since. Moreover, DeeplabV3 probes convolutional features at multiple scales corresponding to image-level features that encode the global context [16]. The feature pyramid network (FPN) is the most popular model for segmentation. It was designed with a pyramid of features to learn features at low and high resolutions [17]. E-net was designed from the ground up specifically for segmentation [18]. Mask R-CNN, which was introduced in 2017, has shown superior performance in instance segmentation tasks, where the goal is to detect and segment objects within an image [19].

The basic structure of all of these deep learning models consists of two main parts: an encoder and a decoder, also known as autoencoders. The encoder takes an input and maps it to a lower-dimensional representation, also known as a bottleneck. The decoder then takes this lower-dimensional representation and reconstructs the original input as closely as possible. The bottleneck is a crucial part of an autoencoder, as it forces the network to learn a compressed representation of the input data. This compressed representation can be used for various downstream tasks, such as clustering, classification, or anomaly detection. Autoencoders have several applications in various fields, such as the segmentation of images, language generation, translation, and sentiment analysis. In addition, autoencoders can be used for feature extraction, where the bottleneck is used to extract the most important

features of the input data. This is particularly useful in domains such as computer vision and signal processing, where the input data can be highly dimensional. Although these models have primarily focused on improving the quality of datasets and segmentation efficiency for skin lesion segmentation, some of these studies have relied heavily on color cues, which can limit performance and lead to bias related to skin tones. Additionally, the availability of powerful GPUs has enabled more efficient deep learning processes. However, deep models with large numbers of parameters can be computationally intensive and impractical for use on personal computers. Therefore, there is a growing need for a lightweight model with fewer parameters that can be easily run on a personal computer. This problem is considered a subset of object segmentation and has been less addressed in previous studies.

2. Implement Deep Learning for Segmentation

Traditional machine learning and deep learning methods are the two main approaches to skin lesion segmentation. For example, the authors used an ant colony as the traditional machine learning method to extract three lesion features: texture, color, and geometrical properties. They then used the K-nearest neighbour (KNN) technique to segment skin lesions [20]. Jaisakthi et al. [21] introduced combining the grab-cut algorithm and KNN for skin lesion segmentation. Adjed et al. [22] proposed the use of wavelet transforms and curvelet transforms to extract features, and then used SVM to classify skin lesion. Arroyo et al. [23] presented fuzzy histogram thresholding for segmenting skin lesions. Do et al. [24] applied hierarchical segmentation for skin lesion segmentation. Alfed et al. [25] integrated the HOG and SVM methods for skin lesion segmentation. Zhou et al. [26] segmented skin lesions using a gradient flow-based algorithm. Although these traditional machine learning methods can segment skin lesions, these methods need help with large datasets.

In truth, the development of computers and available big data urged researchers to implement deep learning for segmentation. Deep learning methods have achieved better performances in skin lesion segmentation. U-net, which is built upon FCNs with encoder and decoder paths, was considered as the state-of-the-art model for image segmentation. The total number of parameters in the u-net mode is approximately 31.03 million. The success of u-net in image segmentation led to the development of several updated models that built upon its architecture and introduced further advancements. For instance, DenseNetUnet [27] was built with numerous dilated convolutions with different atrous rates to observe a wide field of vision and utilized an attention mechanism with a better performance than that of u-net. The cons of the DenseUnet model are its huge number of parameters (it has 251.84 million parameters). Another example, called u-net 3+, is a modification of u-net that utilizes comprehensive skip connections and deep supervision [28]. The full-scale skip connections combine low-level details with high-level semantics from feature maps in multiple sizes. ResUnet++ is an extension of the u-net architecture that incorporates residual connections in the encoder and the decoder paths [29]. DoubleU-Net is another variant of the u-net architecture that includes Atrous Spatial Pyramid Pooling (ASPP) to selectively highlight contextual features [30]. Deep Residual u-net is a variant of the u-net architecture that contains residual connections and deep supervision to improve segmentation accuracy [31]. Furthermore, mask-RCNN is also a recent state-of-the-art model for segmentation. A mask-RCNN architecture includes an extension of a faster R-CNN that combines an

existing path for bounding box identification with an approach for estimating an object mask. Segnet uses pooling in upsampling path indices calculated in the maximum pooling stage of the matching downsampling path [32]. VGG-16 focuses on an increased depth with a small filter size that has 138 million parameters [33]. Laddernet has skip connections between each pair of adjacent upsampling and downsampling paths in each level and many pairings of downsampling and upsampling paths. Deeplab v3 is NMF with an L0-constraint on the H columns (NMF L0-H) [34]. The total number of parameters in Deep lab v3 is 43.9 million parameters. Inception V3 utilizes a small filter size and has better feature representation [35]. The AASC model also has empirical results for skin lesion segmentation [36]. Recently, a novel boundary-aware transformer (BAT) has also shown impressive results for skin segmentation based on the development of transformer-based systems and a boundary-wise level of knowledge [37]. When comparing the diagnostic accuracy of an experienced dermatologist, a clinician with minimal dermoscopy training, and a computer-aided diagnosis, it was found that the former two had lower sensitivity (69%) than the computer (92%) [38]. Overall, these deep learning methods perform better skin segmentation than traditional machine learning methods, but the model sizes are large with computational complexity. Thus, these deep learning models have slow training and inference times. Furthermore, these models require significant preprocessing to handle varying input sizes. These deep learning models are prone to overfitting when the training dataset is small.

