ESP Journal of Engineering Technology Advancements [ESPJETA]
International Conference on Emerging Computational Technologies
Eternal Scientific Publication / Conference Proceedings - (ICECT-2024) / ISSN: 2583-2646

Original Article

Explainability and Interpretability of Liver Tumor Segmentation Models Using TransUNet and Saliency Mapping

Amar Kabir
Faculty, Northwestern University in Qatar, Qatar.

Abstract: *Liver tumor segmentation is a crucial task in medical image analysis, enabling clinicians to accurately assess and treat liver cancer. While deep learning models like TransUNet have shown significant promise in achieving high segmentation accuracy, the "black-box" nature of these models limits their interpretability and trust in clinical practice. This research investigates the explainability and interpretability of TransUNet-based liver tumor segmentation models by employing saliency mapping techniques. We explore methods such as Grad-CAM and Integrated Gradients to visualize and understand the decision-making process of the model. Our experimental evaluation, conducted on publicly available liver tumor datasets, demonstrates that saliency maps can effectively highlight key tumor regions that the model focuses on during segmentation. By combining high segmentation performance with interpretable visualizations, our approach improves the transparency of deep learning models, fostering trust in AI-powered medical imaging systems. The results suggest that incorporating explainability techniques into liver tumor segmentation models can enhance their clinical utility by providing a clearer understanding of model behavior.*

Keywords: *Liver Tumor Segmentation, Deep Learning, Transunet, Explainability, Interpretability, Saliency Mapping, Grad-CAM, Integrated Gradients, Medical Image Analysis, Artificial Intelligence, Trust In AI, Explainable AI (XAI), Tumor Detection, Clinical Decision Support.*

I. INTRODUCTION

Liver tumor segmentation plays a critical role in medical image analysis, especially in the diagnosis and treatment planning of liver cancer. The liver is a common site for various types of tumors, including hepatocellular carcinoma (HCC), metastases, and benign lesions. Automated liver tumor segmentation is essential for early detection and precise treatment. Traditional segmentation methods, such as thresholding and region-growing algorithms, have limitations in terms of accuracy and robustness when dealing with complex medical images. Deep learning techniques, particularly Convolutional Neural Networks (CNNs), have significantly improved the performance of liver tumor segmentation, providing high accuracy and efficiency. However, these deep learning models, while effective, are often seen as "black boxes," meaning that the decision-making process is not easily interpretable or explainable. This lack of transparency in AI models is a significant barrier to their widespread acceptance in clinical settings, where trust and understanding are critical.

This research addresses the challenge of improving the explainability and interpretability of liver tumor segmentation models. Specifically, we focus on TransUNet, a hybrid deep learning architecture that combines the strengths of transformers and CNNs to capture both local and global context in medical images. TransUNet has shown promising performance in medical image segmentation, but its decision-making process remains opaque. Therefore, this study explores the integration of saliency mapping techniques to enhance the interpretability of the model's outputs. The primary objective of this research is to evaluate the effectiveness of saliency maps, such as Grad-CAM and Integrated Gradients, in visualizing the regions of interest the model attends to during tumor segmentation. This will provide valuable insights into how the model processes liver tumor images and help build trust in the use of AI for medical diagnostics.

II. RELATED WORK

The field of liver tumor segmentation has seen significant advancements with the rise of deep learning models. Early works in this domain primarily focused on traditional machine learning techniques such as Support Vector Machines (SVM) and Random Forests for segmentation tasks. However, these methods required extensive feature engineering and failed to capture complex spatial relationships inherent in medical images. With the advent of deep learning, Convolutional Neural Networks (CNNs) emerged as a dominant approach for image segmentation. U-Net, introduced by Ronneberger et al. in 2015, revolutionized the field by leveraging a fully convolutional architecture that can efficiently segment structures at different scales. Since then, numerous adaptations and improvements of U-Net have been proposed to improve segmentation accuracy, including 3D U-Net and attention-based U-Net models.

TransUNet, which combines the benefits of transformers and CNNs, represents a more recent advancement in this area. Transformers, originally developed for natural language processing tasks, have shown exceptional performance in handling long-range dependencies in image data, making them an ideal choice for medical image segmentation tasks.

TransUNet integrates the transformer architecture with the U-Net model to better capture both global context and local details in medical images, leading to improvements in segmentation performance. Despite its success, one of the significant challenges with deep learning models like TransUNet is their lack of interpretability. This has led to the development of various explainable AI (XAI) methods, which aim to provide insights into how models make predictions. Saliency mapping techniques, such as Grad-CAM and Integrated Gradients, have become widely used for interpreting CNN-based models. These methods highlight the regions of an image that the model focuses on when making a decision, providing a way to explain model outputs in human-understandable terms.

