Computer Engineering and Applications Vol. 13, No. 3, October 2024

MRI-Based Brain Tumor Instance Segmentation Using Mask R-CNN

Muhammad Nasrudin

Department of Data Science, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur
**nasrudin.fasilkom@upnjatim.ac.id.*

ABSTRACT

Brain tumor segmentation is a crucial step in medical image analysis for the accurate diagnosis and treatment of patients. Traditional methods for tumor segmentation often require extensive manual effort and are prone to variability. In this study, we propose an automated approach for brain tumor segmentation using Mask R-CNN, a state-of-the-art deep learning model for instance segmentation. Our method leverages MRI images to identify and delineate brain tumors with high precision. We trained the Mask R-CNN model on a dataset of annotated MRI images and evaluated its performance using the mean Average Precision (mAP) metric. The results demonstrate that our model achieves a high mAP of 90.3%, indicating its effectiveness in accurately segmenting brain tumors. This automated approach not only reduces the manual effort required for tumor segmentation but also provides consistent and reliable results, potentially improving clinical outcomes.

Keywords: Brain Tumor, Instance segmentation, Mask R-CNN

1. INTRODUCTION

A brain tumor ranks as the 12th most lethal disease globally among all cancer types. The World Health Organization (WHO) reported 321,731 cases with a mortality rate of 77% in 2020 [1]. Consequently, brain tumors are deemed a critical health concern. Brain tumor detection can be performed using a medical device known as Magnetic Resonance Imaging (MRI). The segmentation of brain tumors in MRI scans is a critical task in medical image analysis. Accurate segmentation is essential for diagnosis, treatment planning, and monitoring of brain tumors. Traditional methods for brain tumor segmentation often involve manual delineation by radiologists, which is time-consuming and subject to inter-observer variability. With the advent of advanced machine learning techniques, automated segmentation has become a viable and increasingly accurate alternative [2], [3].

Recent advancements in deep learning have significantly improved the accuracy and efficiency of medical image analysis [4], [5]. Convolutional Neural Networks (CNNs), in particular, have demonstrated remarkable performance in various image processing tasks, including segmentation [6]-[8]. Among these, the Mask R-CNN (Region-based Convolutional Neural Network) has emerged as a highly effective architecture for object detection and segmentation. Originally designed for general object segmentation tasks, Mask R-CNN has been successfully adapted for medical imaging applications, including brain tumor segmentation in MRI scans [9].

Mask R-CNN extends the Faster R-CNN architecture by adding a branch for predicting segmentation masks on each region of interest (RoI). This makes it particularly well-suited for tasks that require precise delineation of object boundaries, such as tumor segmentation. The architecture consists of a backbone network for feature extraction, typically a ResNet or ResNeXt, followed by a Region Proposal Network (RPN) that generates candidate object proposals. These proposals are refined, classified, and pixel-wise masks are generated for each detected object, enabling precise segmentation [10].

Applying Mask R-CNN to brain tumor segmentation in MRI images involves several challenges, including the heterogeneity of tumor appearance, varying sizes and shapes of tumors, and the presence of similar-looking healthy tissues. Despite these challenges, Mask R-CNN's adaptability and robustness make it a promising approach for this task. Recent studies have shown that Mask R-CNN can achieve high segmentation accuracy, which is critical for effective clinical decision-making.

This research aims to evaluate the effectiveness of Mask R-CNN for MRI-based brain tumor segmentation, focusing on improving segmentation accuracy and robustness. By leveraging the capabilities of Mask R-CNN and incorporating domain-specific adaptations, this study seeks to advance the state-of-the-art in automated brain tumor segmentation.

2. MATERIAL AND METHODS

2.1 DATASET

The data used in this study comes from a Kaggle dataset, which includes 253 MRI brain images with a resolution of 1024×1024 pixels. The data is categorized into two classes, tumor and non-tumor. The data is divided into training, validation, and testing sets with a ratio of 70:20:10. The amount of data for each set is shown in Figure 1.

