International Journal of
INTELLIGENT SYSTEMS AND APPLICATIONS IN ENGINEERING
ISSN:2147-6799 www.ijisae.org Original Research Paper

Deep Learning Approaches for Medical Images Segmentation: A Systematic Review

Taryn Michael[1], Ibidun Christiana Obagbuwa*[2]

Submitted: 17/05/2024 **Revised:** 27/06/2024 **Accepted:** 11/07/2024

Abstract*:* Medical image segmentation is a critical component of computer-aided diagnosis and treatment planning, with advancements driven by continuous research in deep learning architectures. This systematic review explores state-of-the-art models for medical image segmentation, focusing on recent developments and innovations. The review covers prominent architectures, such as UNet++, nnU-Net, HRNet, Vision Transformer (ViT), and DUCK-Net, each contributing to improved segmentation accuracy and efficiency. Additionally, it discusses traditional and deep learning-based approaches, highlighting the effectiveness of convolutional neural networks (CNNs) and fully convolutional networks (FCNs). The integration of uncertainty quantification methodologies, particularly Bayesian neural networks (BNNs), is examined to enhance interpretability and reliability in medical imaging models. Evaluation metrics, including F- measure-based metrics, sensitivity, specificity, and the impact of class imbalance, are thoroughly analyzed to ensure robust algorithmic assessment. The review emphasizes the significance of metrics like the Dice Similarity Coefficient (DSC) and Intersection-over-Union (IoU) in addressing challenges posed by class imbalance in medical image segmentation. The comprehensive synthesis aims to provide a detailed overview of the current landscape, aiding researchers and practitioners in navigating the evolving field of medical image segmentation. The systematic search and selection methodology ensures a rigorous examination of relevant literature, contributing to a comprehensive understanding of the advancements in this critical domain.

Keywords: *Medical Image Segmentation, Medical Image Analysis, Computer-Aided Diagnosis and Treatment, Deep Learning Architectures, Uncertainty Quantification, Evaluation Metrics, State-Of-The-Art Architecture, Systematic Review*

1. Introduction

Medical image analysis, explicitly identifying structures from medical images through computational techniques, constitutes a vast and evolving field. Among the myriad techniques, image segmentation is a pivotal method for delineating and extracting regions of interest within complex biomedical data.

Image segmentation, a process of partitioning an image into distinct segments, is pivotal in extracting regions of interest that are crucial for diagnostic and treatment planning. By assigning labels to pixels within the image, segmentation aids in identifying and understanding the elements encapsulated in the medical images. However, despite the extensive study and development of segmentation techniques, achieving accurate and reliable results remains a formidable challenge, primarily due to the inherent variations in the shape and size of an individual's anatomy, as highlighted by [1].

The advent of deep neural networks has ushered in a new era in medical image segmentation, offering state-of-the-art results. However, these networks exhibit limitations—relying solely on deterministic estimates, resulting in unreliable predictions, a lack of interpretability, and a dependency on extensive datasets and regularization to prevent overfitting, as outlined by [2]. The reliability of predictions is particularly critical in the medical field, where inaccuracies can have life-threatening consequences.

This systematic literature review delves into recent advancements in state-of-the-art architectures for medical image segmentation, exploring innovative models that extend beyond the traditional U-Net framework. Researchers have been committed to refining segmentation accuracy, addressing challenges in volumetric medical image segmentation, and incorporating high-resolution features to enhance localization accuracy. The landscape has witnessed the emergence of novel architectures, such as UNet++, nnU-Net, HRNet, Vision Transformer (ViT), and DUCK-Net, each contributing distinct innovations to the field.

Furthermore, the review navigates the transformative landscape of deep learning approaches for medical image segmentation, differentiating traditional methods like active contours from paradigm-shifting convolutional

1,2Department of Computer Science & Information Technology,

Sol Plaatjie University, South Africa

Email: ibidun.obagbuwa@spu.ac.za

neural networks (CNNs) and fully convolutional networks (FCNs). Examples such as U-Net for liver segmentation and DeepLab for brain tumor segmentation exemplify the prowess of deep learning in achieving superior segmentation results, surpassing traditional methodologies.

