# A Hybrid Unet-Transformer Architecture for Medical Image Segmentation

**Team**

*Zhekai Han & Shihao Wang*

Georgetown University Computer Science Department

# Background

## Medical Image Segmentation

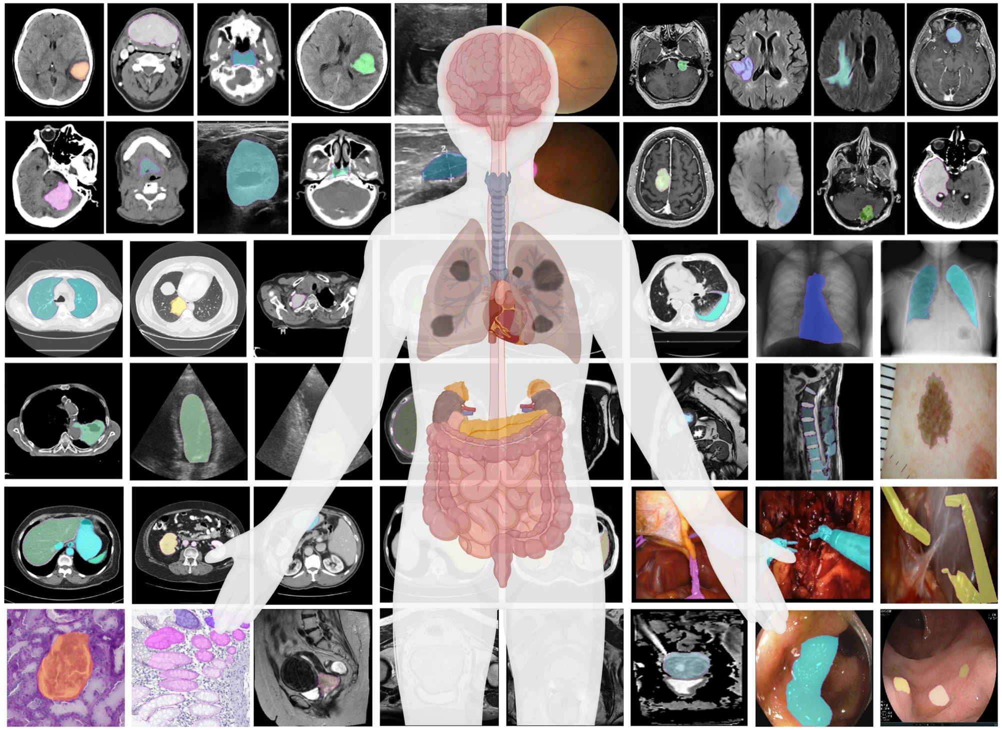


Image source: Ma, Jun, et al. "Segment anything in medical images." Nature Communications 15.1 (2024): 654.

# Background

## Unet

- \+ Excellent at handling images with noise and contrast variations.
- \+ Great for capturing detailed local features like tumors.
- \- Limited in understanding the overall image context.

## Transformer

- \+ Strength in understanding the overall layout of images.
- \+ Helps in grasping global information across the image.
- \- Requires substantial computational resources.
- \- May lack precision for very detailed tasks.

# Motivation

- Inspired by the pioneering work of Ali Hatamizadeh from their 2022 study, titled “Unetr: Transformers for 3d medical image segmentation”.
- While this existing method addressed some challenges in medical image segmentation, we saw room for significant enhancements.
- Our team has innovated upon the original structure to significantly increase the accuracy and efficiency of the segmentation process.

# Goal

- To refine and expand upon the methodology proposed by Ali Hatamizadeh in their influential work.
- We have redesigned the structural components of the proposed model.
- Our enhancements better suit the complex demands of medical imaging, enhancing the model's performance and clinical applicability.

# Main Reference Work(s)

UNETR: Transformers for 3D Medical Image Segmentation

BEFUnet: A Hybrid CNN-Transformer Architecture for Precise Medical Image Segmentation

Sepvit: Separable vision transformer

Fully Convolutional Networks for Semantic Segmentation

https://github.com/Project-MONAI/MONAI

# Dataset

**Target:**

Gliomas segmentation necrotic/active tumour and oedema

**Modality:**

Multimodal multisite MRI data (FLAIR, T1w, T1gd, T2w)

**Size:**

750 4D volumes (484 Training + 266 Testing)