FIGURE 1. Division of the Data

2.2 MASK R-CNN

Mask R-CNN is a deep learning model that integrates object detection and instance segmentation, building upon the Faster R-CNN architecture. The key innovation of Mask R-CNN is its ability to perform pixel-wise instance segmentation in addition to object detection. This is accomplished through the addition of a "mask head" branch, which produces precise segmentation masks for each detected object, enabling detailed and accurate pixel-level boundaries.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

Two critical enhancements in Mask R-CNN are ROIAlign and the Feature Pyramid Network (FPN). ROIAlign addresses the limitations of traditional ROI pooling by using bilinear interpolation during the pooling process, mitigating misalignment issues and ensuring accurate spatial information capture from the input feature map, which improves segmentation accuracy, especially for small objects.

FPN is crucial for feature extraction, constructing a multi-scale feature pyramid that incorporates features from different scales. This allows the model to understand object context more comprehensively and enhances object detection and segmentation across a wide range of object sizes [6].

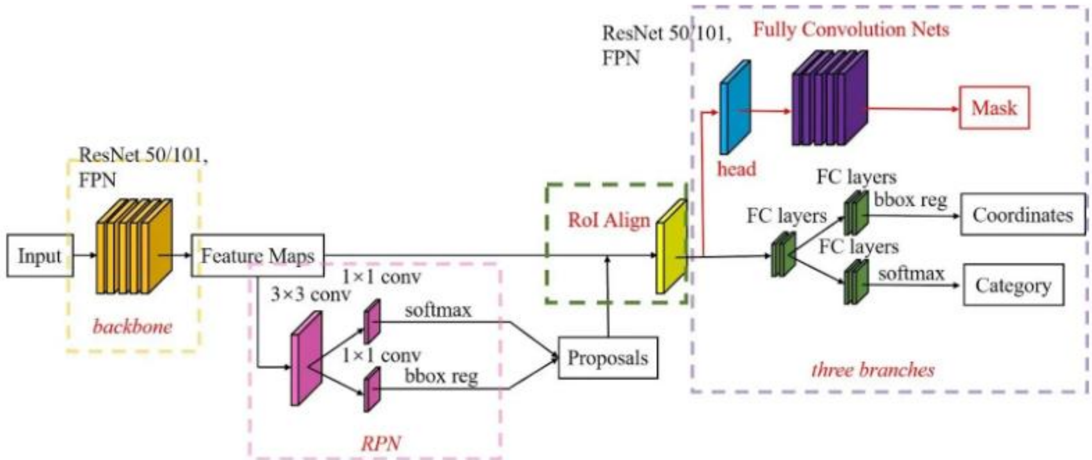

FIGURE 2. Mask R-CNN Framework for Instance Segmentation

However, in the second stage, alongside predicting the class and box offset, Mask R-CNN also produces a binary mask for each RoI. During training, a multi-task loss is applied to each sampled RoI, defined as $L = L_{cls} + L_{box} + L_{mask}$, where L_{cls} represents the classification loss, L_{box} represents the bounding-box loss, and L_{mask} represents the average binary cross-entropy loss.

The architecture of Mask R-CNN is divided into two parts: first, the convolutional backbone architecture used for feature extraction over the entire image, and second, the network head for bounding-box recognition (classification and regression) and mask prediction, which is applied separately to each RoI.

In Mask R-CNN, the backbone network is typically a pre-trained convolutional neural network, such as ResNet or ResNeXt, which processes the input image to extract high-level features. An FPN is then added to the top of this backbone network to construct a feature pyramid. FPNs are designed to handle objects of different sizes and scales within an image. This architecture forms a multi-scale feature pyramid by merging features from various levels of the backbone network, incorporating features with different spatial resolutions. It includes high-resolution features rich in semantic information and low-resolution features with finer spatial details. The feature pyramid generated by FPN enables Mask R-CNN to handle objects of various sizes effectively. This multi-scale representation allows the model to capture contextual information and accurately detect objects at different scales within the image.

