

Brain Tumor Image Dataset: Semantic Segmentation

Final Project for the Introduction to Deep Learning Course

Fan-Pin Tseng, Sen-Hao Ma, Jing-Jing Liao, Ming-Yi Huang

December 26, 2024



① Introduction

② Methods

③ Results

④ Conclusion

1 Introduction

2 Methods

3 Results

④ Conclusion

Introduction

- Problem Definition
- Dataset Overview
- Background: MRI Images

Problem Definition

Task

- Identify brain tumor regions on MRI images with **semantic segmentation**.

Semantic Segmentation

- The process of taking an image and labeling each pixel in that image with a certain class. In our case, **1: Tumor** (foreground), **0: non-tumor** (background).
- **Input:** An MRI image of size 640×640 pixels.
- **Output:** A binary mask of the same size.

Goal

- Accurately classify each pixel as **tumor or non-tumor**.
- Improve segmentation performance for medical imaging tasks.

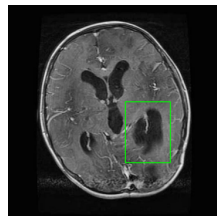
Dataset Overview: Brain Tumor Image Dataset

Dataset

- **Name:** Brain Tumor Image DataSet : Semantic Segmentation
- **Kaggle:** <https://www.kaggle.com/datasets/pkdarabi/brain-tumor-image-dataset-semantic-segmentation>
- **Preprocessing:** auto-oriented, stretched to 640×640 pixels, and unaugmented.

Data Split

- **Training set:** 1502 Images (70%)
- **Validation set:** 429 Images (20%)
- **Testing set:** 215 Images (10%)



Features of Brain MRI

Major Imaging Sequences:

- **T1-weighted Imaging:**

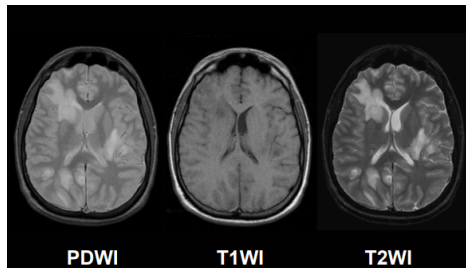
- Clear anatomical details
- Cerebrospinal fluid (CSF) appears dark
- fatty tissues appear bright

- **T2-weighted Imaging:**

- Highlights water content
- CSF and lesions often appear bright

- **PDWI (Proton Density-Weighted Imaging):**

- Highlights the distribution of hydrogen protons in tissues, reflecting proton density
- Provides detailed anatomical information, particularly effective for visualizing cartilage, gray matter, and white matter



1

¹Image source: http://home.ee.ntu.edu.tw/classnotes/bme2/2006/10_23_06.pdf

Reading 3D MRI Brain Images

Radiologists read images from three planes:

- **Axial Plane:** Horizontal slices (top-to-bottom view)
- **Coronal Plane:** Perpendicular to axial plane (front-to-back view)
- **Sagittal Plane:** Midline of the body (left-to-right view)

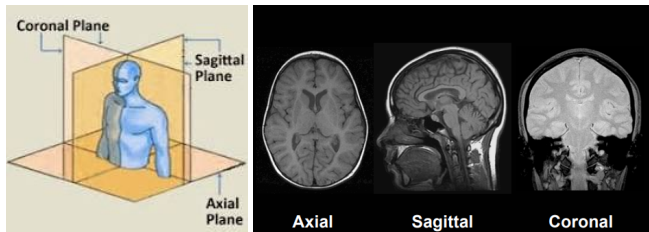


Figure 1: Illustration of axial, sagittal, and coronal planes.^{2 3}

²<https://pro.boehringer-ingelheim.com/us/ipfradiologyrounds/hrct-primer/image-reconstruction>

³Image source: http://home.ee.ntu.edu.tw/classnotes/bme2/2006/10_23_06.pdf

Issues of the Dataset (1/2)

Issues in images

- The provided images are in 2 dimensional, and
- the dataset is the mixture of axial view, coronal view, and sagittal view.

Why is this a problem?