7.09GB

**Source:**

BRATS 2016 and 2017 datasets

**Challenge:**

Complex and heterogeneously-located targets

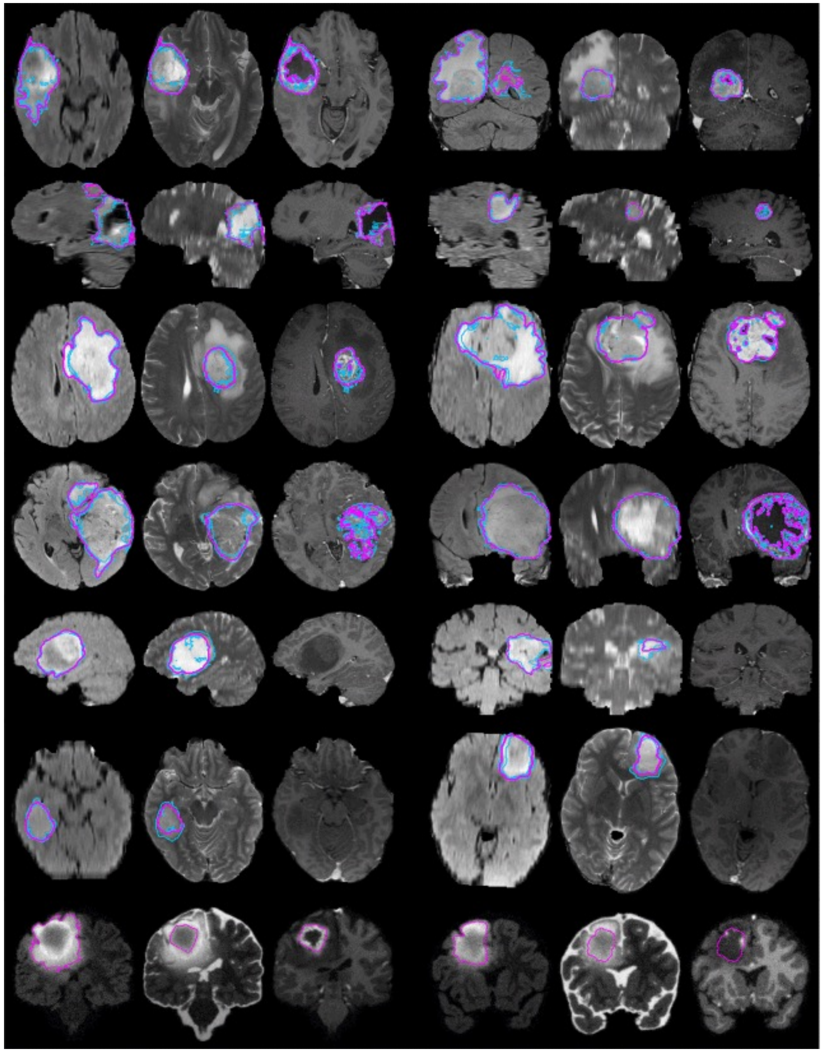


Image source: Menze, Bjoern H., et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." IEEE transactions on medical imaging 34.10 (2014): 1993-2024.

# Dataset

- the whole tumor (yellow) visible in T2-FLAIR (Fig. A).
- the tumor core (red) visible in T2 (Fig. B).
- the enhancing tumor structures (light blue) visible in T1Gd, surrounding the cystic/necrotic components of the core (green) (Fig. C).
- The segmentations are combined to generate the final labels of the tumor sub-regions (Fig. D): edema (yellow), non-enhancing solid core (red), necrotic/cystic core (green), enhancing core (blue).

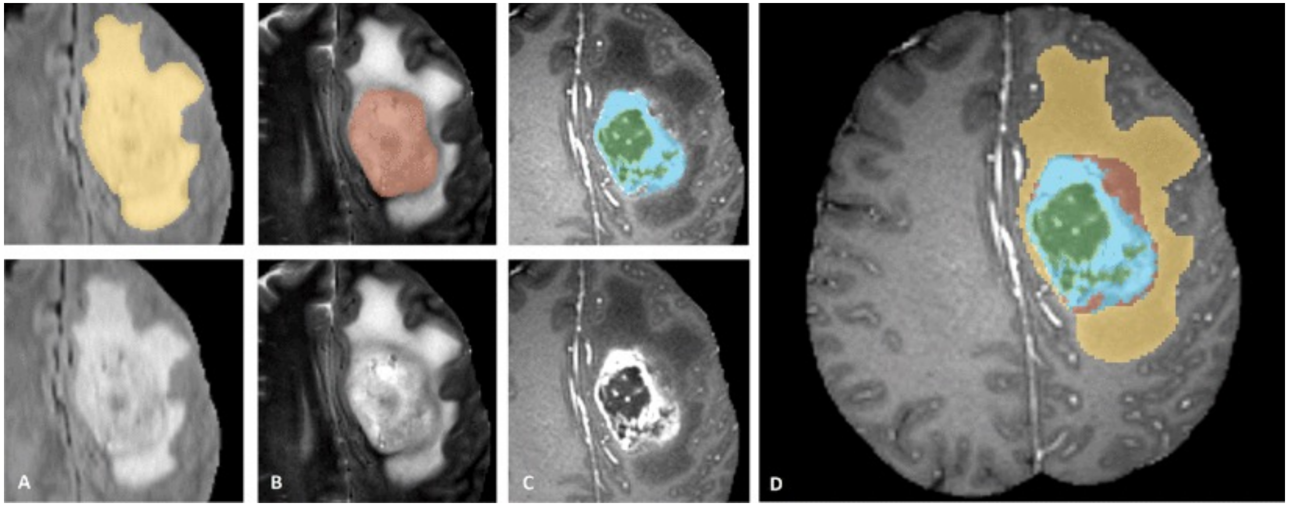



Image source: Menze, Bjoern H., et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." IEEE transactions on medical imaging 34.10 (2014): 1993-2024.

