Combining XAI and graph cuts for skin-lesion segmentation

Zyad Husni, Uwe Jaekel and Babette Dellen

University of Applied Sciences Koblenz,
Faculty of Mathematics, Informatics, Technology,
53424 Remagen, Germany
{zhusni,jaekel,dellen}@hs-koblenz.de

Abstract. Deep neural networks and supervised machine learning for medical image segmentation, including dermatology [10], require large pixel-wise annotated datasets for training, which can be difficult to obtain. Image classification, on the other hand, only requires a label for each image, which is often automatically provided with a medical diagnosis, but does not provide segmentation maps. However, in image-classification tasks, Explainable-AI (XAI) algorithms provide a means of identifying pixels in the original image that are part of the object or relevant structure. We propose to exploit this information for segmenting the images by building a network graph from XAI explanations and using the graph-cut algorithm for segmentation. Our approach is evaluated using the HAM10k [26] dataset, demonstrating its ability to segment skin lesions in dermatoscopic images without requiring pixel-annotated data for training. This makes our approach a cost-effective alternative in scenarios where annotated images are not available.

Keywords: Image Segmentation · Supervised Machine Learning · Explainable AI · Graph-Cut · SHAP · Grad-CAM · GMM · ResNet152 · CBAM.

1 Introduction

Medical image segmentation is a crucial task in the diagnosis and treatment of diseases, especially in dermatology, where accurate delineation of skin lesions is essential for early detection of conditions such as melanoma [4,13]. Supervised machine learning has become an important and frequent paradigm in dermatology [10]. One of the most widely used methods is the U-Net [21], which shows exceptional performance in medical image segmentation. However, it relies heavily on large, annotated datasets, which are costly and time-consuming to produce.

Clustering-based approaches have been proposed to address this challenge [23,11], but a drawback is their dependence on pixel intensity, which often requires careful tuning of parameters. Other approaches rely on user-defined inputs that provide markers for initialization and are therefore not automatic [25]. More recent unsupervised deep learning techniques aim to learn feature representations

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

directly from the data but can be computationally intensive and require large datasets [29].

Skin-lesion segmentation has also been recently formulated as an anomaly detection problem [8]. This has the advantage that no annotated data is required, because the network is trained on images showing only healthy skin. However, this approach does not provide information about the type of anomaly that has been detected, e.g., different stages of a disease or classes.

Our approach addresses these challenges by proposing a fully automated segmentation framework that integrates Explainable-AI techniques [3], specifically SHAP [15,24] and Grad-CAM [22,17], with Gaussian Mixture Models [14,19] (GMM) and Graph-Cut [6,5]. Images are annotated with a single label, representing different classes, and then used to train a classification network. XAI-driven feature maps provide pixel-wise information on the importance of the pixel for the correct classification of the image. These feature maps are used to initialize a graph-cuts method for image segmentation [6]. The XAI-driven feature maps generated by SHAP and Grad-CAM guide the segmentation process, while the GMM model the intensity distributions of the foreground and background. Graph-Cut further refines the segmentation by minimizing the energy function, ensuring that the object boundaries are accurately delineated. This approach avoids tedious pixel-wise annotations of images and therefore provides an automatic and scalable solution for medical image-segmentation tasks where labeled data are scarce or unavailable.

The primary objectives of this study are (i) the development and training of a medical image-segmentation framework combining XAI-driven feature maps from SHAP and Grad-CAM with GMM and graph cuts, (ii) the evaluation of the proposed approach on the HAM10k [26] dataset as well as comparison of the results with segmentation results obtained by U-Net [21], and (iii) the investigation of the impact of CBAM [27] integrations on segmentation quality and model transparency.

2 Methods

The proposed image-segmentation framework consists of the following parts:

(i) The core of the method is a ResNet152 model for image classification, trained with a subset of the HAM10K dataset from the ISIC_2018 challenge [2]. During training, the network is provided only with the class label for each image, not ground-truth pixel-wise annotations. Consequently, the network outputs only a single label prediction for each image.
(ii) SHAP and Grad-CAM explanations, obtained for the ResNet15 model of step (i), are used to create binary masks for each image. Grad-CAM explanations are further improved by integrating CBAM into ResNet152.
(iii) The binary masks from step (ii) are merged and used to initialize the graph-cuts method. The resulting graph-cuts segmentations are the XAI-driven segmentations produced by our method.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

A schematic overview of the method is provided in Fig. 1. In the following, we provide further details of the different techniques employed in our framework.

