MRI BRAIN TUMOR CAPTIONING USING VGG IMAGE ANNOTATOR

M.Senthamilselvi[1], Kalavathi Palanisamy[2]

[1,2]The Gandhigram Rural Institute (Deemed to be University),
Gandhigram, Dindigul – 624 302, Tamil Nadu, India.

Abstract

Digital medical images have become increasingly important in the last several years for the detection of more disorders. The health care industry needs an accurate and efficient AI diagnostic system for the compilation of medical reports. This aims for object detection technique, which uses the computer vision (CV) and deep learning(DL) based pre trained Convolutional Neural Networks (CNNs) architectures to extract more accurate features and to classify the objects in the MR Brain images. Natural language processing, and Recurrent Neural Networks (RNNs) techniques were utilized to complete the task of generating high quality captions from these identified features. Automatic image captioning has made a significant progress in recent years, and it is more difficult problem. In order to implement NLP into a medical image captioning system using DL techniques, certain image annotation tools are available as an open source tools. Some popular AI image caption generator tools available which includes LabelImg, MakeSense.ai, SuperAnnotate, CVAT (Computer Vision Annotation Tool) etc. Visual Geometry Group Image Annotator tool (VGGIA), is one of the effective tools available in open source, which helps to define the caption for the regions affected in the MR Brain images. This paper explores the usage of VGGIA for captioning brain tumor image to illustrate NLP approach for deep learning based models.

Keywords Deep learning (DL), Natural Language Processing (NLP), Magnetic Resonance Images (MRI), Visual Geometry Group Image Annotator tool (VGGIA), Artificial Intelligence (AI).

1. INTRODUCTION

The field of medical image captioning has garnered a lot of attention recently. One significant issue facing this subject is the lack of large, high-quality datasets with medical image reports attached. In order to get over this restriction, scientists have looked at the usage of general-purpose foundation models, which can be modified using methods like parameter-efficient transfer learning to tailored for particular medical applications. This proposed study which has the potential to improve patient outcomes and streamline clinical operations.

https://doi.org/10.5281/zenodo.13982032

Medical imaging is a crucial part of modern healthcare, giving non-invasive insights into the human body's interior structures. Clinicians can view tissues and organs with the use of techniques like computed tomography (CT), ultrasound, and magnetic resonance imaging (MRI). This makes it easier to identify and diagnose a variety of medical disorders, including brain tumors. In particular, magnetic resonance imaging (MRI) is the modality of choice for brain imaging because it provides high-resolution images with superior contrast between various tissue types. Efficient processing approaches are necessary to support clinical decision-making due to the increasing volume of medical imaging data.

Algorithms are used in medical image processing to improve, analyze, and interpret images. Image processing techniques are important in the field of brain tumors because they help with tasks like segmentation, which involves defining the boundaries of a tumor, and classification, which involves identifying the type of tumor. The intricacy of brain anatomy and the minute fluctuations in tumor appearance present formidable obstacles that necessitate sophisticated processing techniques to guarantee precision. Deep learning (DL), a branch of AI, has transformed medical image processing by offering strong automated analysis tools. In tasks including brain image classification, brain image segmentation, and brain image tumor detection, CNNs have shown great promise. When it comes to brain tumor analysis, deep learning models can identify complex patterns from large datasets, which makes it possible to diagnose tumors automatically and with a high degree of accuracy. The requirement for scalable, precise, and reliable image analysis which is essential for managing the expanding volumes of medical imaging data gives rise to the need for deep learning.

Because Natural Language Processing (NLP) makes it possible for machines to comprehend and produce human language and deemed it an important to the healthcare industry. NLP can be applied to medical imaging to create descriptive reports from imaging data, which will help radiologists evaluate the findings. The goal of automating report generation to lessen clinician effort and lower the possibility of interpretation errors is what motivates the use of NLP in medical imaging. Image captioning in brain medical imaging can provide meaningful captions for segmented tumors, offering context and information that are essential for diagnosis and therapy planning. When picture captioning is combined with deep learning models such as Mask R-CNN, accurate and clinically meaningful descriptions can be produced, improving the use of medical images in clinical practice.