References

1. Zhou, M.; Liang, L.; Sun, J.; Wang, Y. AAM based face tracking with temporal matching and face segmentation. In Proceedings of the 2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, San Francisco, CA, USA, 13–18 June 2010; IEEE: Piscataway, NJ, USA, 2010; pp. 701–708.
2. Viola, P.; Jones, M.J. Robust real-time face detection. *Int. J. Comput. Vis.* 2004, 57, 137–154.
3. Mitra, S.; Acharya, T. Gesture recognition: A survey. *IEEE Trans. Syst. Man Cybern. Part C Appl. Rev.* 2007, 37, 311–324.
4. Tagare, H.D.; Jaffe, C.C.; Duncan, J. Medical image databases: A content-based retrieval approach. *J. Am. Med. Inform. Assoc.* 1997, 4, 184–198.
5. Garcia, E.; Jimenez, M.A.; De Santos, P.G.; Armada, M. The evolution of robotics research. *IEEE Robot. Autom. Mag.* 2007, 14, 90–103.
6. Burdea, G.C.; Coiffet, P. *Virtual Reality Technology*; John Wiley Sons: Hoboken, NJ, USA, 2003.
7. Zhao, W.; Chellappa, R.; Phillips, P.J.; Rosenfeld, A. Face recognition: A literature survey. *ACM Comput. Surv.* 2003, 35, 399–458.
8. Mitsunaga, N.; Miyashita, T.; Ishiguro, H.; Kogure, K.; Hagita, N. Robovie-IV: A communication robot interacting with people daily in an office. In Proceedings of the 2006 IEEE/RSJ International

- Conference on Intelligent Robots and Systems, Beijing, China, 9–15 October 2006; IEEE: Piscataway, NJ, USA, 2006; pp. 5066–5072.
9. Masood, A.; Al-Jumaily, A. Computer aided diagnostic support system for skin cancer: A review of techniques and algorithms. *Int. J. Biomed. Imaging* 2013, 2013, 323268.
 10. Kakumanu, P.; Makrogiannis, S.; Bourbakis, N. A survey of skin-color modeling and detection methods. *Pattern Recognit.* 2007, 40, 1106–1122.
 11. Kohler, R. A segmentation system based on thresholding. *Comput. Graph. Image Process.* 1981, 15, 319–338.
 12. Tayal, Y.; Lamba, R.; Padhee, S. Automatic face detection using color based segmentation. *Int. J. Sci. Res. Publ.* 2012, 2, 1–7.
 13. Joshi, G.D.; Sivaswamy, J.; Krishnadas, S.R. Depth discontinuity-based cup segmentation from multiview color retinal images. *IEEE Trans. Biomed. Eng.* 2012, 59, 1523–1531.
 14. Gould, S.; Gao, T.; Koller, D. Region-based segmentation and object detection. *Adv. Neural Inf. Process. Syst.* 2009, 22, 655–663.
 15. Guth, F.; de Campos, T.E. Skin lesion segmentation using U-Net and good training strategies. *arXiv* 2018, arXiv:1811.11314.
 16. Chen, L.C.; Papandreou, G.; Schroff, F.; Adam, H. Rethinking atrous convolution for semantic image segmentation. *arXiv* 2017, arXiv:1706.05587.
 17. Lin, T.Y.; Dollár, P.; Girshick, R.; He, K.; Hariharan, B.; Belongie, S. Feature pyramid networks for object detection. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, Honolulu, HI, USA, 21–26 July 2017; pp. 2117–2125.
 18. Paszke, A.; Chaurasia, A.; Kim, S.; Culurciello, E. Enet: A deep neural network architecture for real-time semantic segmentation. *arXiv* 2016, arXiv:1606.02147.
 19. Vuola, A.O.; Akram, S.U.; Kannala, J. Mask-RCNN and U-net ensembled for nuclei segmentation. In *Proceedings of the 2019 IEEE 16th International Symposium on Biomedical Imaging (ISBI 2019)*, Venice, Italy, 8–11 April 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 208–212.
 20. Ma, Z.; Tavares, J.M.R. Effective features to classify skin lesions in dermoscopic images. *Expert Syst. Appl.* 2017, 84, 92–101.
 21. Jaisakthi, S.M.; Mirunalini, P.; Aravindan, C. Automated skin lesion segmentation of dermoscopic images using GrabCut and k-means algorithms. *IET Comput. Vis.* 2018, 12, 1088–1095.
 22. Adjed, F.; Safdar Gardezi, S.J.; Ababsa, F.; Faye, I.; Chandra Dass, S. Fusion of structural and textural features for melanoma recognition. *IET Comput. Vis.* 2018, 12, 185–195.