# Architecture

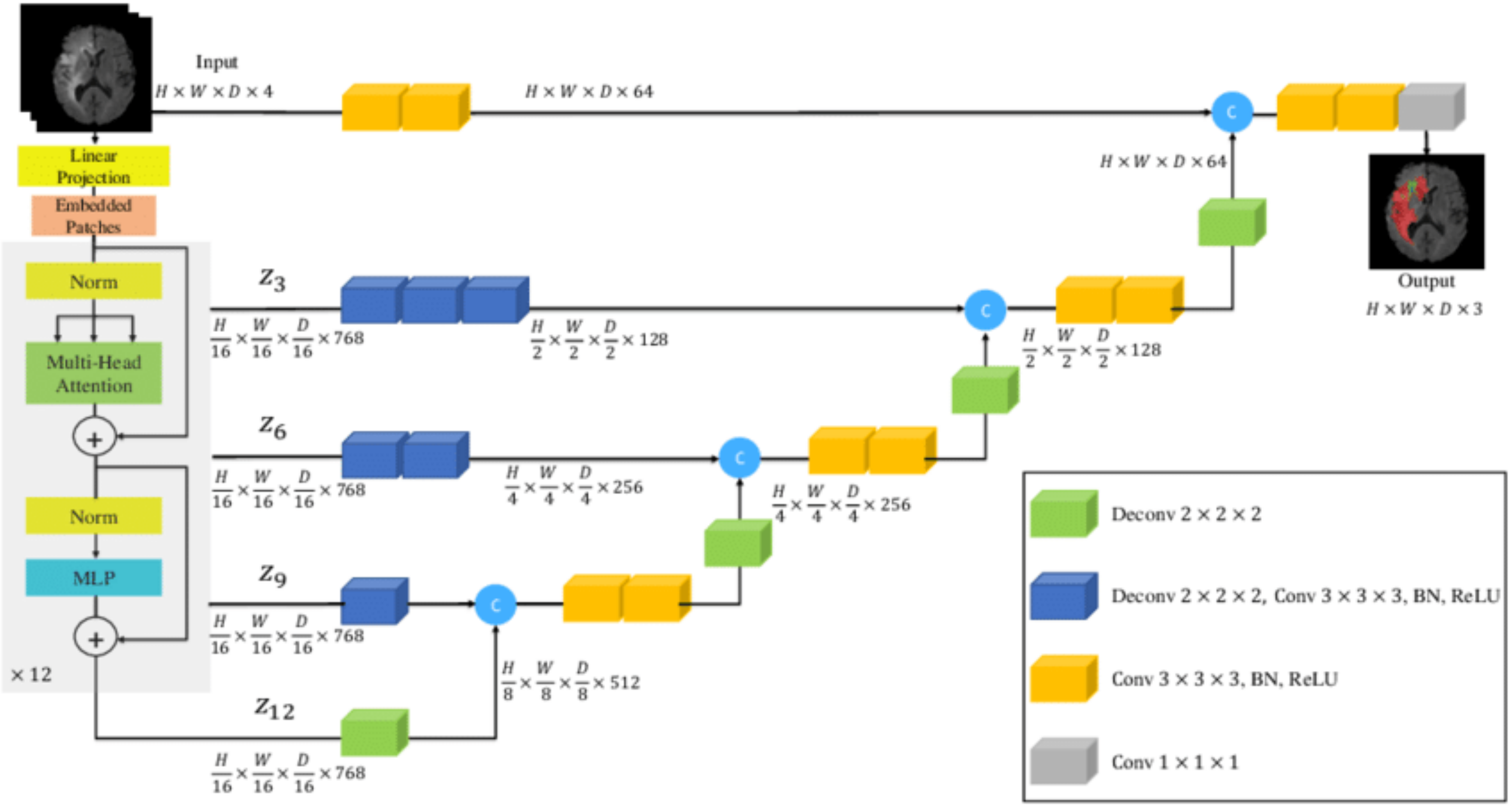
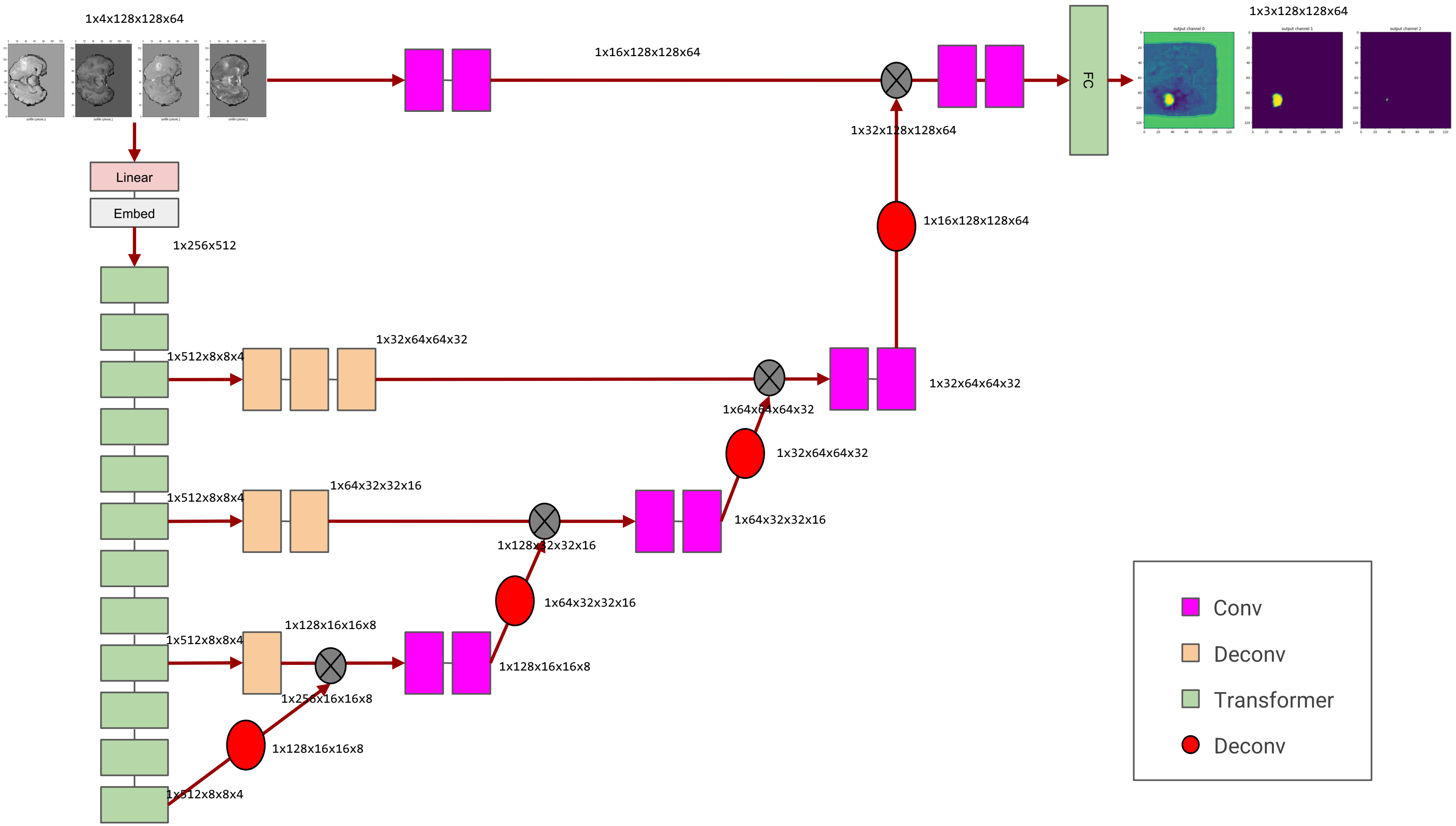