- Difficult to interpret and analyze consistently
- Misalignment of viewpoints affects segmentation and diagnosis

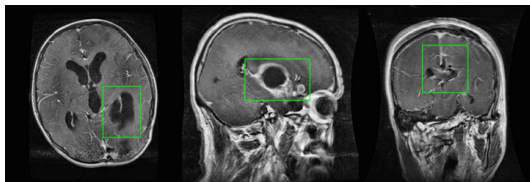


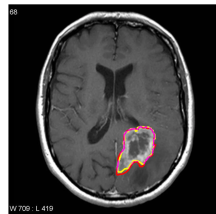
Figure 2: Images of axial view, sagittal view, and coronal view. ⁴

⁴Source: **Training dataset**

Issues of the Dataset (2/2)

Consequences:

- Reduced diagnostic accuracy
- Incorrect segmentation of brain regions
- Difficult to spot the region of interest (red segment)



Annotation Issues:

- In medical practice, **region of interest (ROI)** is the choice of annotation.
- **Bounding boxes (BBox)** lack precision and include unnecessary regions.

1 Introduction

2 Methods

3 Results

4 Conclusion

Annotation Approach and Modeling Strategy

Annotation Preference: ROI over BBox

- Although our dataset provides BBox annotations, we adhere to real-world medical scenarios and avoid using BBox-based methods, such as **YOLO**.
- Instead, we adopt an **ROI-based approach** for tumor segmentation.

Modeling Strategy:

- Four NN architectures: **U-Net, Residual U-Net, Attention U-Net, and Nested U-Net (UNet++)**.
- Final prediction of each pixel by the ensemble votes.
- U-Net-like architecture is suitable for precise medical image segmentation, particularly for tasks involving ROI detection.

U-Net Architecture

Key Features:

- Encoder-Decoder structure:
 - **Encoder:** Contracting path (downsampling), extracts features.
 - **Decoder:** Expanding path (upsampling) uses transpose convolution to upsample features and reconstruct image to original size.
- Skip Connection
 - In upsampling, **concatenates** features from corresponding downsampling layers to **alleviate vanishing gradient**
- 1×1 convolution
 - Transform multiple channels to a single-channel mask

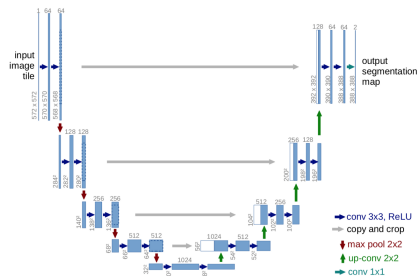


Figure 3: U-Net Architecture⁵

⁵Image source: <https://arxiv.org/pdf/1505.04597>

Residual U-Net

Key Features:

- Encoder-Decoder structure with **residual blocks** extrapolated from ResNet
- A **residual block** consists of stacked layers such as batch normalization (BN), ReLU activation, weight layer (i.e. convolutional layer).
- A **shortcut connection (SC)** skips some layers in the NN and goes through a convolutional layer to maintain the dimension with the output of the main convolutional block.
- SC improves feature learning and **mitigate vanishing gradient**.

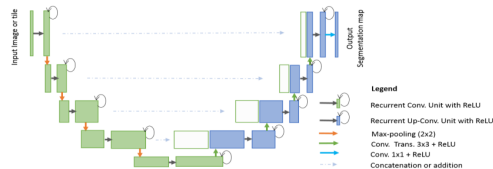


Figure 4: Residual U-Net architecture⁶

⁶Image source: <https://arxiv.org/pdf/1802.06955>

Attention U-Net

Key Features:

- Encoder-Decoder Structure: integrates **attention gates** into UNet.
- **Attention Gates (AGs)**: filter the features propagated through the skip connections and improve model sensitivity to foreground pixels without requiring complicated heuristics.
- Enhance the segmentation performance by focusing on target regions while suppressing irrelevant background.
- Particularly effective for detecting small objects or low-contrast segmentation.

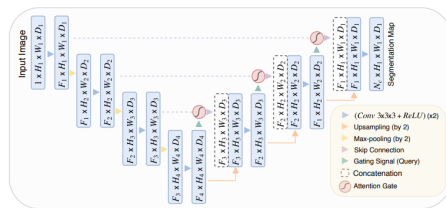


Figure 5: Attention U-Net architecture⁷

⁷Image source: <https://arxiv.org/pdf/1804.03999>

Nested U-Net (UNet++)

Key Features:

- **Main idea** is to bridge the semantic gap between the feature maps of the encoder and decoder prior to fusion.
- **Encoder-decoder** is connected through a series of nested dense convolutional blocks.
 - Each block includes batch normalization after each convolution for better training stability.
- **Dense skip connections:**
 - Aggregate features from shallow layers to deeper layers.
 - Preserve fine-grained details and combine global semantic context.