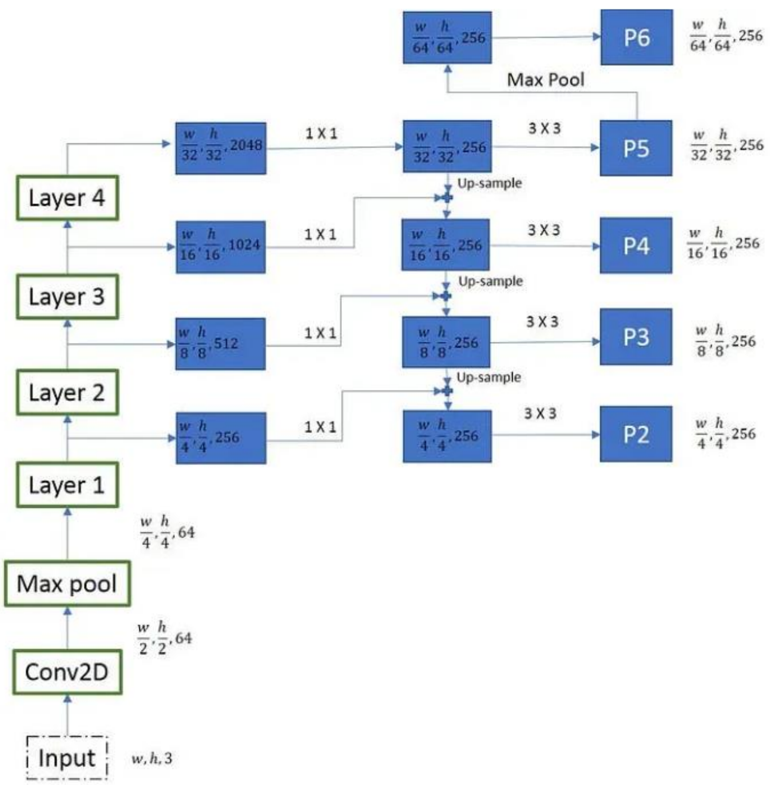

FIGURE 3. Feature Pyramid Networks (FPN) Backbone

The choice of backbone architecture can vary based on specific models, although [10] suggest that using the ResNet-FPN backbone for feature extraction with Mask R-CNN yields significant improvements in both accuracy and speed. The heads for Faster R-CNN with a ResNet backbone are shown in Figure 4.

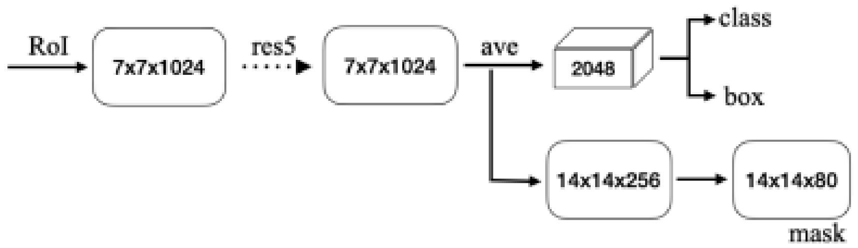

FIGURE 4. Mask R-CNN Head Architecture: Faster R-CNN with ResNet Backbone

2.3 MODEL EVALUATION

Performance evaluation via mean average precision (mAP) determines how well the model is. The bounding box results, from an annotation data detection system, and ground truth that is annotated by researchers with the doctor validation [11], are then processed in the confusion matrix as in Table 1. The data is grouped as True Positive (TP), False Positive (FP), True Negative (TN), and False Negative (FN), which has

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

been defined as Interest of Union (IoU). Performance Parameters Calculation Flow can be illustrated in Figure 5.

