Multimodality Brain Tumour Segmentation With LLM

Yun Zhang | Supervised by Professor Liangqiong Qu

The University of Hong Kong - Department of Statistics and Actuarial Science

Abstract

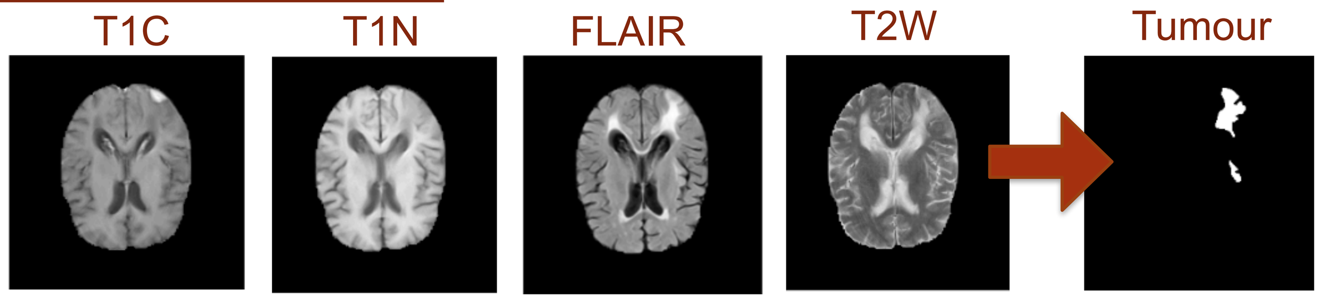
Severity of Brain Tumours[1]:

2 MILLION PEOPLE are living with a brain tumour

35.7% SURVIVAL RATE for malignant brain tumour

37,980 PEOPLE will die from it in 2024

Four Modalities of MRI:



Objective: develop a robust, efficient, multimodal framework for accurate brain tumour segmentation by integrating various MRI modalities using LLM + SAM.

Motivation: Why LLM + SAM[2]?

MANUALLY	nnUNet	LLM + SAM
Tedious	Complex	Simple
Error Prone	Purely Visual	Visual + Contextual
Time Consuming	Static	Dynamic
Multi-modality	Single Modality	Multi-modality

Contributions:

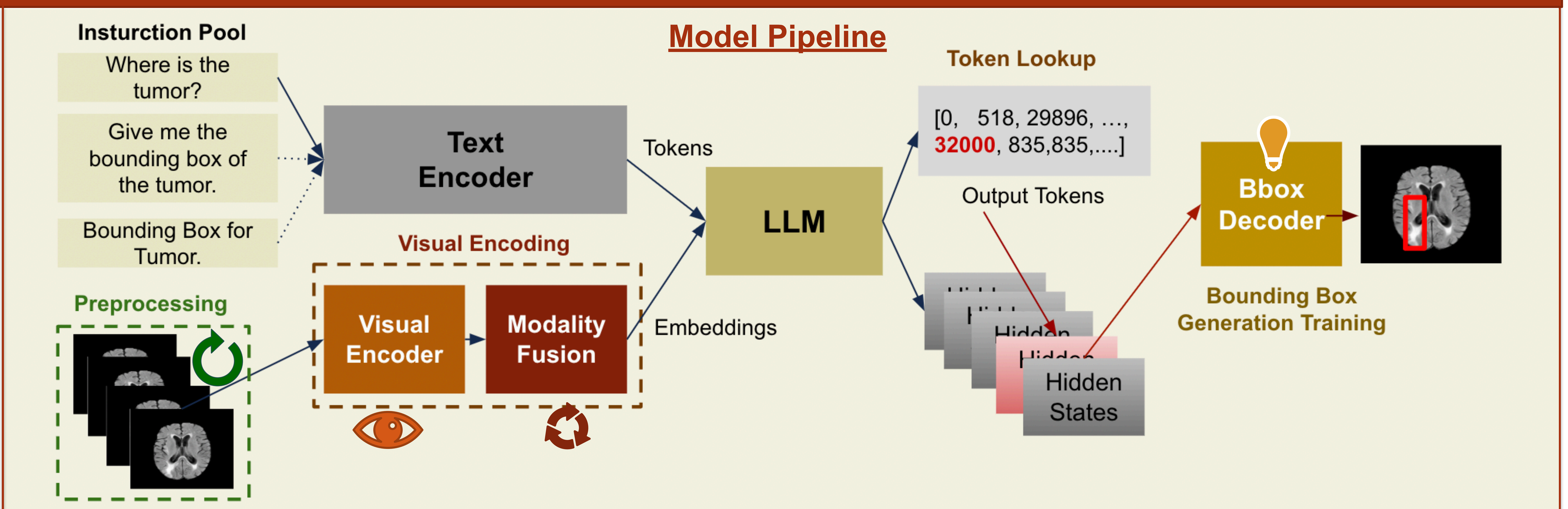
- Established a model using Large Language Models (MiniGPT4[3]) to **generate bounding boxes** for tumours in brain MRI images.
- Developed a comprehensive framework to **integrate multi-modality MRI data** (T1c, T1n, T2, FLAIR) for improved tumour segmentation.
- Established a truly **simple, ready to use** model for users with zero expertise on machine learning.