Image source: Hatamizadeh, Ali, et al. "Unetr: Transformers for 3d medical image segmentation." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022.

# Architecture



# Method

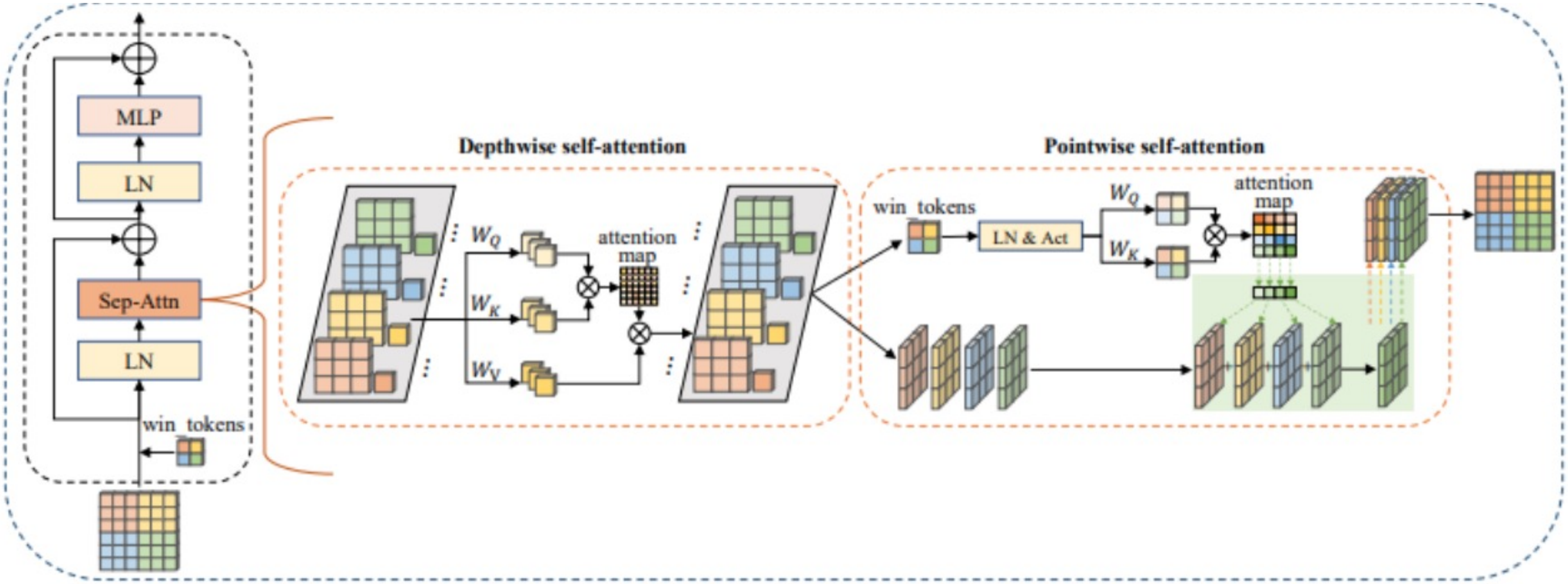



**Fig. 2.** Separable Vision Transformer (SepViT). The top row is the overall hierarchical architecture of SepViT. The bottom row is the SepViT block and the detailed visualization of our depthwise separable self-attention and the window token embedding scheme.