2.1 Datasets and data augmentation

The HAM10K dataset from the ISIC_2018 challenge [2] contains dermatoscopic images of skin lesions, categorized into seven classes: melanoma (MEL), nevus (NV), actinic keratosis(AKIEC), basal cell carcinoma (BCC), benign ker-atosis-like lesions (BKL), vascular lesions (VASC), and dermatofibroma (DF). A key characteristic of this dataset is its class imbalance. For example, the most prevalent class, NV, includes over 58 times more images than the least common, DF. The original dataset contained 10,015 training images (each with a corresponding ground-truth segmentation mask), 194 validation images (without ground-truth segmentations), and 1,503 test images (without ground-truth segmentations). Our goal is to generate U-Net and XAI-driven segmentations and evaluate their accuracy using the Dice [30] and Jaccard [20] scores. These scores compare the generated segmentations to the ground-truth segmentations. Since the test and validation set lacked ground-truth segmentations, we restricted our analysis to the original training set of 10,015 images, which was then divided into 3 parts, defining dataset A, consisting of 8,012 images for testing, 202 images for validation during training, and 1,803 images for testing.

To enhance the dataset and improve model performance, image augmentation techniques were applied to the training and validation dataset using the PyTorch deep learning library [18]. For augmentation, the Albumentations library [1,9] was chosen due to its extensive range of augmentation options and efficient implementation. To address class imbalance in the dataset, a weighted augmentation strategy was employed. Using Albumentations, images from underrepresented classes were specifically targeted for augmentation. This strategy was based on inverse class frequency weighting or adaptive data augmentation, which ensured that minority classes received more augmentation than dominant classes. The goal was to balance the distribution of classes in the training data and improve the generalizability of the model. Not all images were augmented; the focus was primarily on the minority classes. Data augmentation resulted in 29,213 training images and 4,190 validation images. The number of testing images remained unchanged, defining dataset B.

2.2 Model training and feature extraction

We use a ResNet152 [12] network as the backbone for feature extraction. ResNet [12] introduces the concept of skip connections, improving training stability and convergence.

The model was trained on dataset A and on the augmented dataset B. Both trainings were conducted on one of the HPC compute nodes of the University of Applied Sciences Koblenz, equipped with 2 AMD EPYC 7713 CPUs (64 cores, 128 threads each), 1 TB RAM, and 4 NVIDIA A100 GPUs (80 GB each) connected via NVLink. The training configuration included a learning rate of 0.1