23. Garcia-Arroyo, J.L.; Garcia-Zapirain, B. Segmentation of skin lesions in dermoscopy images using fuzzy classification of pixels and histogram thresholding. *Comput. Methods Programs Biomed.* 2019, 168, 11–19.
24. Do, T.T.; Hoang, T.; Pomponiu, V.; Zhou, Y.; Chen, Z.; Cheung, N.M.; Koh, D.; Tan, A.; Tan, S.H. Accessible melanoma detection using smartphones and mobile image analysis. *IEEE Trans. Multimed.* 2018, 20, 2849–2864.
25. Alfed, N.; Khelifi, F. Bagged textural and color features for melanoma skin cancer detection in dermoscopic and standard images. *Expert Syst. Appl.* 2017, 90, 101–110.
26. Zhou, H.; Li, X.; Schaefer, G.; Celebi, M.E.; Miller, P. Mean shift based gradient vector flow for image segmentation. *Comput. Vis. Image Underst.* 2013, 117, 1004–1016.
27. Li, X.; Chen, H.; Qi, X.; Dou, Q.; Fu, C.W.; Heng, P.A. H-DenseUNet: Hybrid densely connected UNet for liver and tumor segmentation from CT volumes. *IEEE Trans. Med. Imaging* 2018, 37, 2663–2674.
28. Huang, H.; Lin, L.; Tong, R.; Hu, H.; Zhang, Q.; Iwamoto, Y.; Han, X.; Chen, Y.W.; Wu, J. Unet 3+: A full-scale connected unet for medical image segmentation. In *Proceedings of the ICASSP 2020—2020 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, Barcelona, Spain, 4–8 May 2020; IEEE: Piscataway, NJ, USA, 2020; pp. 1055–1059.
29. Jha, D.; Smedsrud, P.H.; Riegler, M.A.; Johansen, D.; De Lange, T.; Halvorsen, P.; Johansen, H.D. Resunet++: An advanced architecture for medical image segmentation. In *Proceedings of the 2019 IEEE International Symposium on Multimedia (ISM)*, San Diego, CA, USA, 9–11 December 2019; IEEE: Piscataway, NJ, USA, 2019; pp. 225–2255.
30. Sun, Y.; Bi, F.; Gao, Y.; Chen, L.; Feng, S. A multi-attention UNet for semantic segmentation in remote sensing images. *Symmetry* 2022, 14, 906.
31. Venkatesh, G.M.; Naresh, Y.G.; Little, S.; O'Connor, N.E. A deep residual architecture for skin lesion segmentation. In *OR 2.0 Context-Aware Operating Theaters, Computer Assisted Robotic Endoscopy, Clinical Image-Based Procedures, and Skin Image Analysis: First International Workshop, OR 2.0 2018, 5th International Workshop, CARE 2018, 7th International Workshop, CLIP 2018, Third International Workshop, ISIC 2018, Held in Conjunction with MICCAI 2018*, Granada, Spain, 16–20 September 2018, *Proceedings 5*; Springer International Publishing: Berlin/Heidelberg, Germany, 2018; pp. 277–284.
32. Badrinarayanan, V.; Kendall, A.; Cipolla, R. Segnet: A deep convolutional encoder-decoder architecture for image segmentation. *IEEE Trans. Pattern Anal. Mach. Intell.* 2017, 39, 2481–2495.
33. Geng, L.; Zhang, S.; Tong, J.; Xiao, Z. Lung segmentation method with dilated convolution based on VGG-16 network. *Computer Assisted Surgery. Comput. Assist. Surg.* 2019, 24 (Suppl. S2),

27–33.

34. Zhuang, J. LadderNet: Multi-path networks based on U-Net for medical image segmentation. arXiv 2018, arXiv:1810.07810.
35. Punn, N.S.; Agarwal, S. Inception u-net architecture for semantic segmentation to identify nuclei in microscopy cell images. *ACM Trans. Multimed. Comput. Commun. Appl.* 2020, 16, 1–15.
36. Le, P.T.; Chang, C.C.; Li, Y.H.; Hsu, Y.C.; Wang, J.C. Antialiasing Attention Spatial Convolution Model for Skin Lesion Segmentation with Applications in the Medical IoT. *Wirel. Commun. Mob. Comput.* 2022, 2022, 1278515.
37. Wang, J.; Wei, L.; Wang, L.; Zhou, Q.; Zhu, L.; Qin, J. Boundary-aware transformers for skin lesion segmentation. In *Proceedings of the International Conference on Medical Image Computing and Computer-Assisted Intervention, Strasbourg, France, 27 September–1 October 2021*; Springer: Cham, Switzerland, 2021; pp. 206–216.
38. Piccolo, D.; Ferrari, A.; Peris, K.E.T.T.Y.; Daidone, R.; Ruggeri, B.; Chimenti, S. Dermoscopic diagnosis by a trained clinician vs. a clinician with minimal dermoscopy training vs. computer-aided diagnosis of 341 pigmented skin lesions: A comparative study. *Br. J. Dermatol.* 2002, 147, 481–486.

Retrieved from <https://encyclopedia.pub/entry/history/show/107195>