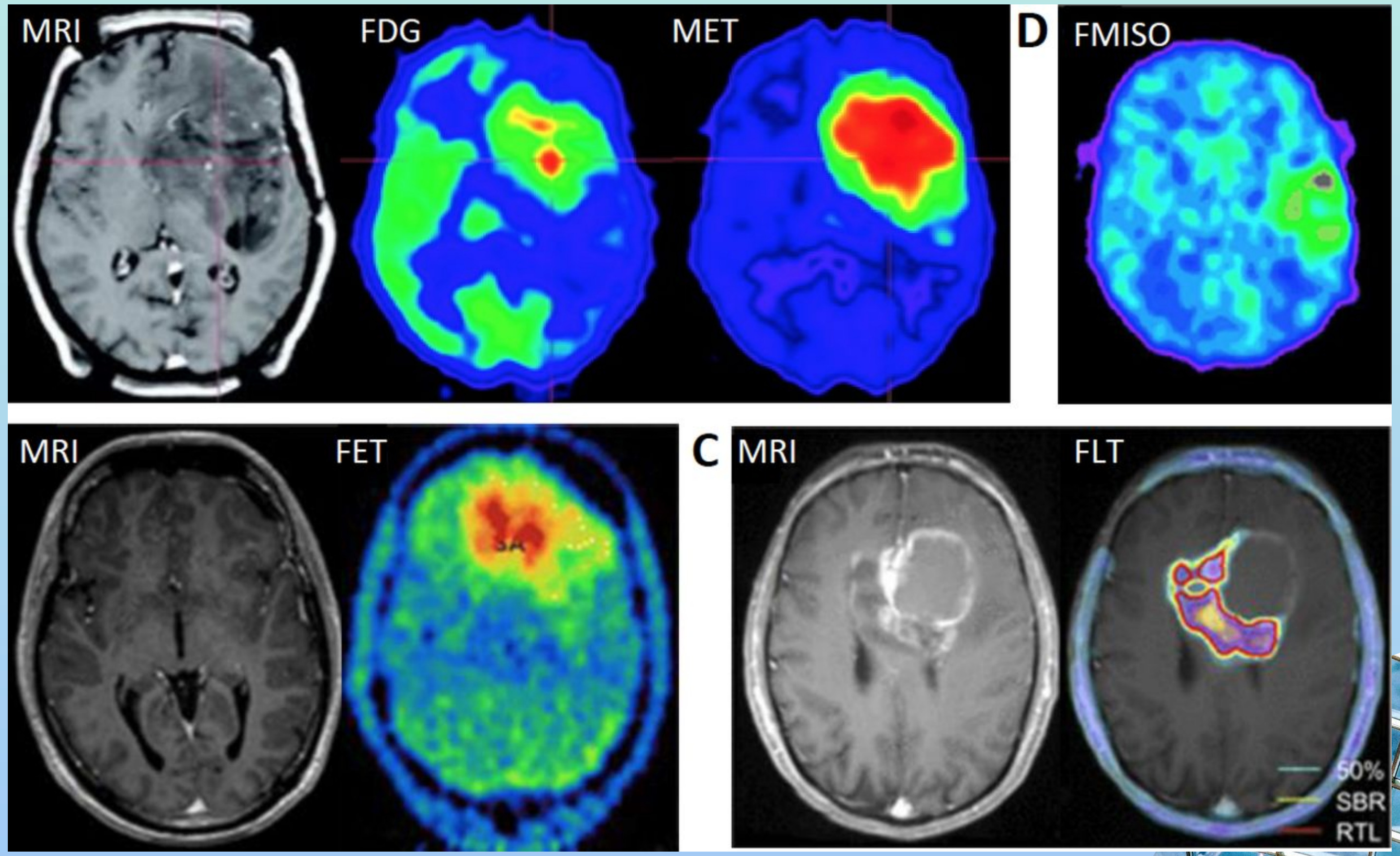
Swin-Unet: Unet-like Transformer for Brain Tumor Segmentation