2. Literature Review

Recent developments in medical brain imaging, particularly in the field of DL, have greatly increased the accuracy and efficiency of brain tumor analysis and detection. According to Pereira et al. [1], who used CNNs for brain tumor segmentation in MRI images. Stated that CNNs have become the mainstay of medical brain image segmentation, setting the foundation for DL applications in medical imaging. Expanding upon this, Ronneberger et al.[2] presented the U-Net architecture, which works especially well for biomedical image segmentation, even in situations with a dearth of labeled data. The design of U-Net has shown to be essential for precise segmentation of structures such as tumors, particularly in situations when data availability is limited.

He et al. [3] improved on this by introducing the Mask R-CNN model, which has become a standard for instance, segmentation tasks. The approach is well suited for fine-grained tumor segmentation since it can provide pixel-wise masks for every object in an image. The use of models like DeepLab (Chen et al.,)[4], which introduces atrous convolution and CRFs boost segmentation accuracy by collecting detailed nuances in medical images, is a response to the difficulty of fine-grained segmentation in complicated images. In addition to these segmentation advances, the automatic production of descriptive captions for segmented tumors is made possible by the incorporation of image captioning models, as proposed by Vinyals et al. [5] with their "Show and Tell" model.

The inclusion of visual attention mechanisms, as proposed by Xu et al. [6], improves the relevance and accuracy of these captions even more by making sure that the generated descriptions emphasise the image regions that are most crucial for diagnosis. Milletari et al. [7] proposed V-Net, a completely CNN designed for 3D medical image segmentation, in the context of volumetric data. The use of this model with MRI scans is especially pertinent to comprehending the three-dimensional architecture of medical brain imaging tumors, which is necessary for precise diagnosis and treatment strategizing. Furthermore, Zhang et al.'s [8] developed a deep residual U-Net which illustrates how residual learning may be incorporated into segmentation models to further increase segmentation accuracy and reliability, making these methods resilient to the difficulties presented by complex medical images.

Efficiency in model architecture has also been a focus in parallel to these advancements, as demonstrated by Iandola et al. [9] using SqueezeNet. With much fewer parameters needed to achieve AlexNet-level accuracy, this model provides a mechanism to implement potent DL models in contexts with limited resources, including remote healthcare practice or mobile diagnostics.

Lastly, Radford et al. [10] have investigated the possibility of incorporating NLP into medical imaging with their CLIP (Contrastive Language-Image Pretraining) model, which acquires visual concepts through natural language supervision. These models' capacity to produce contextually relevant descriptions of medical images creates new opportunities for medical report creation automation, which lessens the cognitive burden on physicians and enhances the uniformity of diagnostic interpretations.

When taken as a whole, these studies offer a thorough foundation for developing sophisticated DL systems for brain tumor identification, detection, and captioning. They also show how CNN-based detection models and NLP-driven captioning frameworks work in concert to improve the usefulness of medical imaging in clinical practice.

3. Captioning Brain Tumor MR Image using VGG Image Annotator

There are various procedures involved in using the VGG Image Annotator (VIA) to annotate MR brain images. After importing MR images into VIA, tumor locations are annotated with the relevant class names and bounding box drawings utilizing tools. Following that, these annotations are exported and stored in JSON format. The JSON data needs to be transformed to Common Objects in Context (COCO) format, which contains segmentation masks, category labels, and image metadata, in order to employ these annotations with a Mask R-CNN model. Using Matterport's implementation architecture, the gathered datasets are split into training and validation sets in order to train the Mask R-CNN model [11]. Tumor detection and segmentation are taught to the model, which is then assessed and adjusted. In order to explain new MR images, the trained Mask R-CNN model is used to extract tumor properties. Either a custom solution or an image captioning model such as “Show and Tell” is then used to describe the new images. Lastly, post-processing entails superimposing the generated captions for display and interpretation over the segmented tumors on MR brain images. The stages and procedures used in the VGG Image Annotator are as follows.