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

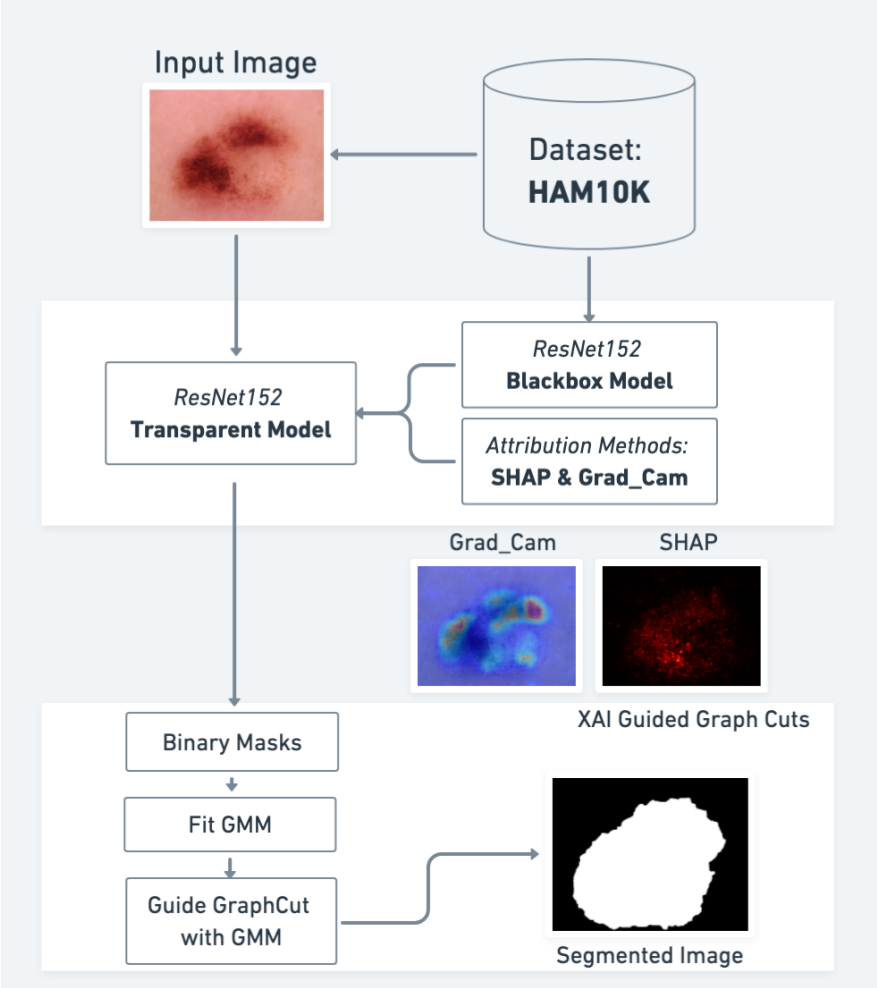

Fig. 1: Overview of the pipeline: Images are augmented and used to train a ResNet-152 as a black-box classifier. The transparent model processes each input to produce Grad-CAM and SHAP attribution maps, which provide an initial binary mask. A Gaussian Mixture Model is then fitted to both foreground and background regions and used to guide a graph-cut algorithm for image segmentation.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

with a learning rate scheduler, momentum set to 0.8, and a weight decay of 0.001. An early stopping strategy was employed to save the best model during training, preventing overfitting and ensuring better generalization.

Class	**Precision**	**Recall**	**F1-Score**	**Support**
Class 0 (MEL)	0.71	0.57	0.63	158
Class 1 (NV)	0.93	0.97	0.95	1248
Class 2 (AKEIC)	0.72	0.85	0.78	81
Class 3 (BCC)	0.81	0.57	0.67	68
Class 4 (BKL)	0.83	0.73	0.78	201
Class 5 (DF)	0.78	0.74	0.76	19
Class 6 (VASC)	0.82	0.96	0.89	28
Accuracy			0.88	1803
Macro avg	0.80	0.77	0.78	1803
Weighted avg	0.88	0.88	0.88	1803
Test Loss			0.5447	

Table 1: Classification results of the ResNet152 model, trained with dataset B, for the test set.

2.3 Initial segmentation-mask generation using XAI

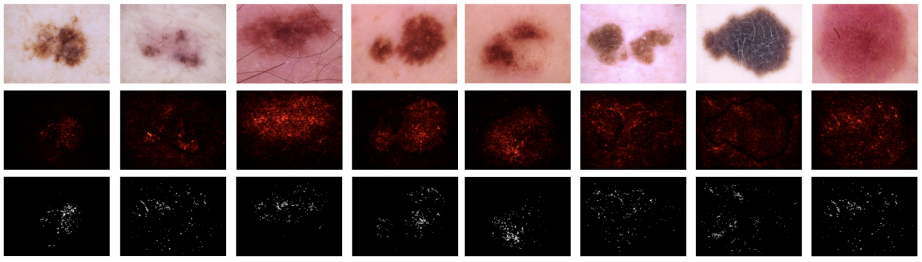

Fig. 2: The first row displays the original images from the HAM10K dataset, followed by the raw SHAP attributions extracted from the transparent model in the second row. The third row shows the corresponding binary masks generated from the SHAP values.

The learned feature maps from the ResNet152 model serve as input for the XAI methods, SHAP and Grad-CAM. SHAP values are used to generate the importance of pixel-level features, highlighting the most influential regions for image classification [15,24]. Grad-CAM utilizes the feature maps generated by