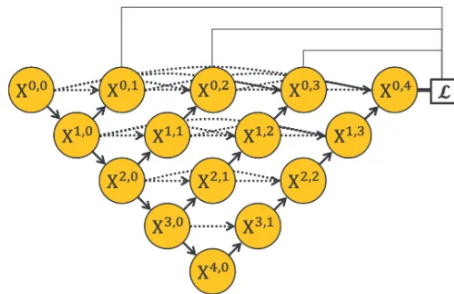


Figure 6: Nested U-Net architecture⁸

Vote Strategy

Overview

- **Ensemble voting**
 - Models independently predict masks
 - Count predictions of each pixel from models of U-Net, Residual U-Net, Attention U-Net, and Nested U-Net
 - A pixel is marked as tumor if total votes for this pixel is greater than threshold
- **Voting thresholds**
 - Threshold = 2, achieves better segmentation performance
 - Threshold = 3, more aligned with medical practices, slightly ineffective

3

Loss Function

- **Dice Loss Function**

- Designed for image segmentation tasks
- Measuring the **overlap between predicted and ground truth masks**
- **Handle imbalanced data** by focusing equally on foreground and background regions
- **Boundary-aware** ensures better segmentation along ROI boundaries
- **Formulas:**

$$\text{Dice Loss} = 1 - \text{Dice Coefficient}$$

$$\text{Dice Coefficient} = \frac{2 \cdot |P \cap T| + \epsilon}{|P| + |T| + \epsilon}$$

Where:

- P : Predicted binary mask.
- T : Target binary mask.
- $|P \cap T|$: Intersection of P and T .
- $|P|$, $|T|$: Total pixels in P and T .
- ϵ : A small positive constant to avoid division by zero.

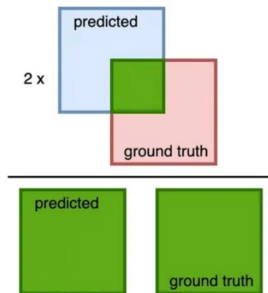


Figure 7.9

Evaluation Metrics

- **Dice Coefficient**

- **Accuracy**

- Measures the proportion of correctly predicted pixels out of the total number of pixels

- **Formula:**

$$\text{Accuracy} = \frac{|P = T|}{N}$$

- P : Predicted binary mask. T : Target binary mask. N : Total number of pixels.
- $|P = T|$: Number of pixels where P equals T .

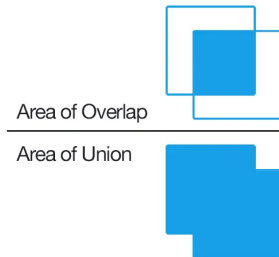
- **Intersection over Union (IoU)**

- Quantifies the overlap between the predicted binary mask P and the target binary mask T relative to their union.

- **Formula:**

$$\text{IoU} = \frac{|P \cap T| + \epsilon}{|P \cup T| + \epsilon}$$

- $|P \cup T|$: Union of P and T , calculated as $|P| + |T| - |P \cap T|$.



Benchmarks of IoU and DICE

Intersection over Union (IoU):¹¹

- **IoU < 0.5:** Poor segmentation; requires significant improvement.
- **IoU 0.5 – 0.7:** Acceptable but could be improved, suitable for general tasks.
- **IoU > 0.7:** Good segmentation; high accuracy.

Dice Coefficient:¹²

- **Dice < 0.7:** Poor performance; not ideal for most applications.
- **Dice 0.7 – 0.85:** Reasonable performance; acceptable for many tasks.
- **Dice > 0.85:** Excellent performance; near-perfect segmentation.