3.1 Steps involved in the VGG Image Annotator

The following below steps are involved for brain tumor detection, segmentation, and captioning.

```
Step 1: Annotation with VGG Image Annotator (VIA), which involves the following sub steps
    - Load MR Images into VIA Tool
    - For each image:
        • Annotate Tumors using Bounding Boxes/Polygons
    - Save Annotations in JSON format
```

```
Step 2: Convert Annotations to Mask R-CNN Format
    - Parse JSON Annotations
    - For each annotation:
        • Convert to COCO Format (Include Image Metadata, Category Labels, Segmentation)
Step 3: Train Mask R-CNN Model
    - Prepare Dataset for Training
    - Train Mask R-CNN Model using the Prepared Dataset
    - Evaluate Model Performance
    - Fine-Tune Model as Necessary
Step 4: Generate Image Captioning
    - For each new MR Image:
        • Detect Tumors using Trained Mask R-CNN Model
        • Segment Tumors (Create Masks)
        • Extract Features from Segmented Images
        • Generate Descriptive Captions based on Extracted Features
Step 5: Post-Processing and Visualization
    - For each processed image:
        • Overlay Segmentation Masks on Original Images
        • Add Generated Captions to Images
```

3.2 Components of VGG Image Annotator (VIA)

3.2.1. *User Interface (UI): Components and Interactions:*

- *Image/Video Loader*: Allows users to select and load images or videos from local storage.
- *Annotation Tools:* Provides tools like bounding boxes, polygons, and points for image annotation.
- *Attribute Editor*: Enables users to add labels, tags, and other metadata to annotations.
- *Save/Export Button:* Allows saving and exporting annotations in formats like JSON or CSV.
- *Review/Refinement Panel:* Provides options to review, modify, or refine existing annotations.
- The user interacts with these components to perform annotation tasks.

3.2.2. Backend Processing: Components and Interactions

- *Annotation Engine:* Manages the creation and manipulation of annotations.
- *Data Storage:* Handles temporary storage of annotations before they are saved or exported.
- *Export Handler:* Converts annotations into formats like JSON or CSV for export.
- The UI sends annotation data to the Annotation Engine.
- The Annotation Engine communicates with Data Storage to save and retrieve annotation data.
- The Export Handler retrieves data from Data Storage to create exportable files.

3.3.3. Data Flow

- *Importing Data:* User loads images/videos through the UI.
- *Annotation Data:* User annotations are sent to the Annotation Engine and stored temporarily in Data Storage.
- *Saving Data:* Upon saving, the data is processed by the Export Handler and saved in the desired format.

3.2.4. Output: Components

- *JSON/CSV Files:* Exported annotation files.
- *Annotated Images/Videos:* Files ready for use in machine learning tasks.

The following architecture diagram (Figure 1) captures the key components and their interactions within the VGG Image Annotator tool.