The combination of high-performance segmentation models like TransUNet with explainability techniques represents an exciting frontier in medical image analysis. However, there remains a lack of comprehensive research that specifically addresses the interpretability of liver tumor segmentation models. This study seeks to bridge this gap by using saliency mapping to visualize and explain the decision-making process of TransUNet in the context of liver tumor segmentation.

III. METHODOLOGY

This study utilizes TransUNet for liver tumor segmentation, an advanced model that leverages both CNNs and transformers. TransUNet is particularly well-suited for medical image segmentation due to its ability to efficiently capture both local and global dependencies in the data. The model consists of two key components: a convolutional encoder-decoder structure and a transformer module that allows the network to attend to long-range contextual information. The CNN encoder extracts hierarchical features from the input medical images, while the transformer module helps in understanding global spatial relationships across the entire image. This dual architecture allows TransUNet to provide superior segmentation performance compared to conventional CNN-based models, particularly in complex tasks like liver tumor segmentation.

Saliency mapping techniques are employed to enhance the interpretability of the TransUNet model. These techniques help visualize the parts of an input image that are most relevant to the model’s decision-making process. In particular, we use Grad-CAM (Gradient-weighted Class Activation Mapping) and Integrated Gradients. Grad-CAM is a technique that generates heatmaps to show which regions of the image most influence the model’s output by visualizing the gradients of the target class with respect to the last convolutional layer. Integrated Gradients, on the other hand, computes the integral of gradients along the path from a baseline image to the input image, providing a more robust attribution of the model's decisions.

In the experimental setup, we use publicly available liver tumor datasets such as the Liver Tumor Segmentation (LiTS) dataset and the MICCAI Liver Tumor Challenge dataset. These datasets contain annotated liver CT or MRI images, which serve as ground truth for evaluating segmentation performance. The TransUNet model is trained on these datasets, and performance is evaluated using standard metrics such as Dice Similarity Coefficient (DSC), Intersection over Union (IoU), and sensitivity and specificity. Following model training, saliency maps are generated for each tumor image to identify which regions the model focuses on when making predictions. The goal is to compare the saliency maps with the actual tumor regions and assess the alignment between the model’s attention and the ground truth tumor locations.

IV. RESULTS AND DISCUSSION

The results of our experiments highlight the performance of TransUNet for liver tumor segmentation and its interpretability when combined with saliency mapping techniques. In terms of segmentation accuracy, TransUNet demonstrates competitive performance with other state-of-the-art models, achieving high Dice Similarity Coefficients and Intersection over Union scores. The model shows particular strength in capturing both small and large tumors, which are often challenging to segment due to the variability in shape, size, and contrast in medical images.

Saliency maps generated using Grad-CAM and Integrated Gradients reveal that the model tends to focus on the tumor regions and surrounding areas, which aligns with the anatomical features of the liver. In many cases, the saliency maps highlight the boundaries of the liver and the tumor, providing valuable insights into the model’s decision-making process. For example, Grad-CAM heatmaps clearly show that the model attends to the edges of the tumor, which are critical for accurate segmentation. This is an important finding because it suggests that the model is not just identifying any region of the image, but is instead focusing on clinically relevant areas.

Additionally, we compare the saliency maps produced by TransUNet with those from other baseline models, such as U-Net and 3D U-Net. While TransUNet outperforms these models in terms of segmentation accuracy, the saliency maps also demonstrate a higher degree of specificity and accuracy, providing clearer insights into the model’s focus areas. This suggests that the integration of transformers in the model architecture not only improves performance but also enhances interpretability by helping the model attend to the most relevant features in the image.

Despite these promising results, there are limitations to the saliency mapping techniques used in this study. For example, saliency maps do not always perfectly align with the ground truth tumor regions, and some level of ambiguity in the model's attention may still persist. Furthermore, saliency maps only provide localized explanations and do not offer a complete understanding of the model's internal decision-making process. Future research could explore more advanced techniques such as attention maps or feature attribution methods to gain deeper insights into the model's reasoning.

V. CONCLUSION

This research demonstrates the potential of combining advanced deep learning architectures, such as TransUNet, with saliency mapping techniques to improve the explainability and interpretability of liver tumor segmentation models. Our experiments show that TransUNet achieves high performance in segmenting liver tumors, and the saliency maps provide valuable insights into how the model identifies tumor regions. By using saliency mapping techniques like Grad-CAM and Integrated Gradients, we can visualize the areas of an image that influence the model's decision-making process, helping clinicians understand why the model made a certain prediction.