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

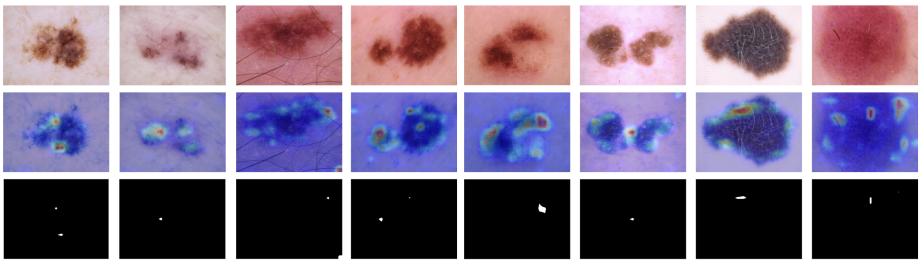

Fig. 3: The first row presents the original images from the HAM10K dataset, followed by the raw Grad-CAM attributions extracted from the transparent model in the second row. The third row displays the corresponding binary masks generated from the Grad-CAM explanations.

the final convolutional layers of the neural network to create coarse localization maps that highlight areas of interest (see Fig. 3, second row). These maps indicate regions that contributed the most to the predictions, providing valuable insights into where the model is focusing its attention [17]. The outputs from SHAP and Grad-CAM are combined to create an initial binary mask, leveraging SHAP's pixel-level importance and Grad-CAM's spatial localization. This complementary approach addresses the limitations of each method individually, with Grad-CAM providing broader context and SHAP capturing finer details. A binary bitwise AND operation using the OpenCV [7] library is applied to merge the masks, ensuring that only pixels significant to both methods contribute to the final combined mask, improving segmentation accuracy.

2.4 CBAM

To further enhance the feature extraction capabilities of the ResNet152 backbone, we integrated the Convolutional Block Attention Module (CBAM) [27] into the architecture to improve Grad-CAM explanations. CBAM is a lightweight and effective attention mechanism that improves a network's ability to focus on the most informative features within an image. It operates by applying attention mechanisms sequentially along the channel and spatial dimensions of the feature maps, refining the extracted features at each stage. The *channel attention module* identifies the importance of each channel by aggregating spatial information using global average pooling and max pooling. The resulting descriptors are passed through a shared multi-layer perceptron (MLP) that includes a bottleneck layer, where the dimensionality is reduced by a factor known as the reduction ratio. This reduction controls the trade-off between model complexity and the ability to capture fine-grained details. The outputs of the MLP are combined and activated using a sigmoid function to produce the channel attention map, which emphasizes the most relevant channels in the feature map. The *spatial attention module* focuses on the most critical regions in the image. It applies pooling operations along the channel axis to generate spatial descriptors, which

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

are concatenated and passed through a convolutional layer to produce a spatial attention map. This map highlights the important regions in the feature map, guiding the network to focus on the areas most relevant for segmentation. Integrating CBAM into ResNet152 enhances the network's ability to focus on critical areas, such as lesion boundaries, in dermatoscopic images. To accommodate the complexity of the dataset, the reduction ratio in CBAM is set to 4 instead of the default 16, and the kernel size for spatial attention is increased to 5 [16].

2.5 Segmenting with Graph-Cut

The binary masks obtained from SHAP and Grad-CAM are merged and then used to initialize Gaussian Mixture Models (GMM) for foreground and background modeling. GMMs are fitted to the pixel distributions of the image to compute foreground and background probabilities, where the GMM models the intensity and color distribution of the foreground and background regions in the combined mask. This is determined by extracting the coordinates of the black (background) and white (foreground) pixels from the mask and mapping these pixels back onto the original image to identify their respective regions. The resulting foreground and background probability distributions allow estimating the likelihood of each pixel belonging to the foreground or background. The image is treated as a graph where each pixel corresponds to a node. The source and the sink of the graph (terminal nodes) are provided by the foreground and background probabilities computed earlier via the binary XAI mask [6]. Edge weights are calculated from intensity, texture, and gradient similarities between neighboring pixels. These weights influence the energy-minimization process and ensure accurate segmentation boundaries. The Graph-Cut algorithm then minimizes the energy function that defines the relationship of foreground and background regions. This iterative minimization process refines the segmentation mask, effectively separating the object of interest from the background [6].

3 Results

To obtain XAI-guided image segmentation, the ResNet152 model is first trained using the augmented dataset B. The combined outputs of SHAP and Grad-CAM explanations are used to initialize the graph-cut method for image segmentation. We report intermediate results for the different parts of our framework as well as the final XAI-guided segmentation performance.