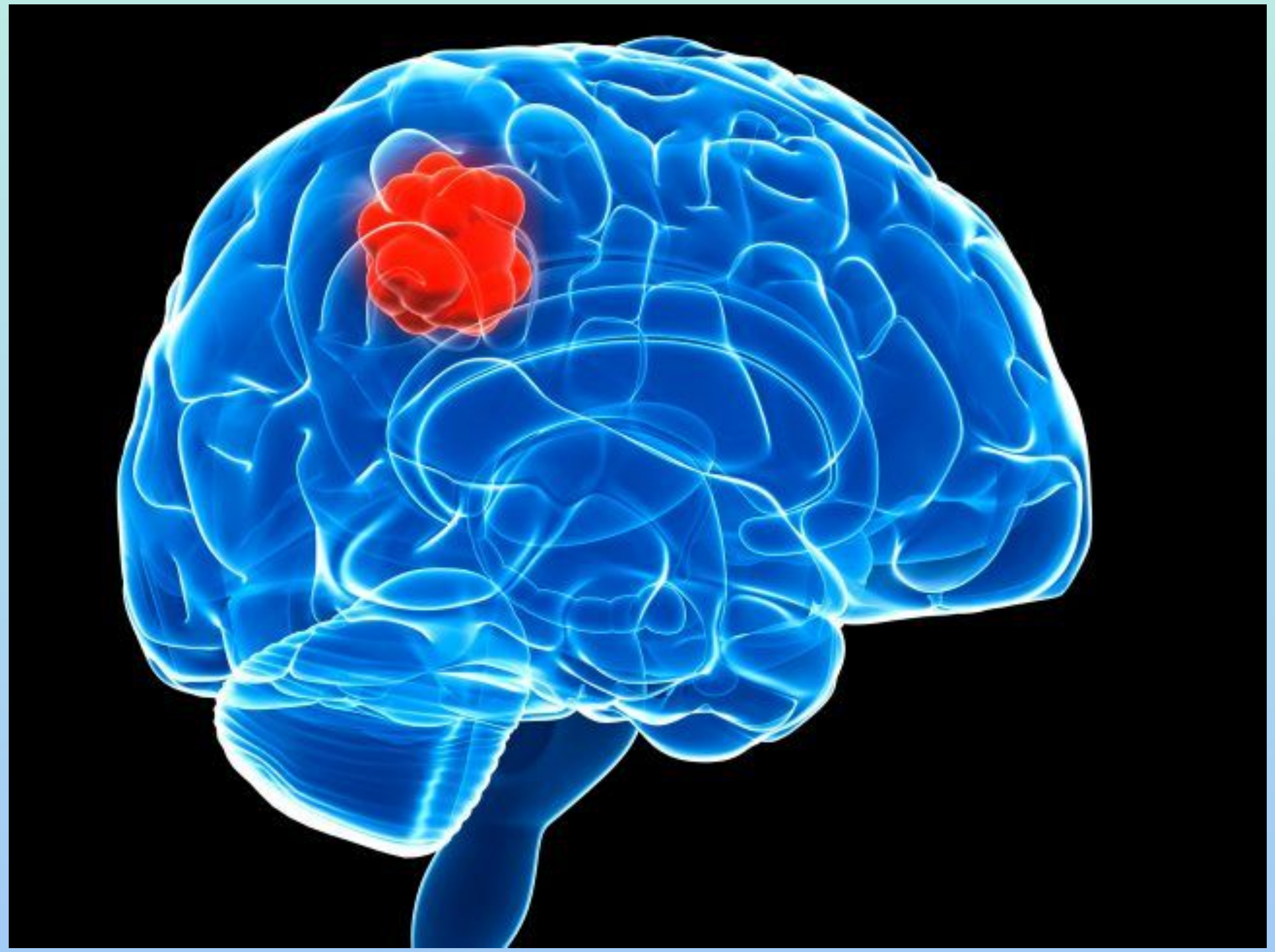


- Omar Ayman
- Muhammed Essam
- Ahmed Abdelgawad
- Yahya Mohamed

Problem Definition

Brain Tumor - Importance of Early Detection

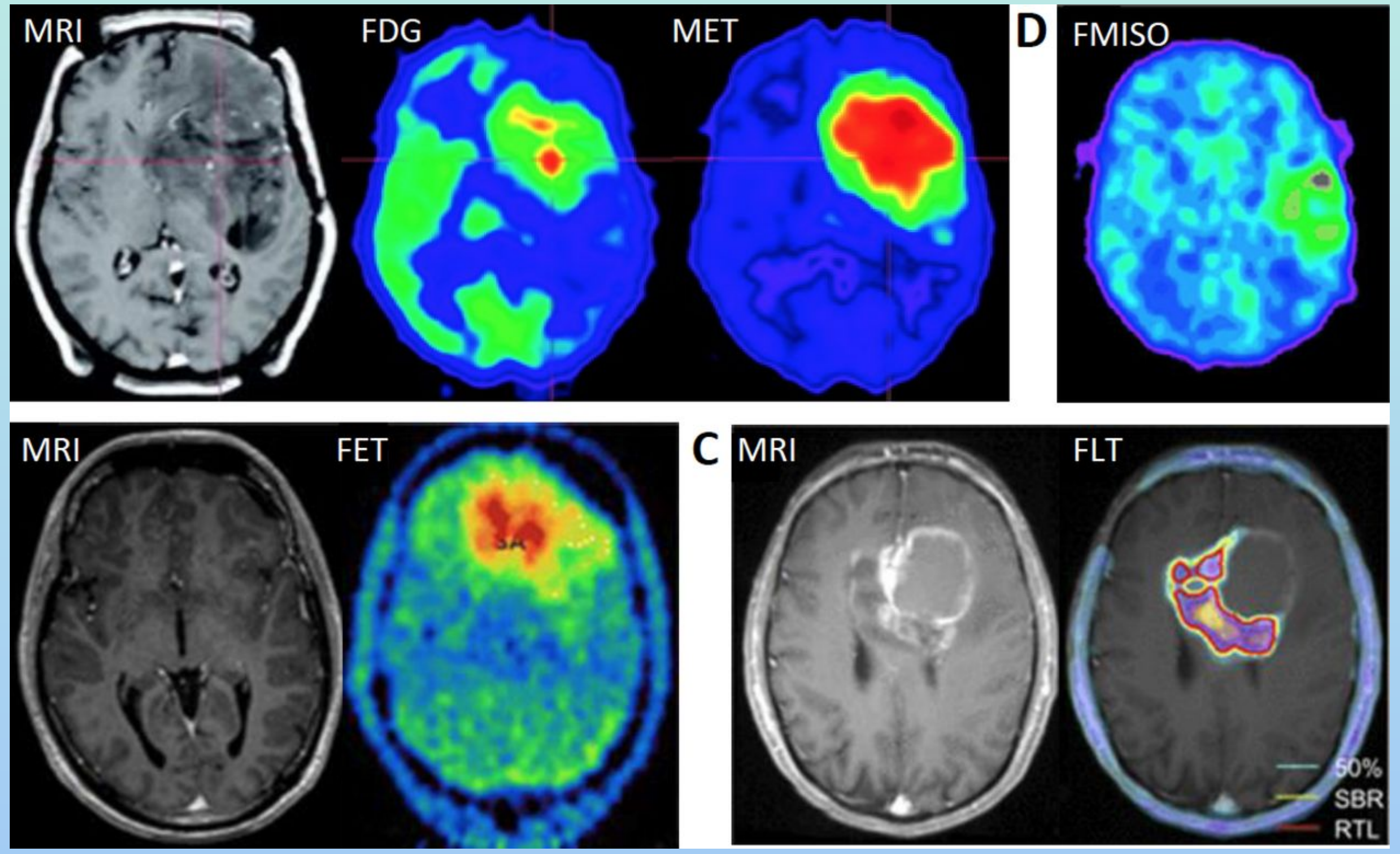
- Automated and Precise Segmentation
- Critical for Effective Treatment
- Life-Threatening Nature
- Depending on type, size, and location.
- Early detection is critical for effective treatment planning.
- Segmentation Focus