In parallel, medical imaging models have an escalating emphasis on uncertainty quantification. Researchers are actively developing models that deliver accurate segmentation, provide insights into the reasoning behind segmentation decisions, and quantify uncertainty levels. Bayesian Neural Networks (BNNs) have emerged as a promising avenue, offering robust uncertainty estimates and contributing to improved model performance in medical image analysis.

Robust evaluation metrics are imperative to ensure the efficacy of these methodologies. The review scrutinizes various evaluation metrics, emphasizing the challenges of class imbalances in medical images. F-measure-based metrics, sensitivity and specificity, and the caution against using accuracy due to class imbalance are discussed, focusing on the suitability of metrics like Dice Similarity Coefficient (DSC) and Intersection-over-Union (IoU) in medical image segmentation.

By synthesizing insights from state-of-the-art architectures, traditional and deep learning approaches, uncertainty quantification, and evaluation metrics, this systematic literature review aims to contribute to the ongoing refinement of techniques in medical image segmentation, with the ultimate goal of improving diagnostic methods and ensuring the reliability of predictions in critical medical applications.

The subsequent sections are structured as follows:

Materials and Methods details the search and selection methods employed to curate relevant articles.

Cutting-edge Architectures for Medical Image Segmentation, which provides an overview of cutting-edge architectures, including UNet++, nnU-Net, HRNet, Vision Transformer (ViT), and DUCK-Net, each contributing unique innovations to enhance segmentation accuracy and efficiency.

Revolutionary Deep Learning Approaches for Medical Image Segmentation, which provides a comprehensive exploration of traditional and deep learning-based approaches, showcasing the evolution from methods like active contours to the revolutionary impact of convolutional neural networks (CNNs) and fully convolutional networks (FCNs).

Quantifying Uncertainty in Medical Imaging Models, which places a growing emphasis on interpretability and uncertainty quantification, with a focus on Bayesian inference as an alternative approach to address the limitations of deterministic estimates in deep neural networks.

Evaluation Metrics for Medical Image Segmentation: An in-depth analysis of evaluation metrics, emphasizing the necessity for metrics like Dice Similarity Coefficient (DSC) and Intersection-over-Union (IoU) that provide a precise representation in the context of class imbalances inherent in medical images.

Finally, the conclusion summarizes and concludes this study.

Materials And Methods

The systematic methodology employed in this review aligns with established guidelines outlined by [3] and [10]. To achieve the survey's objectives, distinct research questions were formulated. A structured protocol was rigorously adhered to, ensuring a meticulous and comprehensive approach to identifying pertinent scientific literature.

This methodological framework encompassed the following key components:

1. Definition of Research Questions and Search Queries: Pertinent search queries were precisely formulated, aligned with the research questions, and applied systematically across appropriate research portals. This step ensured a thorough exploration of the existing scientific literature.

2. Inclusion and Exclusion Criteria: Well-defined criteria were established to guide the selection of studies. This included clear parameters for determining which studies were relevant and should be included, as well as those that did not meet the specified criteria for exclusion. This meticulous approach aimed to enhance the quality and relevance of the identified research.

3. Study Selection Method: A methodical process was employed for selecting studies, encompassing the extraction of pertinent information from each selected study. This step facilitated retrieving valuable insights and data necessary for the subsequent analysis.

4. Analysis of Selected Studies: The selected studies underwent a comprehensive analysis, ensuring a robust evaluation of their methodologies, findings, and contributions. This systematic approach allowed for a nuanced understanding of the existing body of literature.

By adhering to this systematic protocol, the review aimed to provide a rigorous, structured, and comprehensive synthesis of the relevant scientific literature, offering valuable insights into the landscape of the chosen research domain.

Research Questions:

The initial phase of the systematic review process involved formulating specific research questions tailored to the study's objectives. In accordance with the study's overarching aim, the following research questions were defined:

1. What is some current state-of-the-art architecture for medical image segmentation?
2. How have deep learning approaches for medical image segmentation evolved from traditional approaches?
3. How have uncertainty quantifications been utilized in medical image segmentation models to address their limitations?
4. How is the performance of medical image segmentation models evaluated?