Methodology

Challenges

- Need to implement a mechanism to train LLM to **generate bounding box** for images with given instructions.
- LLM is trained on natural images, which struggles with understanding **medical images.**
- Need to develop an innovative approach to let the LLM to generate a **synchronised bounding box** for four different input pictures.
- Need to **increase the correctness** (measured by Intersection Over Union) of the LLM predicted bounding boxes.

Model Pipeline

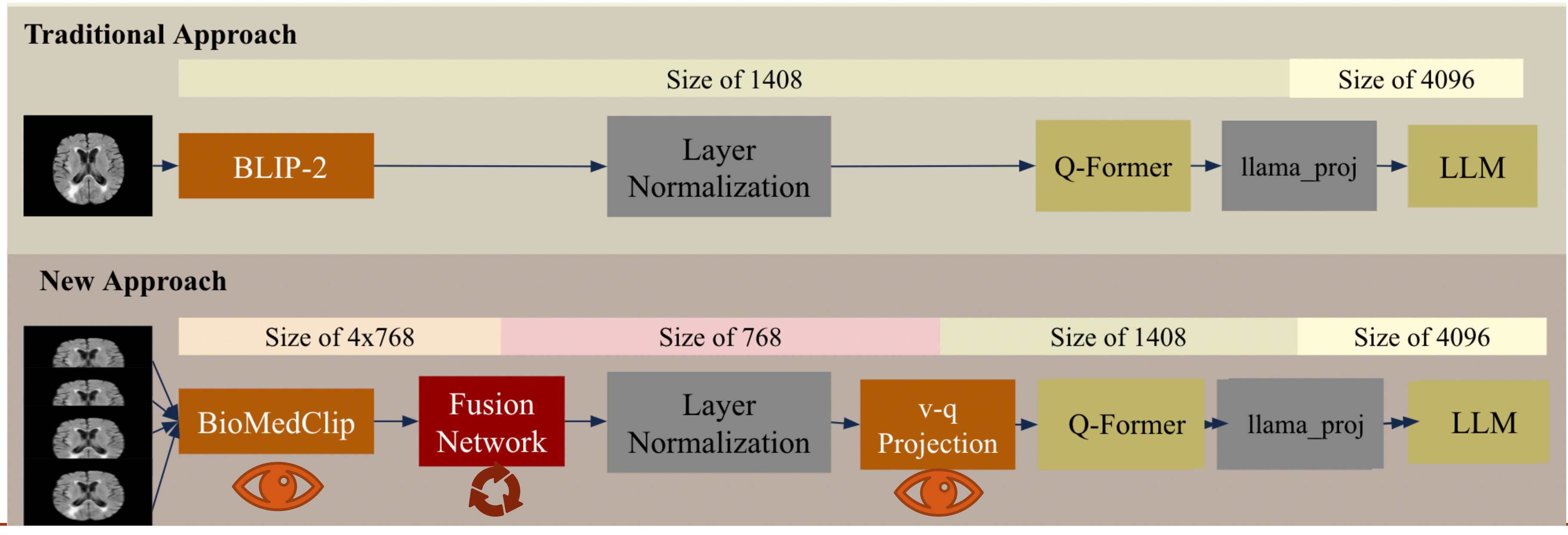


Preprocessing: Except for resizing and normalization transformation, add random rotation (p = 0.4) and flipping (p = 0.2) during prepossessing to enhance model understanding.

Visual Encoding

Specialised Visual Encoder: Utilized BioMedClip[4] to enable LLM to understand medical images.

Modality Fusion: Established a fusion network that enables LLM to accept multiple images as input.



Bounding Box Generation Training

Used the following loss functions to evaluate the model's performance, guiding it towards the desired direction of achieving higher accuracy in bounding box prediction.

GIoU Loss:

$$L_{GIoU}(\hat{t}, t) = 1 - GIoU(\hat{t}, t)$$

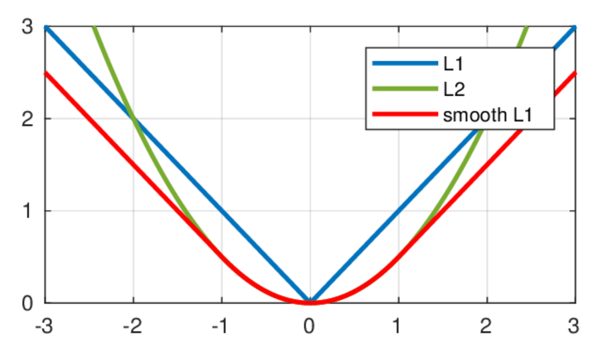