¹¹<https://yy-programer.blogspot.com/2020/06/iouapmap.html>

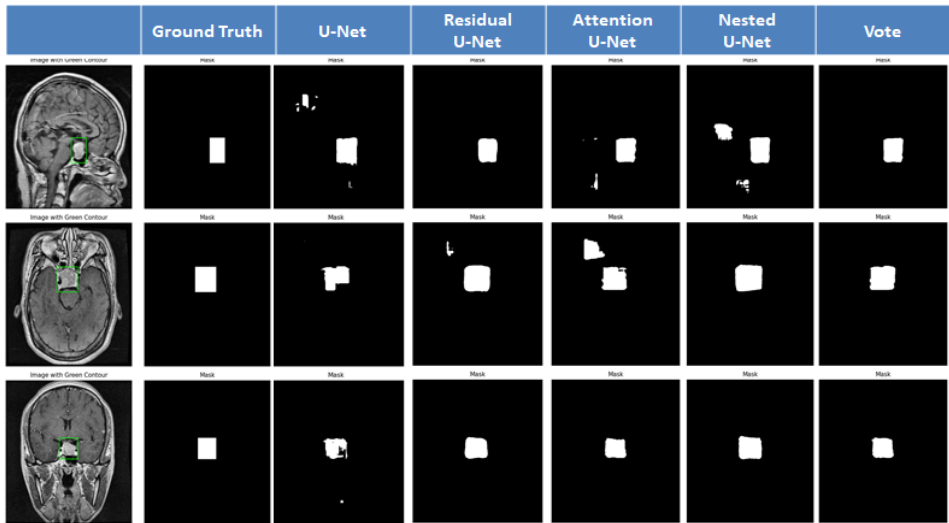
¹²<https://arxiv.org/pdf/2301.06358>

Model Evaluation

Table 1: Model Evaluation Metrics on Test Set

Evaluation Metric(%)	U-Net	Residual U-Net	Attention U-Net	Nested U-Net	Vote
Average Dice Coefficient	57.93	59.80	63.34	63.71	67.94
Average Accuracy	97.14	97.80	97.58	97.46	97.86
Average IoU	44.71	49.33	51.32	51.54	55.90

Illustration of Results (The first three images of the test set, threshold = 3)



4

Conclusion

- Findings
- Discussion
- Future Work
- GitHub: code, documents, and more...

Findings (1/3)

1. Challenges with Bounding Box Labels

- Segmentation labels were provided as bounding boxes rather than pixel-level annotations, making precise segmentation challenging.
- Thus, our models can only assist physicians by quickly identifying suspected tumors and marking them, serving as a CAD (computer-aided diagnosis) tool.

2. Model Performance

- The trained models achieved **IoU = 0.5** and **Dice = 0.6**, indicating that the models could capture the approximate tumor region.
- Based on ensemble voting, we anticipate that **IoU could increase to 4.36%** and **Dice to 4.23%**, potentially improving performance.

Findings (2/3)

3. Reasons for Limited Accuracy

- Mixed MRI viewpoints in the dataset increased segmentation difficulty, as the models could not focus on a consistent anatomical structure.
- The inherently 3D nature of MRI data limited context, as 2D slices lack information from adjacent slices.

4. Model Variants Comparison

- **ResU-Net:** Showed advantages in handling boundary sharpness.
- **AttU-Net:** Effectively differentiated between background and tumor.
- **U-Net++:** Improved segmentation of fine structures.

Findings (3/3)

5. Data Augmentation

- We implemented the **albumentations library** for data augmentation. Due to time constraints, we could not finish full training and hyperparameter tuning.
- We expect data augmentation to improve model performance further.

6. GPU Resource Limitations

- Processing medical images is computationally intensive. We had to reduce batch sizes to 2 on standard GPUs, including those accessed via Colab.

Discussion

Model Performance and Feasibility

- Our models semantic segment tumor regions on MRI images with four UNet-like architectures.
- We achieved an acceptable level of performance (refer to IoU and Dice metrics) that demonstrates the feasibility of our approach.

Implications

- Our model can serve as an auxiliary tool for medical professionals, such as neurologists and radiologists, as quick identification and labeling of tumors on MRI slices.

Future Work

Viewpoint Separation

- Manually annotate and separate data by viewpoints (**axial, sagittal, coronal**) to allow each view to be processed individually.
- Low-level features could share common weights to optimize training.

Data Augmentation

- Complete training and fine-tuning with augmented data
- Validate its impact on model performance.

Resource Optimization

- Explore methods to improve computational efficiency and allow larger batch sizes, e.g., mixed precision training.

Collaboration with medical professionals

- Interview domain experts (qualitative analysis)
- Compare model predictions with clinical evaluation

Code Availability and Model Deployability

Code and Documents Repository

- The source code and comprehensive documentation for our project are publicly accessible on GitHub.
- Link: https://github.com/peaceway10608/MRI_project

Model Deployability

- We use **Docker** to ensure deployability across heterogeneous platforms.
- Instructions for building and running the Docker container are included in the repository.



Q & A