TABLE 1.
Confusion Matrix

		Actual	
		Positive	Negative
Predict	Positive	TP	FP
	Negative	FN	TN

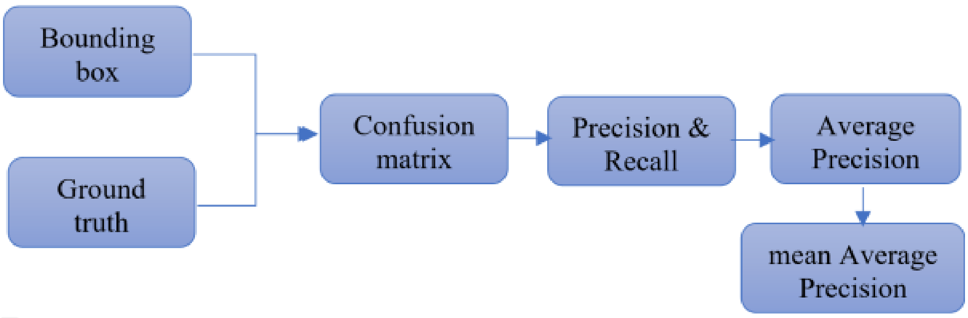

FIGURE 5. Performance Parameters Calculation Flow

As illustrated in Figure 5, various evaluation metrics can be derived from the values in the confusion matrix. The formulas for Precision and Recall are provided in equations (1) and (2), respectively.

$$Precision = \frac{TP}{TP+FP} \quad (1)$$

$$Recall = \frac{TP}{TP+FN} \quad (2)$$

Precision indicates the model's capability to correctly identify relevant objects, expressed as a percentage of correct predictions. Recall measures the model's ability to detect all relevant objects, represented as the percentage of true positives found across all ground truths. The mean Average Precision (mAP) is the average of the Average Precision (AP) values, serving as the performance metric for object detection. The AP value is derived from the Precision calculated using Equation (1) and the Recall using Equation (2), subsequently computed as outlined in Equation (3).

$$\mathrm{AP} = \sum_{n=0} (r_{n+1} - r_n)\, p_{interp}(r_{n+1})$$
$$p_{interp}(r_{n+1}) = max\ p(\tilde{r}) \quad (3)$$

where p is Precision, p_{interp} is Precision interpolation, while r is Recall and $p(\tilde{r})$ is Precision calculated on the Recall. AP is the area under curve of Recall and Precision. The curve is sampled at all unique recall values ($r_1, r_2, \ldots, r_n$) so r_n and r_{n+1} are according to the Recall values [12].

3. RESULT AND DISCUSSION

The hyperparameter setting in this study is shown in Table 2. Backbone network used Resnet50 and image size for input resize 256×256 pixel.

TABLE 2.
Hyperparameter of Study

Model	Hyperparameter	Value
Resnet50	Image size	256×256
	Batch size	1
	Training step	200
	Learning rate	0.001

The total number of images trained to obtain one model is 60,000 images. Using 30 epochs, the performance matrix training model is shown in Figure 6.