Search Queries, Analysis and Study Selection:

The repositories searched include Google Scholar, papers with code, Science Direct, and Springer. The search queries were as follows:

"Evaluation metrics for medical image segmentation"
"Uncertainty quantification medical image segmentation"
"Bayesian inference medical image segmentation"
"Medical image segmentation"
"State of the art medical image segmentation"
"Deep learning for medical image segmentation"

The preliminary search yielded over 2000 research articles. Subsequently, a thorough evaluation was conducted based on the titles and a brief overview of their abstracts. Through this examination, only articles able to answer the research questions were selected for further consideration. Finally, 20 articles were then selected for the study.

Inclusion and Exclusion Criteria:

The criteria for inclusion and exclusion were meticulously selected to maintain the highest level of formality in this paper. Adhering to these criteria, papers aligned with the specified research perspective criteria were considered for inclusion in the research scope. Table 1 shows the inclusion and exclusion criteria.

Table 1: Inclusion and Exclusion Criteria

Inclusion Criteria	Exclusion Criteria
Peer-reviewed journals and articles.Written in English. No older than 2017. The study addresses a particular dimension of the research questions.	Written in a language other than English.Older than 2017. Does not answer any of the research questions.

Cutting-Edge Architectures For Edical Image Segmentation

In computer-aided diagnosis and treatment planning, the pivotal task of medical image segmentation has witnessed continuous advancements in accuracy and efficiency. Researchers have consistently pushed the boundaries by developing sophisticated deep learning architectures. Predominantly, these architectures draw inspiration from the U-net's encoder- decoder structure, a widely adopted model for medical image segmentation. Presented below is an overview of recent state-of-the-art architectures in this domain.

Addressing the intricacies inherent in biomedical imaging, [5] introduced UNet++, an augmentation of the original U-Net architecture. This innovation incorporates nested skip connections, enhancing feature aggregation across different scales. Proven crucial in capturing nuanced structures, this enhancement elevates segmentation performance, marking the continuous refinement of U-Net-inspired architectures.

Recognizing challenges in volumetric medical image segmentation, [6] proposed nnU-Net, a modular 3D U-Net architecture demonstrating adaptability to volumetric intricacies. Noteworthy performance across various segmentation tasks highlights its efficiency in processing volumetric information.

Yu et al. contributed to HRNet, which is designed for high-resolution representation learning. HRNet's unique feature lies in its ability to maintain detailed information throughout the network, improving localization accuracy [7]. Demonstrating state-of-the-art performance in medical imaging and diverse computer vision tasks underscores the significance of preserving high-resolution features in medical image segmentation.

Dosovitskiy et al. presented the Vision Transformer (ViT), originally designed for image classification but gaining attention in medical image segmentation. Leveraging self - attention mechanisms, ViT shows promise in various visual recognition tasks, indicating a paradigm shift in architectural choices for medical image segmentation [8].

In a recent contribution, Dumitru et al. introduced DUCK-Net, a novel supervised convolutional neural network architecture designed for effective learning and generalization from limited medical images. DUCK-Net's architecture incorporates an encoder-decoder structure with a residual downsampling mechanism and a custom convolutional block. This design facilitates capturing and processing image information at multiple resolutions in the encoder segment, contributing to its versatility and applicability across various segmentation tasks. DUCK-Net achieves state-of-the-art results on popular benchmark datasets, showcasing its performance across key segmentation metrics [9].

Revolutionary Deep Learning Approached For Medical Image Segmentation:

Traditional approaches to medical image segmentation, including thresholding, region growing, active contours, and graph-based methods, have faced challenges in handling complex images with heterogeneous structures and noise. Recent strides in medical image segmentation owe much to deep learning-based approaches, categorized into convolutional neural networks (CNNs) and fully convolutional networks (FCNs).

One exemplary CNN-based approach is the utilization of U-Net for liver segmentation in CT images, as demonstrated by [10]. U-Net's contracting path captures contextual information, while its symmetric expanding path localizes the object, outperforming traditional methods and achieving high accuracy in liver segmentation.