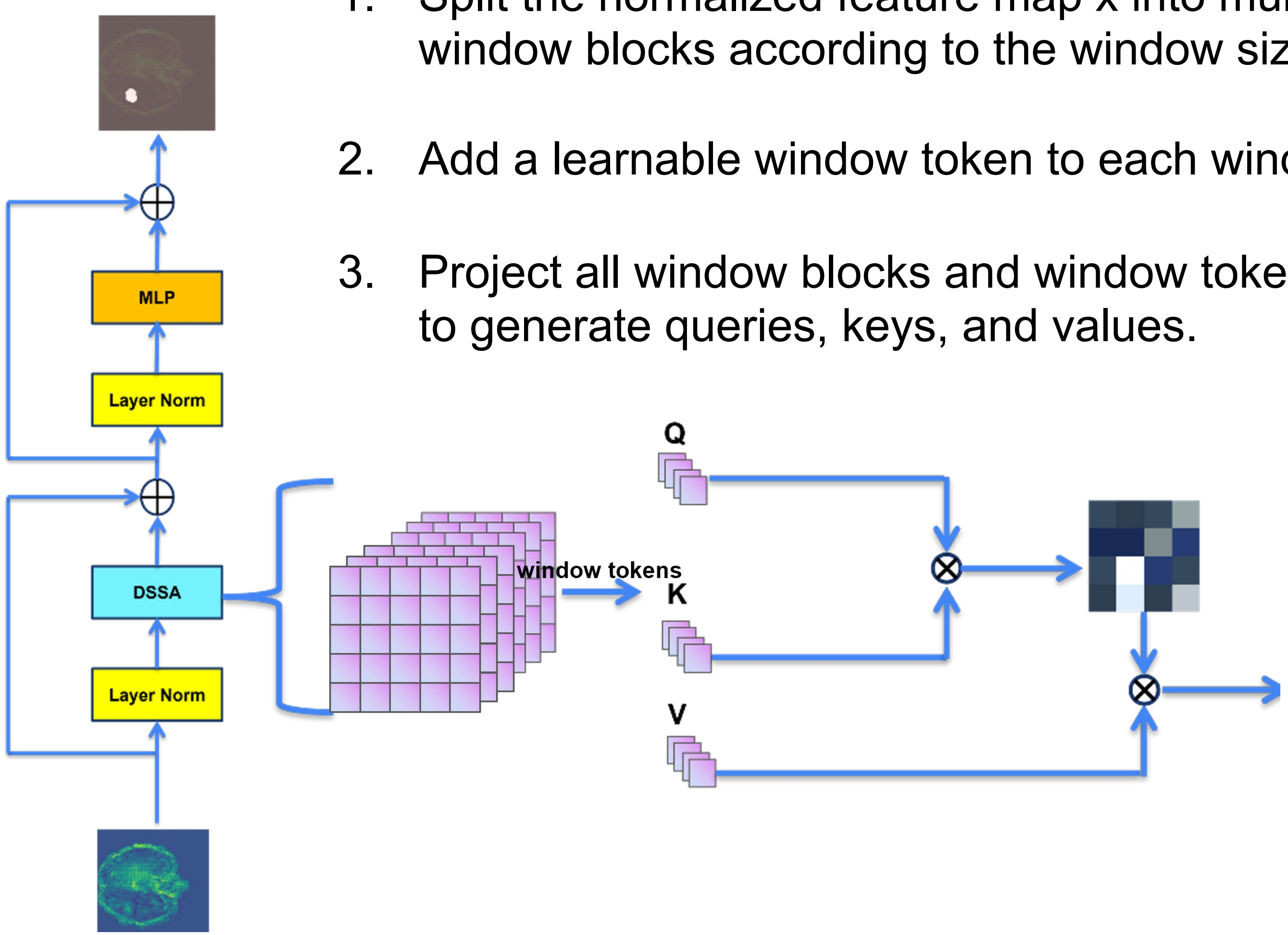
Image source: Li, Wei, et al. "Sepvit: Separable vision transformer." arXiv preprint arXiv:2203.15380 (2022).

# Method

1. Split the normalized feature map x into multiple window blocks according to the window size.

2. Add a learnable window token to each window block.

3. Project all window blocks and window tokens linearly to generate queries, keys, and values.



# Experimental Setup

**Epochs and Dataset:**

Training Duration: 50 epochs

**Computational Resources:**

Computing Platform: Google Colab L4 GPU

**Software and Libraries:**

MONAI v0.7.0, PyTorch v2.2.1+cu121, Numpy v1.25.2, Nibabel v4.0.2, scikit-image v0.19.3, Tensorboard v2.15.2, Transformers v4.38.2, Pillow v9.4.0, tqdm v4.66.2, pandas v2.0.3

**Codebase Information:**

MONAI Revision ID: bfa054b9c3064628a21f4c35bbe3132964e91f43

# Training Details

**Loss Function:**

DiceCELoss = α×Dice Loss + β×Cross-Entropy

**Dice Metric:**

Segmentation accuracy for individual tumor regions and overall.