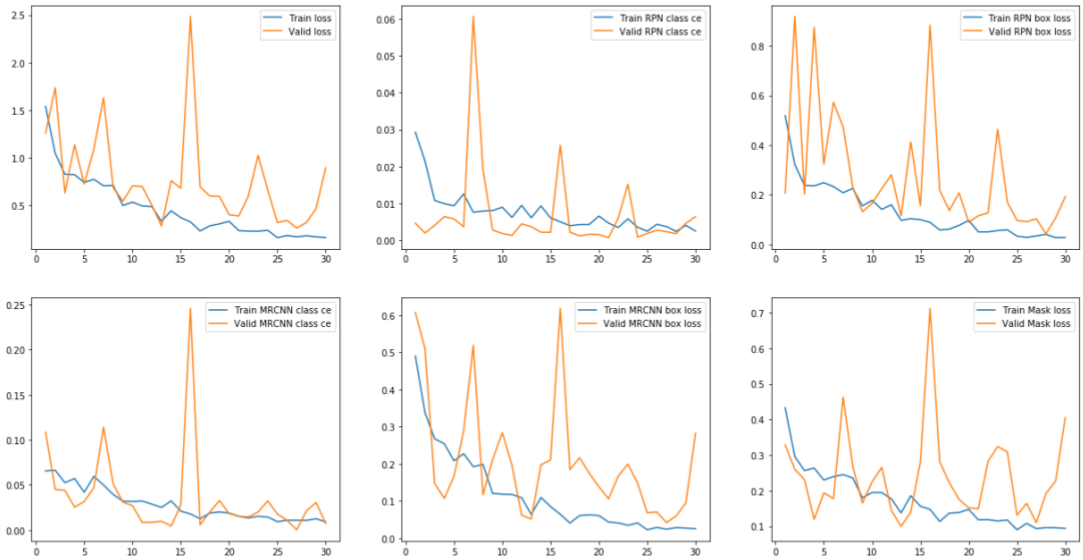

FIGURE 6. Performance Matrix Training Model

The first evaluation metric in this study is the confusion matrix, used to assess the accuracy of the classification results. From this matrix, the precision and recall values can be calculated. The components of classification goodness are detailed in the confusion matrix in Table 3. Precision and Recall values can be derived from the confusion matrix. The Precision, Recall, and mAP values for testing and training are

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

presented in Table 4. Additionally, the visual analysis of the result instance segmentation using Mask R-CNN is displayed in Figure 7.

TABLE 3.
Confusion Matrix

Training		Predicted		Testing		Predicted	
		Positive	Negative			Positive	Negative
Ground-Truth	Positive	102	13	Ground-truth	Positive	15	1
	Negative	12	50		Negative	2	7

So we can calculate precision and recall from the confusion matrix. Precision, Recall, mAP testing, and training data shown in Table 4. In addition, the visual analysis of the result Mask RCNN model in Figure 7.

TABLE 4.
Performance Mask R-CNN

Performance	Training	Testing
Precision	0.887	0.937
Recall	0.910	0.882
F1 score	0.898	0.803
IoU	0.803	0.833
mAP	0.880	0.903