FCN-based approaches, exemplified by DeepLab in brain tumor segmentation [11], replace fully connected layers with convolutional layers, allowing end-to-end learning of the segmentation mask. DeepLab, utilizing atrous convolution to capture multi-scale contextual information, achieves state-of-the-art results in brain tumor segmentation.

Combining traditional and deep learning-based methods, hybrid approaches have gained popularity for achieving more accurate and robust segmentation results. [12] employed convolutional neural networks and graph-based methods for liver segmentation in CT images, demonstrating that the hybrid approach outperformed individual methods.

Quantifying Uncertainty In Medical Imaging Models:

Recent emphasis on the interpretability of models and the quantification of uncertainty in the medical field has led to the development of methods providing insights into a model's segmentation decisions and confidence levels. [13]. The generative Bayesian Deep Learning (GBDL) architecture was introduced, falling under generative models that estimate the joint distribution of medical volumes and their corresponding labels. This architecture, leveraging data distribution and accommodating labelled and unlabelled data, surpasses other contemporary state-of-the-art models.

Kendall et al. and Qin et al. proposed Bayesian versions of SegNet and U-Net architectures for medical image segmentation, introducing Bayesian convolutional encoder-decoder architectures that estimate model uncertainty. Both approaches perform comparably to traditional architectures while significantly reducing uncertainty in segmentation results [14,15].

In contrast, Kohl et al. proposed a Bayesian neural network (BNN) approach that considers epistemic and aleatoric uncertainties for medical image segmentation [16]. Galdran et al. proposed a 3D BNN approach for brain tumor segmentation in MRI [17], while Ghafoorian et al. designed a BNN approach specifically for small data [18]. All three studies demonstrated comparable performance to traditional methods, with [16] showcasing accurate segmentation across multiple medical image datasets.

Evaluation Metrics For Medical Image Segmentation:

Accurate evaluation metrics are imperative for the success of medical image segmentation in diverse clinical applications. The primary goal is to measure the similarity between the predicted and annotated segments (ground truth). Over the past three decades, the Medical Image Segmentation (MIS) field has introduced a diverse range of evaluation metrics. Still, only a select few have demonstrated appropriateness and are consistently utilized in a standardized manner.

F-Measure-Based Metrics:

The F-measure, also referred to as the F-score, stands out as a widely employed metric in computer vision and MIS scientific research. This metric, encompassing sensitivity and precision, quantifies the intersection between predicted segmentation and ground truth. Notably, it addresses the challenge of class-imbalanced datasets in MIS by penalizing false positives. The Dice similarity coefficient (DSC) and Intersection-over-Union (IoU) are popular metrics derived from F-measure, with DSC, introduced by [19], becoming a cornerstone metric known for its simplicity and effectiveness in handling class imbalances.

Sensitivity and Specificity:

In the medical domain, sensitivity (recall) and specificity serve as established standards for performance evaluation. Sensitivity focuses on true positive detection, while specificity evaluates the correct identification of true negative classes like the background. Although sensitivity is a widely used metric in MIS, it may not be as sensitive as F-score-based metrics for accurate evaluation. Specificity, gauging the model's ability to identify the background class, is essential for ensuring the model's functionality. However, its high values may be less indicative of the overall model performance [20].

Accuracy:

Accuracy, also known as pixel accuracy, is a popular statistic metric that calculates the number of correct predictions relative to total predictions. However, using accuracy in MIS is discouraged due to class imbalance. The inclusion of true negatives in the accuracy metric can lead to misleadingly high scores, even when predicting the entire image as the background class [21] [22]. Therefore, accuracy is considered unsuitable for evaluating MIS models in scientific assessments.

Effect of Class Imbalance on Evaluation Metrics:

With their inherent class imbalance, medical images pose a significant challenge for image segmentation tasks. Metrics like Accuracy or Specificity, which equally weigh true positives and true negatives, tend to yield high scores even when any pixel is classified as the Region of Interest (ROI). This biases the interpretation and makes these metrics unsuitable for evaluating segmentation performance in MIS. Metrics concentrating exclusively on true positive classification, excluding true negatives, offer a more precise representation in the medical domain. Hence, metrics like Dice Similarity Coefficient (DSC) and Intersection-over-Union (IoU) are highly favored and recommended in the field of MIS [20].