3.1 ResNet152 model for image classification

The ResNet152 model, trained with dataset A, achieved an accuracy of 88% on the test set. To evaluate model performance, metrics such as precision, recall, and F1-score were used. The minority classes DF and BCC underperformed compared to the other classes. Therefore, we trained with augmented dataset B

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

instead, which improved the results. The overall results are summarized in table 1. This model provides the core of the proposed framework.

3.2 Binary-mask generation with SHAP and Grad-CAM

SHAP and Grad-CAM explanations are generated for each image of the test set using the previously trained ResNet152 model. In Fig. 2 and Fig. 3, the computed SHAP and Grad-CAM explanations are shown in the second row, respectively. In the third row, the respective binary masks generated from the explanations are presented. The binary masks obtained from Grad-CAM are very sparse, making them less suitable for binary-mask generation. By integrating CBAM [27] into the ResNet152 model, the Grad-CAM explanations and the resulting binary mask could be improved (see Fig. 4, last two rows). These adjustments allow the model to capture subtle features more effectively, improving the overall segmentation performance.

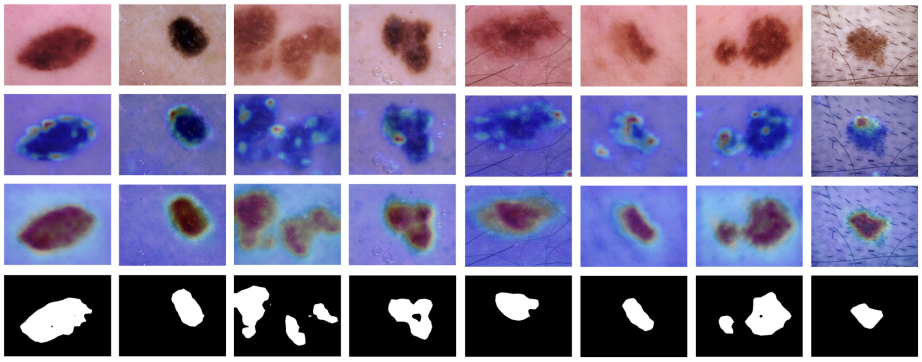

Fig. 4: Improvements of Grad-CAM explanations via integrating CBAM into ResNet152: Original images (first row), Grad-CAM explanations for ResNet152 without CBAM (second row), Grad-CAM explanations for ResNet152 with CBAM (third row), and binary masks generated Grad-CAM explanations for ResNet152 with CBAM (last row).

3.3 XAI-guided image segmentation

Finally, the binary masks obtained in the previous step are merged and used to initialize the graph-cuts method. The resulting binary segmentations are evaluated using standard metrics: Dice Coefficient [30] and Jaccard Index [20]. These metrics quantify the overlap of predicted and ground-truth segmentations provided by experts. Our segmentation framework achieved a Dice score of 0.84 and a Jaccard index of 0.75 (see table 2). In Fig. 5, segmentation results are shown for images of the test set. These results show that our framework is able

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

to classify and segment skin lesions without using pixel-wise annotations during training.

The performance of the proposed method is compared with U-Net, a widely adopted model for medical image segmentation. The U-Net architecture uses ResNet152 as the backbone for its encoder-decoder framework, with the implementation sourced from GitHub [28]. Both networks were trained with the augmented dataset B. However, to train the U-Net, the image class labels and the ground-truth segmentations had to be provided during training, while our method is trained only with image class labels. The U-Net achieved a Dice score of 0.92 and a Jaccard index of 0.85 (see table 2). The score of our method is lower than the one of U-Net, but still in a comparable range. However, a comparison in terms of the Dice and Jaccard score alone is difficult for the following reason: Ground-truth pixel-wise annotations, which are used to train the U-Net and to evaluate segmentation performance of both approaches on the test set, potentially contain a bias. Since the U-Net is trained with this data, it can learn this bias. Therefore, the boundaries of the segments are closer to the precise form of the ground-truth annotations, raising the Dice and the Jaccard scores. When studying the results for individual images in Fig. 5, our method shows, at least visually, competitive results to U-Net.

Method	**Dice Score**	**Jaccard Index**
U-Net	0.92	0.85
XAI segmentation	0.84	0.75

Table 2: Comparison of Dice and Jaccard metrics between U-Net and the proposed approach (XAI segmentation) for the test set.