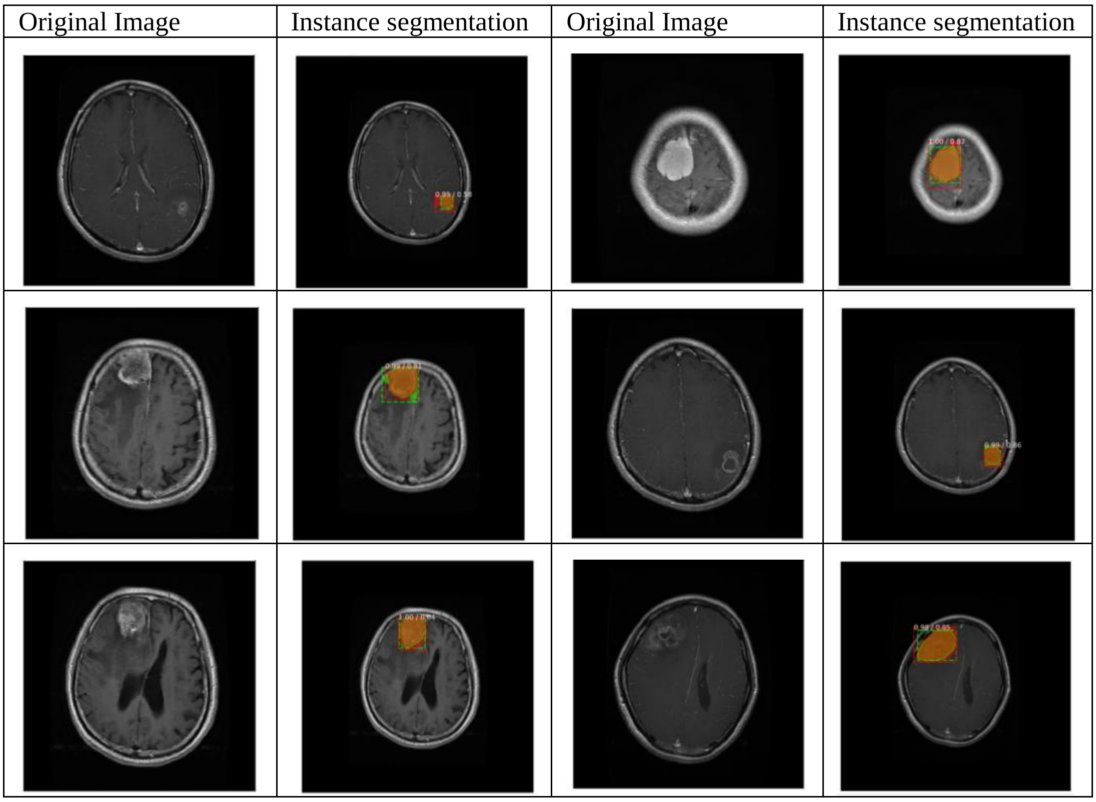

FIGURE 7. Example Brain Tumor Detection with Mask-RCNN

4. RESULT AND DISCUSSION

In this study, we presented an advanced approach for brain tumor segmentation from MRI images using the Mask R-CNN deep learning model. Our method demonstrated a significant improvement in segmentation accuracy, achieving a mean Average Precision (mAP) of 90.3%. This high level of precision underscores the effectiveness of Mask R-CNN in identifying and delineating brain tumors with minimal manual intervention.

The success of our model can be attributed to its robust architecture, which efficiently handles the complexities of MRI-based brain tumor images. By automating the segmentation process, our approach not only enhances the accuracy and consistency of tumor identification but also reduces the time and effort required by medical professionals.

These findings highlight the potential of deep learning models in medical image analysis, paving the way for their broader application in clinical settings. Future work will focus on further improving the model's performance and generalizability by incorporating larger and more diverse datasets, as well as exploring the integration of additional imaging modalities. Our research demonstrates that Mask R-CNN is a promising tool for enhancing the accuracy and efficiency of brain tumor diagnosis and treatment planning.

REFERENCES

[1] Global Cancer Observatory, 360-indonesia-fact sheets, 2019, http://gco.iarc.fr/today/data/factsheets/populations/360-indonesia-fact-sheets.pdf
[2] N. Iriawan, A. A. Pravitasari, U. S. Nuraini et al., "YOLO-UNet Architecture for Detecting and Segmenting the Localized MRI Brain Tumor Image," in Applied Computational Intelligence and Soft Computing, vol. 2024, pp. 1-14, 2024. [Online]. Available: https://doi.org/10.1155/2024/3819801
[3] Pereira, S., et al. "Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images." IEEE Transactions on Medical Imaging, vol. 35, no. 5, pp. 1240-1251, 2016.
[4] Z. Akkus, A. Galimzianova, A. Hoogi, D. L. Rubin, and B. J. Erickson, "Deep learning for brain MRI segmentation: state of the art and future directions," *J. Digit. Imaging*, vol. 30, pp. 449-459, 2017.
[5] D. Shen, G. Wu, and H. I. Suk, "Deep learning in medical image analysis," *Annu. Rev. Biomed. Eng.*, vol. 19, pp. 221-248, 2017.
[6] S. Pereira, A. Pinto, V. Alves, and C. A. Silva, "Brain tumor segmentation using convolutional neural networks in MRI images," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1240-1251, 2016.
[7] Deepak S. and Ameer P. M., Brain tumor classification using deep CNN features via transfer learning, *Computers in Biology and Medicine*. (2019) 111, 103345, https://doi.org/10.1016/j.compbiomed.2019.103345, 2-s2.0-85068241919.
[8] M. Havaei et al., "Brain tumor segmentation with deep neural networks," Med. Image Anal., vol. 35, pp. 18-31, 2017.

ISSN: 2252-4274 (Print)
ISSN: 2252-5459 (Online)

[9] Süleyman Yıldırım M. and Dandıl E., Automatic detection of multiple sclerosis lesions using Mask R-CNN on magnetic resonance scans, *IET Image Processing*. (2020) 14, no. 16, 4277–4290, https://doi.org/10.1049/iet-ipr.2020.1128.
[10] He K., Gkioxari G., Dollar P., and Girshick R., Mask R-CNN, *Proceedings of the IEEE international conference on computer vision*, October 2017, Venice, Italy, 2961–2969.
[11] Mohammed, Bazhdar. Re: How to calculate Precision and Recall?. 2018. Retrieved from: https://www.researchgate.net/ accessed April 2021.
[12] Hui, J. mAP (mean Average Precision) for Object Detection. 2018. Retrieved medium.com accessed April 2021