Conclusions

In conclusion, medical image analysis, particularly segmenting structures from medical images through computational techniques, represents a dynamic and expanding field. Image segmentation, a pivotal method in this domain, is crucial in extracting regions of interest essential for diagnostic and treatment planning. Challenges persist due to inherent anatomical variations despite the advancements in segmentation techniques.

The advent of deep neural networks has brought about a paradigm shift, offering state-of-the- art results in medical image segmentation. However, limitations such as deterministic estimates, lack of interpretability, and dependence on extensive datasets have been acknowledged. In the medical field, where reliability is paramount, prediction inaccuracies can have severe consequences.

This systematic literature review explores recent advancements in state-of-the-art architectures for medical image segmentation, extending beyond traditional frameworks like U-Net. Researchers have dedicated efforts to enhance segmentation accuracy, address challenges in volumetric segmentation, and incorporate high-resolution features for improved localization accuracy. Novel architectures, including UNet++, nnU-Net, HRNet, Vision Transformer (ViT), and DUCK-Net, have emerged with distinct innovations contributing to the field.

The review also navigates the transformative landscape of deep learning approaches for medical image segmentation, highlighting the evolution from traditional methods like active contours to the revolutionary impact of convolutional neural networks (CNNs) and fully convolutional networks (FCNs). Notable examples, such as U-Net for liver segmentation and DeepLab for brain tumor segmentation, underscore the efficacy of deep learning in achieving superior segmentation results.

Uncertainty quantification in medical imaging models has gained prominence, with researchers actively developing models that provide accurate segmentation and insights into the reasoning behind segmentation decisions and quantify uncertainty levels. Bayesian Neural Networks (BNNs) have emerged as a promising avenue, offering robust uncertainty estimates and contributing to improved model performance.

Robust evaluation metrics are imperative for the efficacy of these methodologies. The review scrutinizes various evaluation metrics, emphasizing the challenges of class imbalances in medical images. F-measure-based metrics, sensitivity and specificity, and the caution against using accuracy due to class imbalance are discussed, focusing on the suitability of metrics like Dice Similarity Coefficient (DSC) and Intersection-over-Union (IoU) in medical image segmentation.

By synthesizing insights from state-of-the-art architectures, traditional and deep learning approaches, uncertainty quantification, and evaluation metrics, this systematic literature review aims to contribute to the ongoing refinement of techniques in medical image segmentation. The ultimate goal is to improve diagnostic methods and ensure the reliability of predictions in critical medical applications.

Conflicts of Interest

The authors declare that there is no conflict of interest regarding the publication of this paper.

Funding Statement

This work is funded by the DSCI-NICIS National e-science Postgraduate Teaching and Training Platform (NEPTTP).

References:

[1] Han, Y., Huang, C., Jiang, S., & Chen, X. (2021). A survey on medical image segmentation using deep learning. Neural Computing and Applications, 33(14), 7571 -7584. https://doi.org/10.1007/s00521-021- 06232-1 Havaei, M., et al. (2017). Brain tumour segmentation with Deep Neural Networks. Medical Image Analysis, 35, 18-31.

[2] Gao, C., Yang, Q., Zhao, Y., & Ma, X. (2020). A Deep Learning Framework with Bayesian Optimization for Medical Image Segmentation. IEEE Access, 8, 87140 -87149. https://doi.org/10.1109/access.2020.2990415

[3] Liberati A, Altman DG, Tetzlaff J, Mulrow C, Gøtzsche PC,Ioannidis JP, Clarke M, Devereaux PJ, Kleijnen J, Moher D (2009) The PRISMA statement for reporting systematic reviews and meta-analyses of studies that evaluate health care interventions: explanation and elaboration. J Clin Epidemiol 62(10):e1–e34

[4] Kitchenham B, Charters S (2007) Guidelines for performing systematic literature reviews in software engineering.

[5] Zhou, H., Xu, Y., & Hu, X. (2020). Automated whole-heart segmentation in cardiac MRI using a Bayesian U-Net deep learning network. Computerized Medical Imaging and Graphics, 80.