$$IoU = \frac{Intersection}{Union} \qquad GIoU = IoU - \frac{Area\ of\ the\ smallest\ enclosing\ box - Union}{Area\ of\ the\ smallest\ enclosing\ box}$$

L1 Loss

$$L_{L1}(\hat{t}, t) = \sum_{i \in (minx, miny, maxx, maxy)} smooth_{L1}(\hat{t}_i - t_i)$$

$$\text{smoothL1}(x) = \begin{cases} 0.5\, x^2 & if\ |x| < 1 \\ |x| - 0.5 & otherwise \end{cases}$$

- $\hat{t}$: predicted bounding box $(\widehat{minx}, \widehat{miny}, \widehat{maxx}, \widehat{maxx})$
- t: ground truth bounding box $(minx, miny, maxx, maxx)$



Results

BLIP-2 Vs. BioMedClip | Standard vs. Additional Preprocessing | Single Modality Vs. Multi-Modality

Tumour Type	IoU	Relative Increase	IoU	Relative Increase	IoU	Relative Increase	IoU
GLI	0.208	+ 145.2%	0.582	+ 3.4%	0.602	+ 8.4%	0.653
MEN	0.232	+ 157.8%	0.598	+ 1.2%	0.605	+ 10.7%	0.670
MET	0.219	+ 184.0%	0.622	+ 1.5%	0.631	+ 3.2%	0.651

BLIP-2: Single modality model with standard preprocessing and BLIP-2 as visual encoder

Added BioMedClip: Single modality model with standard preprocessing and BioMedClip as visual encoder

Added Additional Preprocessing: Single modality model with additional preprocessing and BioMedClip as visual encoder

Added Multi-modality: Multi-modality model with BioMedClip as visual encoder, and additional preprocessing