Problem Definition

Brain Tumor - Task

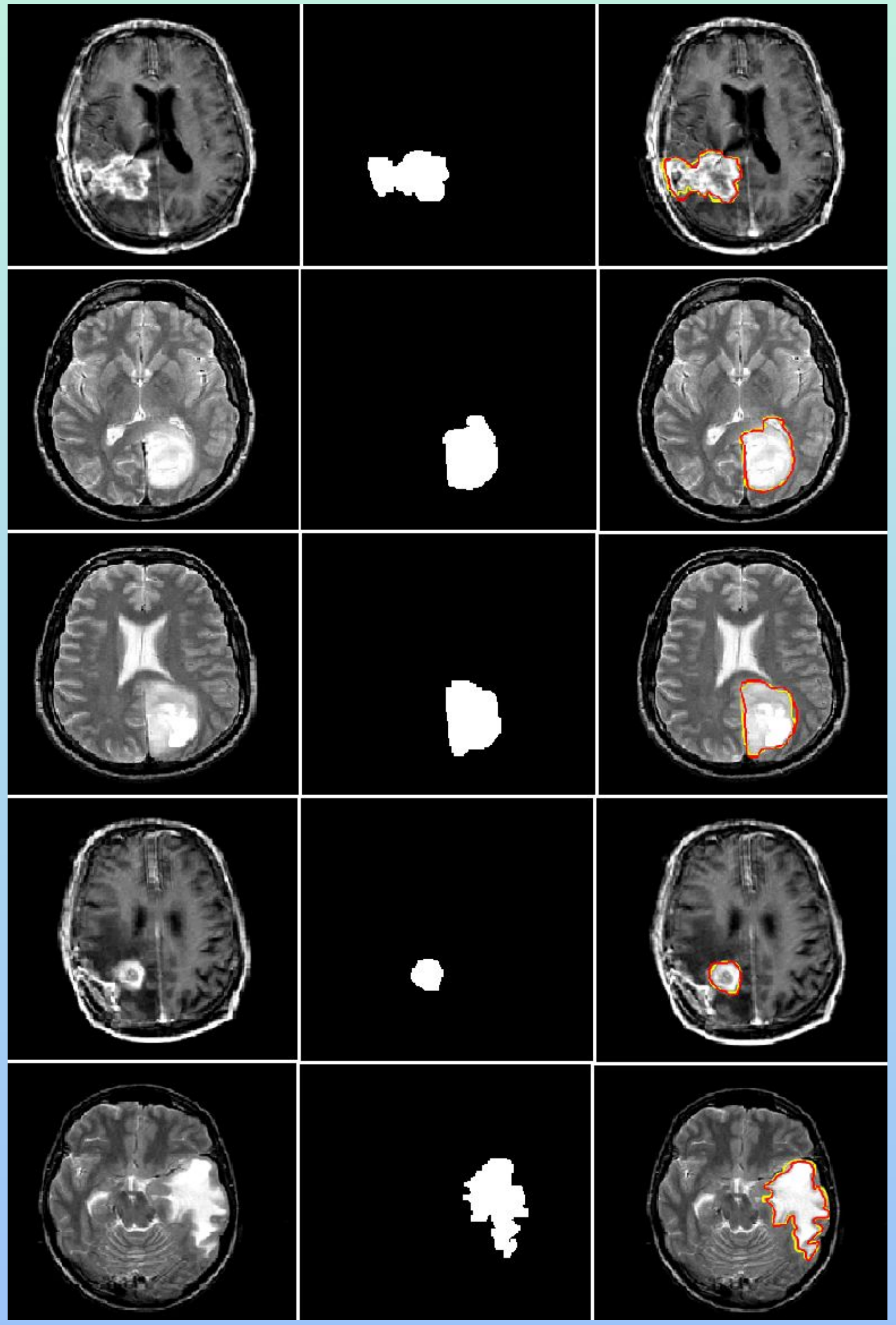
- Focus on the segmentation of brain tumors from MRI images using Deep Learning Techniques.
- Targeting automated and precise segmentation to assist in diagnosis