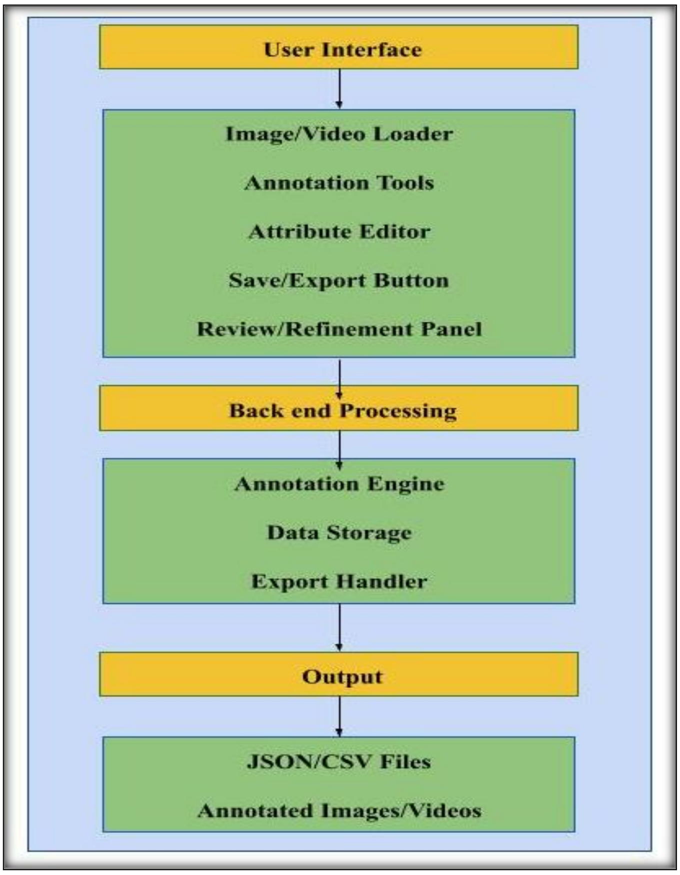

Figure 1. VIA Architecture Diagram

In this proposed study for MR brain tumor analysis and detection, the VGG Image Annotator (VIA) tool is indispensable, especially when bounding box selection is being used to identify tumor locations. By drawing bounding boxes around tumor locations, this technique enables medical professionals to manually and precisely annotate MR images, producing precise training data for AI models. Multi-class annotation is supported by VIA, which makes it possible to identify distinct brain regions and helps construct models that can discriminate between different kinds of tumors and surrounding tissues. big-scale annotation projects benefit greatly from the tool's ability to classify MR images in an organized manner. This is important because training deep learning models requires big datasets. It is simple to export annotations in CSV or JSON formats for usage in machine learning workflows. VIA's support for bounding box annotations guarantees precise segmentation and acts as ground truth for models that are trained to precisely identify and segment tumors. Furthermore, VIA facilitates multi-modal imaging, guaranteeing uniform annotation across several MR modalities a crucial aspect of training resilient models. In addition to offering capabilities for error correction and improvement, the platform promotes expert collaboration and guarantees high annotation accuracy and consistency. VIA is a useful tool for research and large-scale studies centered on brain tumor detection and analysis because of its scalability and customization flexibility [12]. Following are the important actions to be utilizing when using VIA tools:

- *Image Import*: VIA tool to load MRI images.
- *Annotation:* To mark tumor locations with exact forms and designate them as tumors using VIA's capabilities.
- *Export*: TO Store the annotations in a JSON format that includes information on the class and location of the tumor.

3.3 Mask R-CNN Format Conversion

The VIA annotations are transformed into a model-compatible format (Mask R-CNN). This involves the following steps:

- Data extraction from VIA's JSON output is known as parsing JSON.
- *Transforming to COCO Format*: converting the annotations to the COCO format, which consists of segmentation masks, category labels, and image information training and validation of images.

3.4 Training Model (Mask R-CNN)

This employed for tumor detection and segmentation. The training process involves the following steps

- *Dataset Preparation:* Organizing the dataset into training and validation sets using the COCO format.
- *Model Training:* Training the Mask R-CNN model with customized parameters to identify and segment tumors.
- *Evaluation:* Assessing the model's effectiveness and refining it according to validation results.

3.5 Image Captioning

For MR images, captions are generated after segmentation. The steps for captioning images in VIA are as follows:

- *Tumor Detection*: Use the trained (Mask RCNN) model to detect and segment tumors in new MR images.
- *Caption Generation*: Employ an image captioning model to describe the detected tumors based on their features and context.

4. Results and Discussion

In the following Figure 2, illustrates how VIA is used to caption tumors on MR brain images. MR brain image region selection was the main focus of the evaluation of the precision and accuracy of annotations made using the VIA tool. Problems like precisely defining tumor boundaries were tackled, as the marked areas of brain images illustrate in the given diagram below.

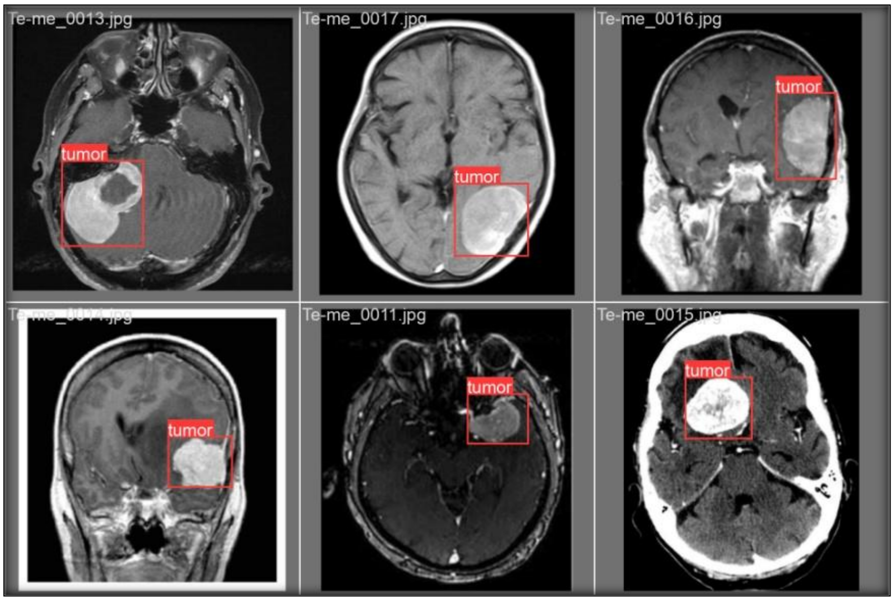

Figure 2. Annotating the region of the brain tumor in MR brain image Using VGG IA