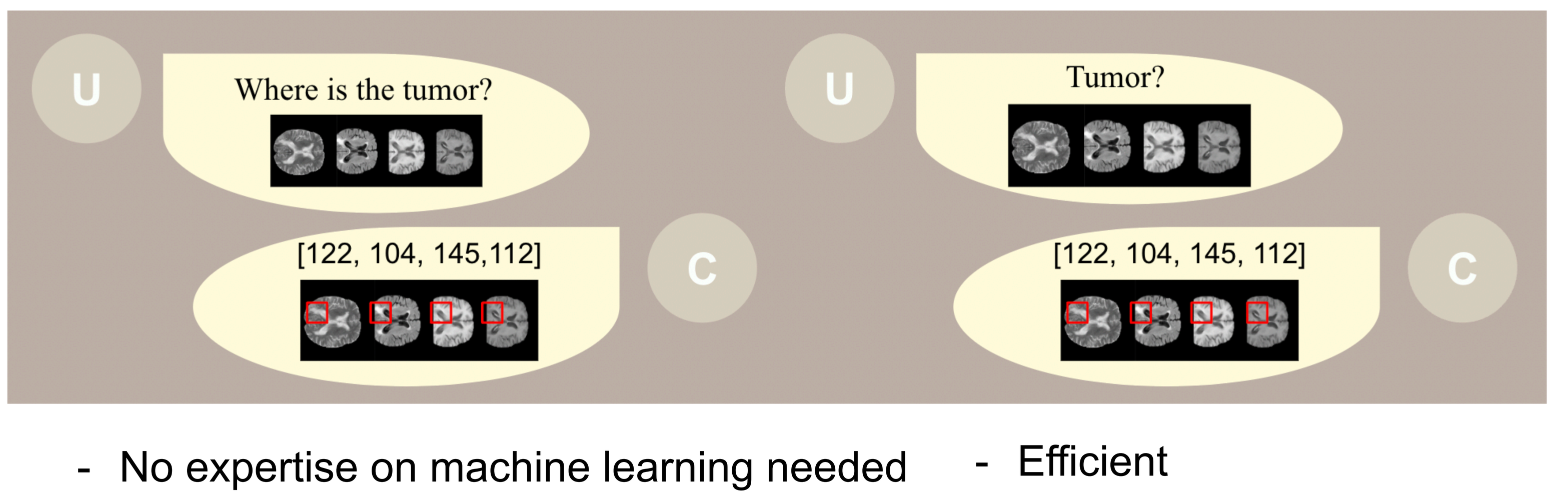
Conclusion & Evaluation

Overall IoU 0.211 → + 212.3% → Overall IoU 0.659

Established a LLM for bounding box prediction in Brain Tumour MRI images.

By adding the specialized visual encoder, additional preprocessing, and incorporating MRI data from all modalities, we achieved a 212.3% increase in overall IoU.

Example Use Case



- No expertise on machine learning needed
- Accept various kind of prompt
- Efficient
- Easy to use

Future Work

1. Connect to **SAM** for more detailed mask generation.
2. Gather **patient data** on contextual information (gender, age) and apply it to the training process.
3. Explore potential adaptations to the **fusion network.**

Reference:

[1] Brain Tumor Facts - National Brain Tumor Society. (2024, February 20). National Brain Tumor Society. https://braintumor.org/brain-tumors/about-brain-tumors/brain-tumor-facts/

[2] Kirillov, A., Mintun, E., Ravi, N., Mao, H., Rolland, A., Gustafson, L., ... & Dollár, P. (2023). Segment Anything. arXiv preprint arXiv:2304.02643. Retrieved from https://arxiv.org/abs/2304.02643

[3] Zheng, K., He, X., & Wang, X. E. (2023). MiniGPT-5: Interleaved Vision-and-Language Generation via Generative Vokens. arXiv preprint arXiv:2310.02239.

[4] Zhang, S., et.all . BiomedCLIP: a multimodal biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv.org. https://arxiv.org/abs/2303.00915biomedical foundation model pretrained from fifteen million scientific image-text pairs. arXiv.org. https://arxiv.org/abs/2303.00915