Dataset Description

Source

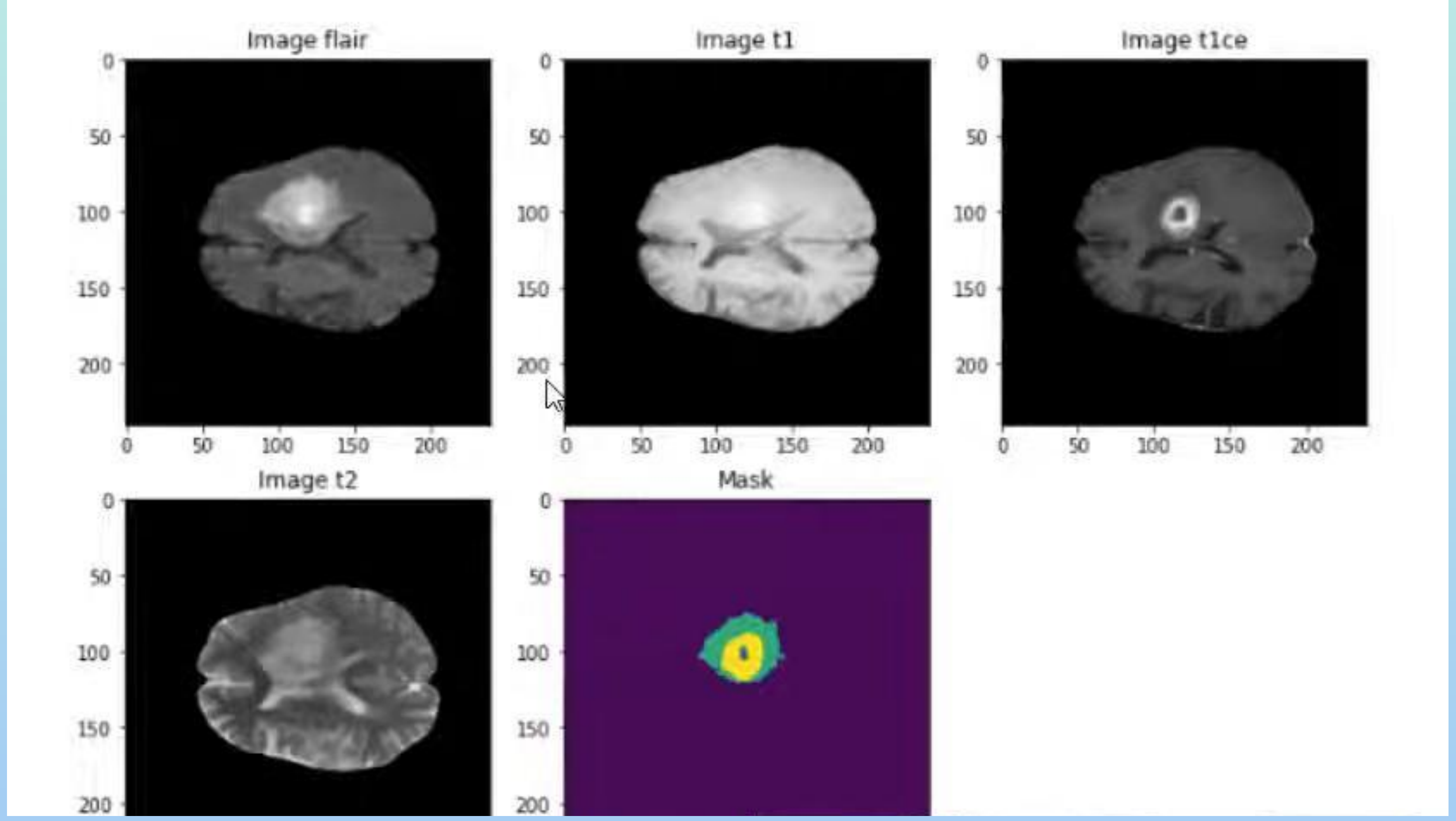
- Brain MRI segmentation dataset.
- https://www.kaggle.com/datasets/mateuszbuda/lgg-mri-segmentation/data.
- The images were obtained from The Cancer Imaging Archive (TCIA).
- They correspond to 110 patients included in The Cancer Genome Atlas (TCGA).



Dataset Description

Dataset

- Multimodal scans available as NIfTI files (.nii.gz)
- Four channels of information:
  - Native (T1).
  - Post-contrast T1-weighted (T1CE).
  - T2-weighted (T2).
  - T2 Fluid Attenuated Inversion Recovery (FLAIR).