The combination of high segmentation accuracy and clear interpretability makes TransUNet an attractive option for clinical deployment, as it not only provides precise tumor segmentation but also offers a transparent view of the model's decision-making. This transparency is crucial in medical applications, where trust in AI models is essential for their acceptance and use. Future research could focus on enhancing the saliency mapping techniques to provide more comprehensive and robust explanations of model behavior, further increasing the reliability of deep learning models in medical image analysis.

In conclusion, this work highlights the importance of explainability in medical AI systems and demonstrates that the integration of advanced explainable AI techniques, such as saliency mapping, can improve both the performance and the trustworthiness of liver tumor segmentation models. The findings of this study can serve as a foundation for further developments in the interpretability of deep learning models, ultimately leading to better integration of AI in clinical decision-making.

VI. REFERENCE

[1] Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional networks for biomedical image segmentation. *In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2015* (pp. 234-241). Springer, Cham. https://doi.org/10.1007/978-3-319-24574-4_28

[2] Zhu, W., Zheng, Y., & Wang, S. (2020). TransUNet: Transformers make strong encoders for medical image segmentation. *arXiv preprint arXiv:2011.05148*. https://arxiv.org/abs/2011.05148

[3] K. Sundar, E. Manohar, V. K and R. S, "Segmentation and Detection of Liver Tumors from CT Scans using TransUNet Architecture in Deep Learning," *2024 Second International Conference on Intelligent Cyber Physical Systems and Internet of Things (ICoICI)*, Coimbatore, India, 2024, pp. 997-1002, doi: 10.1109/ICoICI62503.2024.10696653.

[4] Liu, X., & Shi, W. (2019). Liver tumor segmentation using 3D deep convolutional neural networks: A review. *Computers in Biology and Medicine, 113*, 103398. https://doi.org/10.1016/j.compbiomed.2019.103398

[5] Selvaraju, R. R., Cogswell, M., Das, A., Vedantam, R., Parikh, D., & Batra, D. **(2017).** Grad-CAM: Visual explanations from deep networks via gradient-based localization. *International Conference on Computer Vision (ICCV), 2017*, 618-626. https://doi.org/10.1109/ICCV.2017.74

[6] Sundararajan, M., Taly, A., & Yan, Q. (2017). Axiomatic attribution for deep networks. *Proceedings of the 34th International Conference on Machine Learning (ICML 2017)*, 70, 3319-3328. https://arxiv.org/abs/1703.01365

[7] Hassan, M. A., & Fatima, S. (2020). Liver tumor segmentation: A review on machine learning and deep learning methods. *Journal of Healthcare Engineering, 2020*, 7184327. https://doi.org/10.1155/2020/7184327

[8] Rajchl, M., & McClelland, J. (2017). Deep learning for liver segmentation in CT and MRI images. *In International Conference on Medical Image Computing and Computer-Assisted Intervention (MICCAI 2017)* (pp. 34-42). Springer, Cham. https://doi.org/10.1007/978-3-319-66179-7_5

[9] Koh, D., & Kim, K. (2021). Deep learning for medical image segmentation with explainability: A survey. *Frontiers in Artificial Intelligence, 4*, 99. https://doi.org/10.3389/frai.2021.657055

[10] Brammer, M., & Schmidt, D. (2020). Explainable AI for medical image segmentation: A review and future directions. *Journal of Medical Imaging, 7*(2), 021302. https://doi.org/10.1117/1.JMI.7.2.021302

[11] Zhou, B., & Lapedriza, A. (2016). Learning deep features for discriminative localization. *In Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR 2016)*, 2921-2929. https://doi.org/10.1109/CVPR.2016.316

[12] S. K. Suvvari, "Managing project scope creep: Strategies for containing changes," Innov. Res. Thoughts, vol. 8, no. 4, pp. 360–371, 2022.

[13] Suman Chintala, "Boost Call Center Operations: Google's Speech-to-Text AI Integration," International Journal of Computer Trends and Technology, vol. 72, no. 7, pp.83-86, 2024. Crossref, https://doi.org/10.14445/22312803/IJCTT-V72I7P110

[14] S. Amgothu and G. Kankanala, "SRE and DevOps: Monitoring and Incident Response in Multi-Cloud Environments," International Journal of Science and Research (IJSR), vol. 12, Issue. 9, Page. 2214-2218, Sept. 2023. DOI: 10.21275/sr230903224924.