4 Discussion and Conclusion

This study presents an unsupervised medical image-segmentation approach that leverages XAI techniques, such as SHAP and Grad-CAM, to guide the Gaussian Mixture Model (GMM) initialization for Graph-Cut segmentation. The model was first trained on an augmented dataset, achieving a high accuracy of 88%. XAI explanations were then generated and converted into binary masks to identify key foreground and background pixels. These masks were used to fit GMMs, which provided the initial segmentation. The process was iteratively repeated for a specified number of iterations n or until a convergence threshold was met, measuring the ratio of changed pixels to the total number of pixels. The final output was further refined using morphological operations. For challenging cases, such as lesions that closely resemble healthy skin, improvements were introduced, including the integration of CBAM into the training architecture and adjustments to

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

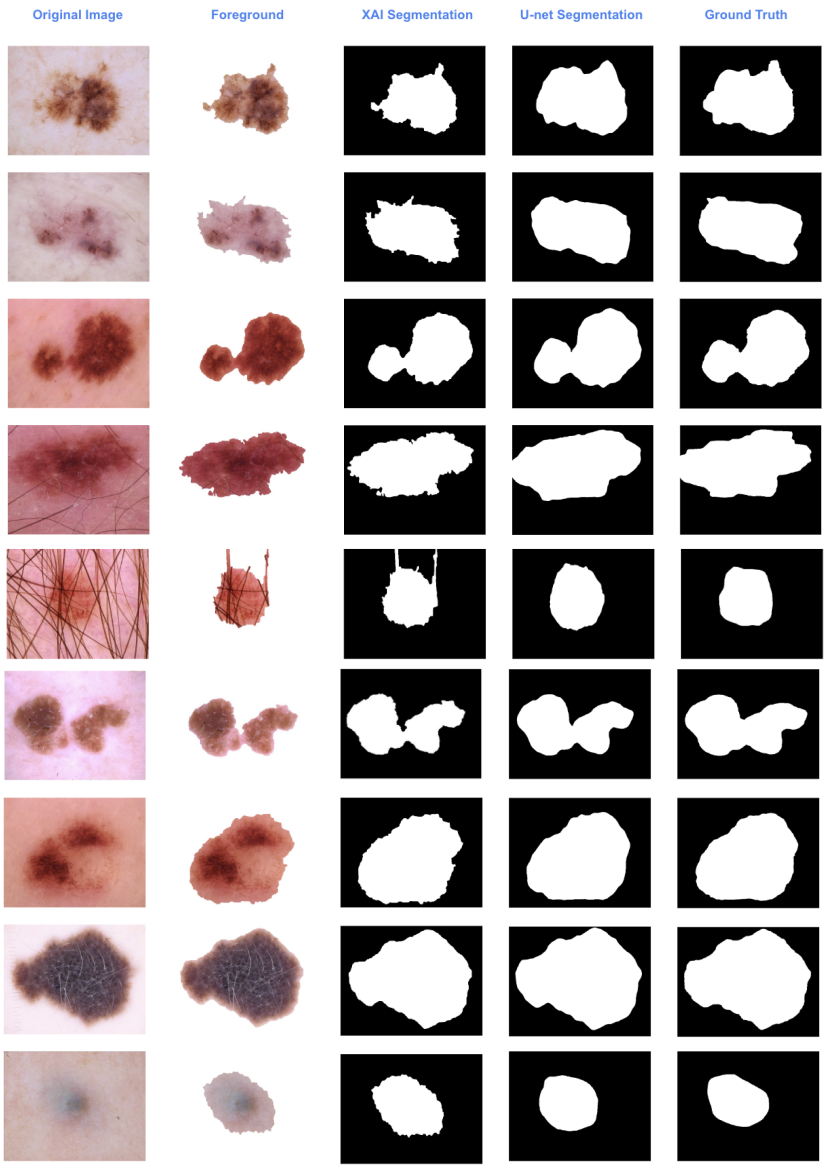

Fig. 5: Example results of our method (XAI-segmentation), U-net and the ground-truth.

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

the number of GMM components for foreground and background modeling. Despite the class imbalance in the HAM10k dataset, the proposed method showed a performance comparable to U-Net. It further eliminates the need for pixel-wise expert annotations, which is an important advantage in scenarios where annotated data are scarce or unavailable.