[6] Isensee, F., Petersen, J., Klein, A., Zimmerer, D., Jaeger, P. F., Kohl, S., ... & Maier-Hein, K. H. (2018). nnu-net: Self-adapting framework for u-net-based medical image segmentation. arXiv preprint arXiv:1809.10486.

[7] Yu, C., Xiao, B., Gao, C., Yuan, L., Zhang, L., Sang, N., & Wang, J. (2021). Lite-hrnet: A lightweight high-resolution network. In Proceedings of the IEEE/CVF conference on computer vision and pattern recognition (pp. 10440-10450).

[8] Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., ... & Houlsby, N. (2020). An image is worth 16x16 words: Transformers for image recognition at scale. arXiv preprint arXiv:2010.11929.

[9] Dumitru, R. G., Peteleaza, D., & Craciun, C. (2023). Using DUCK-Net for polyp image segmentation. Scientific Reports, 13(1), 9803.

[10] Roth, H. R., et al. (2016). Deep learning for liver segmentation in CT scans: a survey of past and present techniques. IEEE Transactions on Medical Imaging, 37(6), 1564 -1573.

[11] Havaei, M., et al. (2017). Brain tumour segmentation with Deep Neural Networks. Medical Image Analysis, 35,18-31.

[12] Christ, P. F., Elshaer, M. E. A., Ettlinger, F., Tatavarty, S., Bickel, M., Bilic, P., ... & Menze, B. H. (2016). Automatic liver and lesion segmentation in CT using cascaded fully convolutional neural networks and 3D conditional random fields. In Medical Image Computing and Computer-Assisted Intervention–MICCAI 2016: 19th International Conference, Athens, Greece, October 17 -21, 2016, Proceedings, Part II 19 (pp. 415-423). Springer International Publishing.

[13] Wang, J. and Lukasiewicz, T., 2022. Rethinking bayesian deep learning methods for semi-supervised volumetric medical image segmentation. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition 182-190.

[14] Kendall, A., Gal, Y., & Cipolla, R. (2017). Multi-task learning using uncertainty to weigh losses for scene geometry and semantics. In Proceedings of the IEEE conference on computer vision and pattern recognition (pp. 7482-7491).

[15] Qin, C., Schlemper, J., Caballero, J., Price, A. N., Hajnal, J. V., & Rueckert, D. (2019). Convolutional recurrent neural networks for dynamic MR image reconstruction. IEEE transactions on medical imaging, 38(1), 280-290.

[16] Kohl, S. A., Romera-Paredes, B., Meyer, C., De Fauw, J., Ledsam, J. R., Maier-Hein, K. H., ... & Ronneberger, O. (2018). A probabilistic U-Net for segmentation of ambiguous images. In Advances in neural information processing systems (pp. 6965-6975).

[17] Galdran, A., Gomez, L., Besga, A., Sanchez, F. J., & Campoy, P. (2019). Bayesian 3D fully convolutional networks for brain tumour segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention (pp.738-746). Springer, Cham.

[18] Ghafoorian, M., Mehrtash, A., Kapur, T., Karssemeijer, N., Marchiori, E., Pesteie, M., ... &

Abolmaesumi, P. (2019). Federated learning of MRI contrasts without sharing patient data: A proof of concept study for brain tumour segmentation. arXiv preprint arXiv:1912.12426.

[19] Dice, L.R. (1945). Measures of the Amount of Ecologic Association Between Species. Ecology, 26, 297-302.

[20] Liu X, Song L, Liu S, Zhang Y (2021). A review of deep-learning-based medical image segmentation methods. Sustain. 13(3):1–29.

[21] Popovic A, de la Fuente M, Engelhardt M, Radermacher K (2007). Statistical validation metric for accuracy assessment in medical image segmentation. Int J Comput Assist Radiol Surg. 2(3–4):169–81. https://doi.org/10.1007/s11548-007-0125-1.

[22] Taha AA, Hanbury A (2015). Metrics for evaluating 3D medical image segmentation: analysis, selection, and tool. BMC Med Imaging. 15(1):29.