Metrics like IoU (Intersection over Union) and mAP (mean average precision) were specifically used to measure the detection accuracy of the model (Mask R-CNN). This model achieved an 84% mAP after 200 epochs [13]. This is seen in Figure 3, where the model successfully detects and localizes the tumor with a confidence score of 0.87 on a testing brain MR image. Further pictures demonstrate the model's resilience in recognizing tumor locations and validate its consistent performance across several MRI scans.

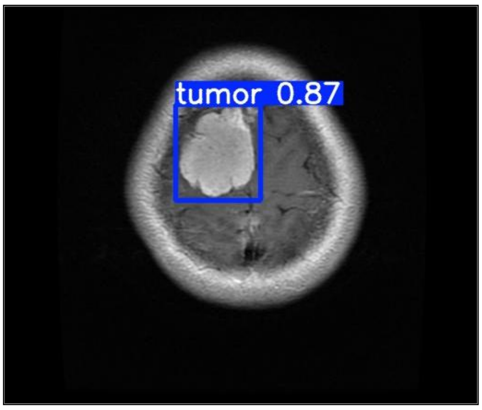

Figure 3. A sample output obtained from Mask R-CNN

4. Conclusion

The benefits of employing Mask R-CNN for exact brain tumor detection and captioning and VIA for correct annotation are outlined in this research. When combined with VIA, YOLOv5 models and file formats may provide comparable advantages including improved speed and effectiveness. By using annotation approaches, optimizing model performance, and assessing the efficacy of substitute tools, this tool can be improved. This model is regarded as one of the general-purpose and medical image processing approach image captioning models.

References

1. Pereira, S., Pinto, A., Alves, V., & Silva, C. A. Brain Tumor Segmentation Using Convolutional Neural Networks in MRI Images.(2016).
2. Ronneberger, O., Fischer, P., & Brox, T. U-Net: Convolutional Networks for Biomedical Image Segmentation. (2015).
3. He, K., Gkioxari, G., Dollár, P., & Girshick, R. Mask R-CNN. (2017).
4. Chen, L. C., Papandreou, G., Kokkinos, I., Murphy, K., & Yuille, A. L. Deeplab: Semantic Image Segmentation with Deep Convolutional Nets, Atrous Convolution, and Fully Connected CRFs. (2017).
5. Vinyals, O., Toshev, A., Bengio, S., & Erhan, D. Show and Tell: A Neural Image Caption Generator(2015).
6. Xu, K., Ba, J. L., Kiros, R., Cho, K., Courville, A., Salakhutdinov, R., ... & Bengio, Y. Show, Attend and Tell: Neural Image Caption Generation with Visual Attention. (2015).
7. Milletari, F., Navab, N., & Ahmadi, S. A. V-Net: Fully Convolutional Neural Networks for Volumetric Medical Image Segmentation. (2016).
8. Zhang, Z., Liu, Q., & Wang, Y. Road Extraction by Deep Residual U-Net. (2018).
9. Iandola, F. N., Han, S., Moskewicz, M. W., Ashraf, K., Dally, W. J., & Keutzer, K. SqueezeNet: AlexNet-level Accuracy with 50x Fewer Parameters and <0.5MB Model Size. (2016).
10. Radford, A., Kim, J. W., Hallacy, C., Ramesh, A., Goh, G., Agarwal, S., & Amodei, D. Learning Transferable Visual Models From Natural Language Supervision. (2021).
11. https://github.com/matterport/Mask_RCNN
12. https://blog.roboflow.com/vgg-image-annotator/
13. He, K., Gkioxari, G., Dollár, P., & Girshick, R. (2017). "Mask R-CNN." *Proceedings of the IEEE International Conference on Computer Vision (ICCV).* https://doi.org/10.1109/ICCV.2017.322