The approach has limitations due to several parameters that influence its performance, including those related to model training, threshold selection for values generated by XAI methods such as SHAP and Grad-CAM, and parameters of GMM components for foreground and background modeling. These limitations become especially apparent for lesions that are difficult to distinguish from healthy skin. In such cases, the raw attribution maps generated by SHAP and Grad-CAM can differ significantly due to the distinct underlying computation methods of each technique. Consequently, merging these masks can result in suboptimal binary masks, which negatively impact the subsequent GMM segmentation performance (see Fig. 6).

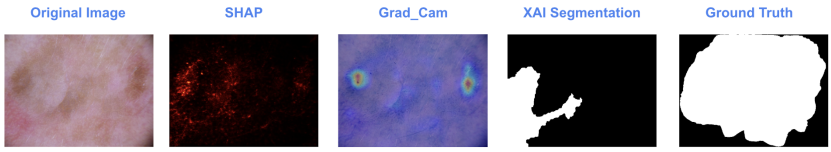

Fig. 6: Limitations of the method: Segmentation fails when the attribution maps differ significantly from one another. This discrepancy leads to a segmented output that does not adequately cover the region of interest.

Future research will focus on improving the binary mask generation process, optimizing GMM fitting by testing multiple configurations with the goal of selecting the best combination of components for foreground and background modeling. Furthermore, multiple models could be trained independently for specific lesion classes to obtain a more detailed understanding of each class. By isolating the segmentation process for particular lesion classes, such as malignant versus benign cases, models could capture the unique characteristics of each category in more detail, potentially enhancing overall performance.

Acknowledgments. This research has received funding from the Ministry of Science and Health of Rhineland-Palatinate, Germany, and the Debeka Krankenversicherungsverein a.G. through the Forschungskolleg Data2Health.

References

1. Albumentations Team: Defining a simple augmentation pipeline for image augmentation (2024), `https://albumentations.ai/docs/examples/example/`, accessed: 2024-08-10

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

2. Archive, I.C.: Isic challenge archive (nd), https://challenge.isic-archive.com/, accessed: December 13, 2024
3. Arrieta, A.B., Díaz-Rodríguez, N., Del Ser, J., Bennetot, A., Tabik, S., Barbado, A., Garcia, S., Gil-Lopez, S., Molina, D., Benjamins, R., Chatila, R., Herrera, F.: Explainable artificial intelligence (xai): Concepts, taxonomies, opportunities and challenges toward responsible ai. Information Fusion **58**, 82–115 (2020). https://doi.org/10.1016/j.inffus.2019.12.012
4. Behera, N., Singh, A.P., Rout, J.K., Balabantaray, B.K.: Melanoma skin cancer detection using deep learning-based lesion segmentation. International Journal of Information Technology **16**(6), 3729–3744 (2024)
5. Boykov, Y., Funka-Lea, G.: Graph cuts and efficient nd image segmentation. International journal of computer vision **70**(2), 109–131 (2006)
6. Boykov, Y., Jolly, M.P.: Interactive graph cuts for optimal boundary and region segmentation of objects in n-d images. Proceedings of IEEE International Conference on Computer Vision (ICCV) **1**, 105–112 (2001)
7. Bradski, G.: The opencv library. Dr. Dobb's Journal of Software Tools (2000)
8. Burgert, A., Dellen, B., Jaekel, U., Paulus, D.: Semi-supervised anomaly detection in skin-lesion images. In: Proceedings of the 20th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications, International Conference on Computer Vision Theory and Applications (Visapp 2025). vol. 2, pp. 535–541. SciTePress (2025)
9. Buslaev, A., Iglovikov, V.I., Khvedchenya, E., Parinov, A., Druzhinin, M., Kalinin, A.A.: Albumentations: Fast and flexible image augmentations. Information **11**(2), 125 (2020). https://doi.org/10.3390/info11020125
10. Chan, S., Reddy, V., Myers, B., Thibodeaux, Q., Brownstone, N., Liao, W.: Machine learning in dermatology: current applications, opportunities, and limitations. Dermatology and therapy **10**, 365–386 (2020)
11. Dhanachandra, N., Manglem, K., Chanu, Y.: Image segmentation using k-means clustering algorithm and subtractive clustering algorithm. Procedia Computer Science **54**, 764–771 (2015)
12. He, K., Zhang, X., Ren, S., Sun, J.: Deep residual learning for image recognition. In: Proceedings of the IEEE conference on computer vision and pattern recognition. pp. 770–778 (2016)
13. Khan, M.Z., Gajendran, M.K., Lee, Y., Khan, M.A.: Deep neural architectures for medical image semantic segmentation: Review. IEEE Access **9**, 83002–83024 (2021). https://doi.org/10.1109/ACCESS.2021.3086530
14. Lei, T., Nandi, A.K.: Image Segmentation: Principles, Techniques, and Applications. John Wiley & Sons (2022)
15. Lundberg, S.M., Lee, S.I.: A unified approach to interpreting model predictions. In: Proceedings of the 31st International Conference on Neural Information Processing Systems (NeurIPS). pp. 4765–4774 (2017)
16. Luuuyi: Cbam pytorch implementation (nd), https://github.com/luuuyi/CBAM.PyTorch/tree/master, accessed: Nov 11, 2024
17. Molnar, C.: Interpretable machine learning. Lulu. com (2020)
18. Paszke, A., Gross, S., Massa, F., Lerer, A., Bradbury, J., Chanan, G., Killeen, T., Lin, Z., Gimelshein, N., Antiga, L., Desmaison, A., Kopf, A., Yang, E., DeVito, Z., Raison, M., Tejani, A., Chilamkurthy, S., Steiner, B., Fang, L., Bai, J., Chintala, S.: Pytorch: An imperative style, high-performance deep learning library. In: Advances in Neural Information Processing Systems (NeurIPS). vol. 32, pp. 8024–8035 (2019)
19. Rasmussen, C.: The infinite gaussian mixture model. Advances in neural information processing systems **12** (1999)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5