$$\text{Dice} = \frac{2 \times |X \cap Y|}{|X| + |Y|}$$

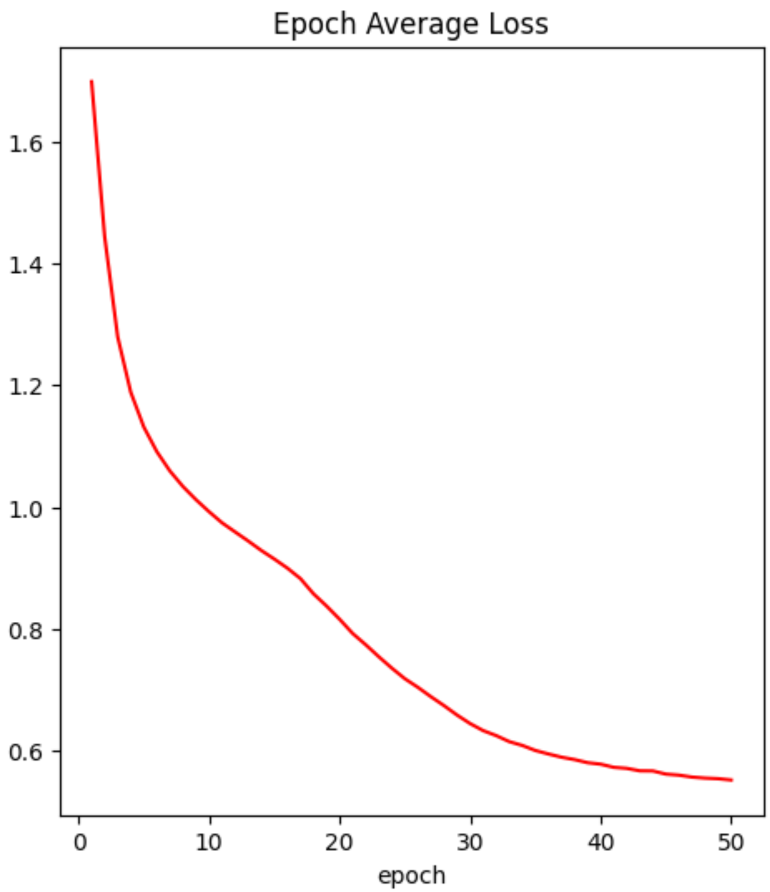
**Cross-Entropy:**

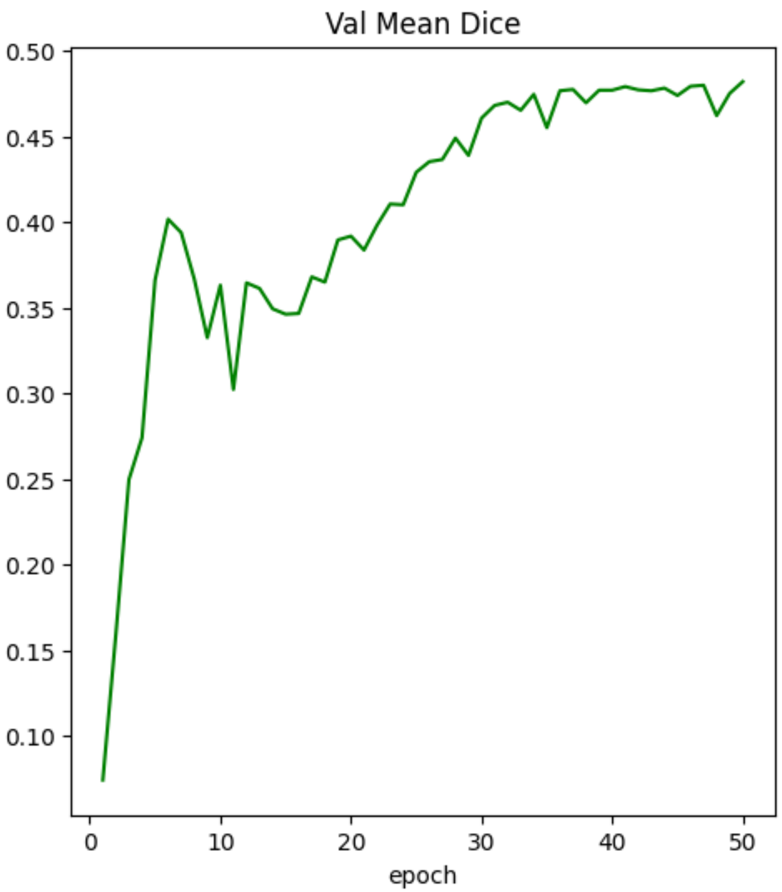
It measures the dissimilarity between the predicted probability distribution and the actual probability distribution.