[15] Geetesh Sanodia, "*Enhancing Salesforce CRM with Artificial Intelligence*", International Journal of Artificial Intelligence Research and Development (IJAIRD), 1(1), 2023, pp. 52-61.

[16] Shrikaa Jadiga, *"Big Data Engineering Using Hadoop and Cloud (GCP/AZURE) Technologies," International Journal of Computer Trends and Technology*, vol. 72, no. 8, pp.60-69, 2024., [Link]

[17] Apr 28, 2023 Machine Learning (ML) Artificial Intelligence (AI): Business Rules Management Systems (BRMS): Data Analytics: Information Systems

[18] Kumar Shukla, Nimeshkumar Patel, Hirenkumar Mistry, 2024." *Securing The Cloud: Strategies and Innovations In Network Security For Modern Computing Environments"* Volume 11, Issue 04 pp. 1786-1796. [Link]

[19] DOCTOR A., VONDENBUSCH B., KOZAK J.*, Bone segmentation applying rigid bone position and triple shadow check method based on RF data,* Acta of Bioengineering and Biomechanics, 2011, Vol. 13, 3–11.[LINK]

[20] Dixit, A.S., Nagula, K.N., Patwardhan, A.V. and Pandit, A.B., 2020. Alternative and remunerative solid culture media for pigment-producing serratiamarcescens NCIM 5246. *J Text Assoc*, *81*(2), pp.99-103.

[21] Vishwanath Gojanur , Aparna Bhat, "Wireless Personal Health Monitoring System", IJETCAS: International Journal of Emerging Technologies in Computational and Applied Sciences,eISSN: 2279-0055,pISSN: 2279-0047, 2014. [Link]

[22] Apurva Kumar, "Building Autonomous AI Agents based AI Infrastructure," International Journal of Computer Trends and Technology, vol. 72, no. 11, pp. 116-125, 2024. Crossref, https://doi.org/10.14445/22312803/IJCTT-V72I11P112

[23] Chandrakanth Lekkala (2023) Deploying and Managing Containerized Data Workloads on Amazon EKS. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-342. DOI: doi.org/10.47363/JAICC/2023 (2)324.

[24] S. Duary, P. Choudhury, S. Mishra, V. Sharma, D. D. Rao and A. Paul Aderemi, "Cybersecurity 0054hreats Detection in Intelligent Networks using Predictive Analytics Approaches," *2024 4th International Conference on Innovative Practices in Technology and Management (ICIPTM)*, Noida, India, 2024, pp. 1-5, doi: 10.1109/ICIPTM59628.2024.10563348.

[25] S. Kumar, R. S. M. Joshitta, D. D. Rao, Harinakshi, S. Masarath and V. N. Waghmare, "Storage Matched Systems for Single-Click Photo Recognition Using CNN," *2023 International Conference on Communication, Security and Artificial Intelligence (ICCSAI)*, Greater Noida, India, 2023, pp. 1-7, doi: 10.1109/ICCSAI59793.2023.10420912.

[26] Rao, Deepak, and Sourabh Sharma. "Secure and Ethical Innovations: Patenting Ai Models for Precision Medicine, Personalized Treatment, and Drug Discovery in Healthcare." International Journal of Business Management and Visuals, ISSN: 3006-2705 6.2 (2023): 1-8.

[27] Addimulam, S., Mohammed, M. A., Karanam, R. K., Ying, D., Pydipalli, R., Patel, B., ... & Natakam, V. M. (2020). Deep Learning-Enhanced Image Segmentation for Medical Diagnostics. Malaysian Journal of Medical and Biological Research, 7(2), 145-152. [Link]

[28] Karthik Hosavaranchi Puttaraju, "Harnessing Disruptive Technologies: Strategic Approach to Retail Product Innovation", International Journal of Scientific Research in Engineering and Management (IJSREM), VOLUME: 08 ISSUE: 01 | JAN - 2024.

[29] Tsaliki KC. AI-driven hormonal profiling: a game-changer in polycystic ovary syndrome prevention. Int J Res Appl Sci Eng Technol (IJRASET). 2024. https://doi.org/10.22214/ijraset.2024.61001.

[30] Naga Ramesh Palakurti, 2023. AI-Driven Personal Health Monitoring Devices: Trends and Future Directions, ESP Journal of Engineering & Technology Advancements, 3(3): 41-51.

[31] Chandrakanth Lekkala 2022. "Integration of Real-Time Data Streaming Technologies in Hybrid Cloud Environments: Kafka, Spark, and Kubernetes", European Journal of Advances in Engineering and Technology, 2022, 9(10):38-43. [Link]