20. Real, E., Shlens, J., Mazzocchi, S., Pan, Y., Le, Q.V.: You might not need higher order penalties for semantic segmentation: Revisiting optimal graph cut with object compatibility. Proceedings of the IEEE International Conference on Computer Vision (ICCV) pp. 2684–2693 (2017)
21. Ronneberger, O., Fischer, P., Brox, T.: U-net: Convolutional networks for biomedical image segmentation. In: Medical Image Computing and Computer-Assisted Intervention (MICCAI). Lecture Notes in Computer Science, vol. 9351, pp. 234–241. Springer (2015). https://doi.org/10.1007/978-3-319-24574-4_28
22. Selvaraju, R.R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., Batra, D.: Grad-cam: Visual explanations from deep networks via gradient-based localization. In: Proceedings of the IEEE International Conference on Computer Vision (ICCV). pp. 618–626 (2017). https://doi.org/10.1109/ICCV.2017.74
23. Siddiqui, F.U., Yahya, A.: Clustering techniques for image segmentation. Springer (2022)
24. Sundararajan, M., Najmi, A.: The many shapley values for model explanation. In: International conference on machine learning. pp. 9269–9278. PMLR (2020)
25. Tang, Y., Li, Y., Zou, H., Zhang, X.: Interactive segmentation for medical images using spatial modeling mamba. Information **15**(10), 633 (2024)
26. Tschandl, P., Rosendahl, C., Kittler, H.: The ham10000 dataset, a large collection of multi-source dermatoscopic images of common pigmented skin lesions. Scientific data **5**(1), 1–9 (2018)
27. Woo, S., Park, J., Lee, J.Y., Kweon, I.S.: Cbam: Convolutional block attention module. In: Proceedings of the European conference on computer vision (ECCV). pp. 3–19 (2018)
28. Yakubovskiy, P.: Segmentation models for pytorch (2024), https://github.com/qubvel-org/segmentation_models.pytorch, accessed: July 18, 2024
29. Zhang, Q., Yang, L.T., Chen, Z.: Deep computation model for unsupervised feature learning on big data. IEEE Transactions on Services Computing **9**(1), 161–171 (2016). https://doi.org/10.1109/TSC.2015.2497705
30. Zou, K.H., Warfield, S.K., Bharatha, A., Tempany, C.M.C., Kaus, M.R., Haker, S.J., Wells, W.M., Jolesz, F.A., Kikinis, R.: Statistical validation of image segmentation quality based on a spatial overlap index: Scientific reports. Academic Radiology **11**(2), 178–189 (2004)

ICCS Camera Ready Version 2025
To cite this paper please use the final published version:
DOI: 10.1007/978-3-031-97557-8_5