**Data Augmentation:**

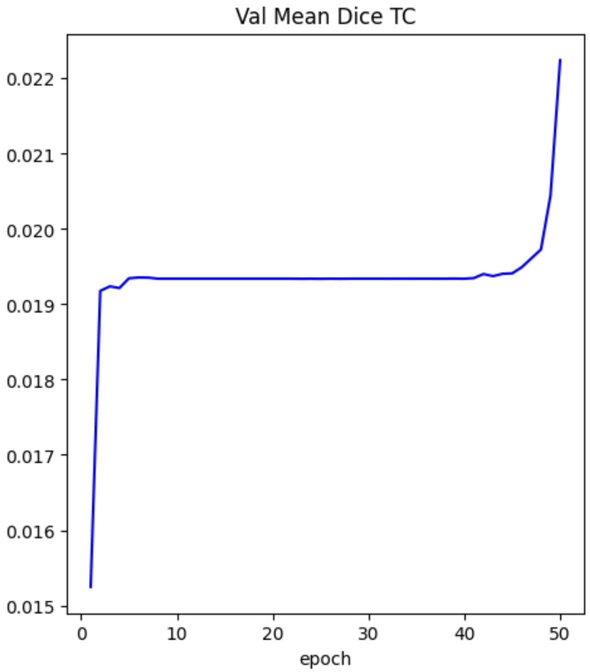
- Random cropping
- Flipping
- Intensity normalization

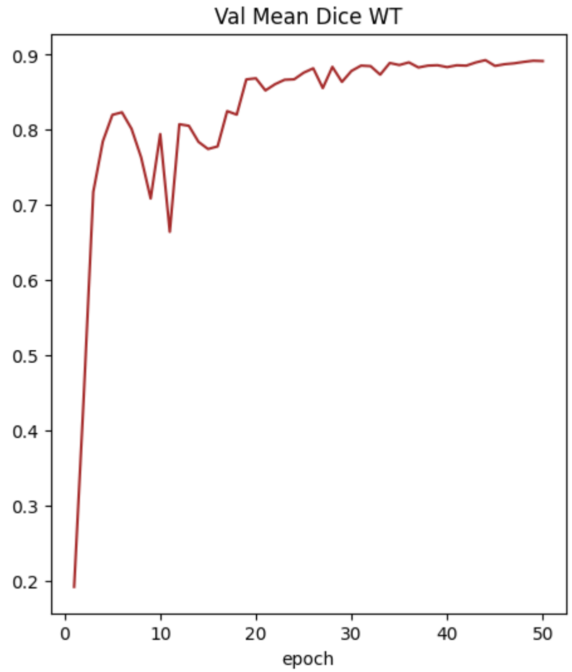
# Experimental Results

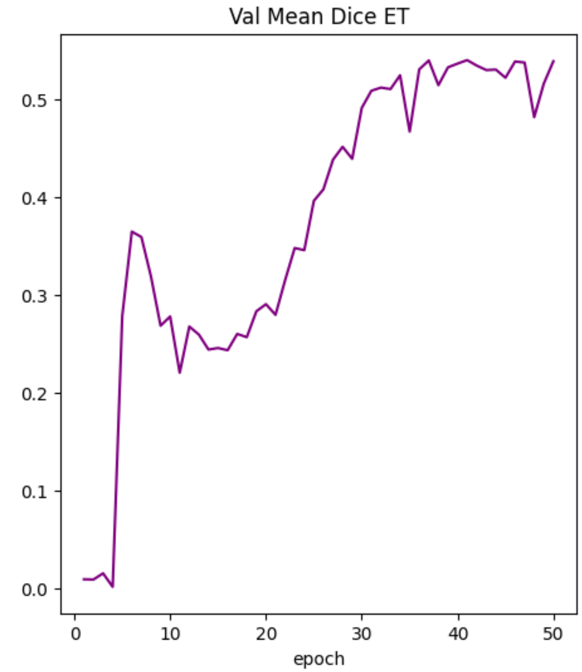


# Experimental Results

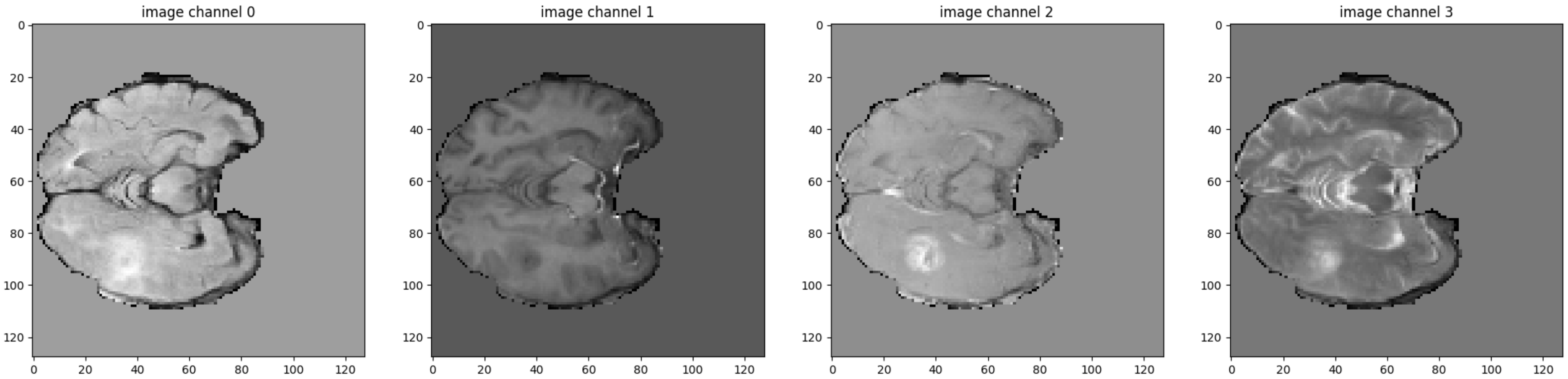


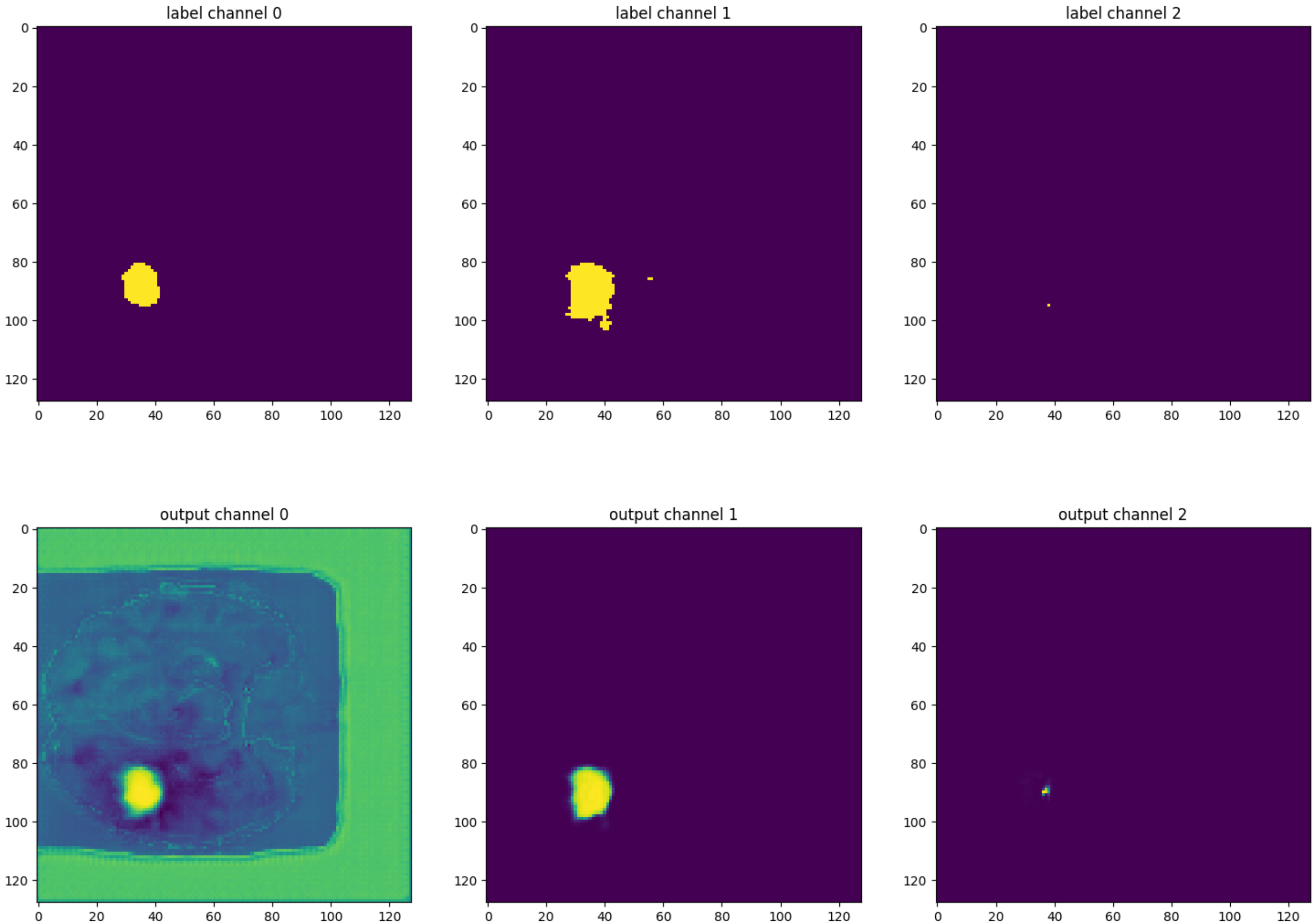

**Tumor Core (TC):** Comprises enhancing and necrotic tumor tissues, excluding edema.
**Whole Tumor (WT):** Includes all tumor tissues.
**Enhancing Tumor (ET):** Refers to the highly active tumor regions in contrast-enhanced scans.

# Demos




Multimodal multisite MRI data (FLAIR, T1w, T1gd, T2w)

# Demos


label channel 0
label channel 1
label channel 2
output channel 0
output channel 1
output channel 2
0
20
40
60
80
100
120

# Github Link

https://github.com/zh249/COSC-5470-01

Thank you!

# QUESTIONS?

# Reference


Ma, Jun, et al. "Segment anything in medical images." Nature Communications 15.1 (2024): 654.

Hatamizadeh, Ali, et al. "Unetr: Transformers for 3d medical image segmentation." Proceedings of the IEEE/CVF winter conference on applications of computer vision. 2022.

Li, Wei, et al. "Sepvit: Separable vision transformer." arXiv preprint arXiv:2203.15380 (2022).

Menze, Bjoern H., et al. "The multimodal brain tumor image segmentation benchmark (BRATS)." IEEE transactions on medical imaging 34.10 (2014): 1993-2024.