Dataset Description

- All the dataset have been segmented manually and were approved by experienced neuro-radiologists.
- Annotations (Labels):

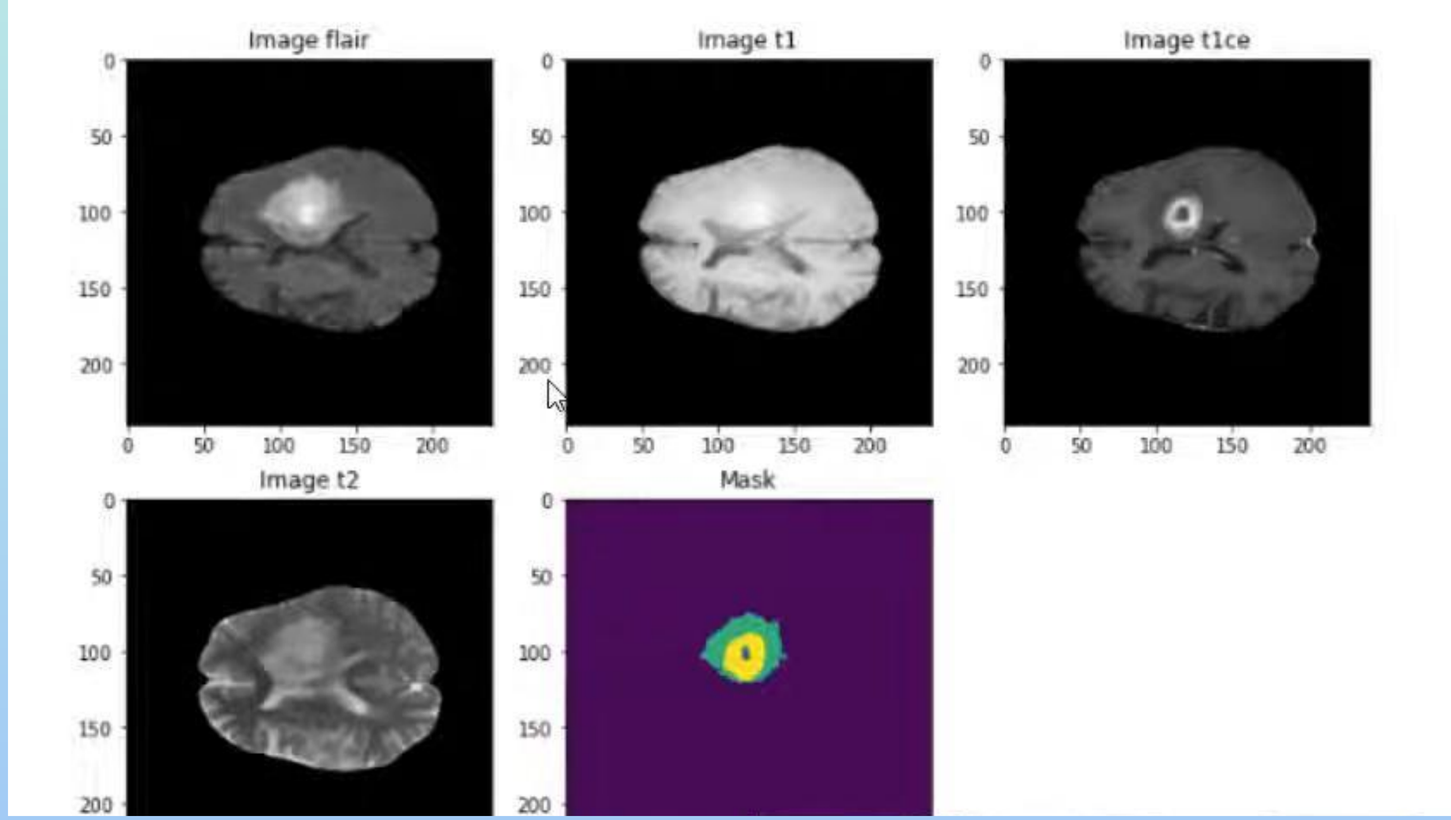
  - Label 0: unlabeled volume.

  - Label 1: Necrotic non-enhancing tumor core (NCR/NET).

  - Label 2: Peritumoral Edema (ED).

  - Label 3: Missing (No pixels for label 3).

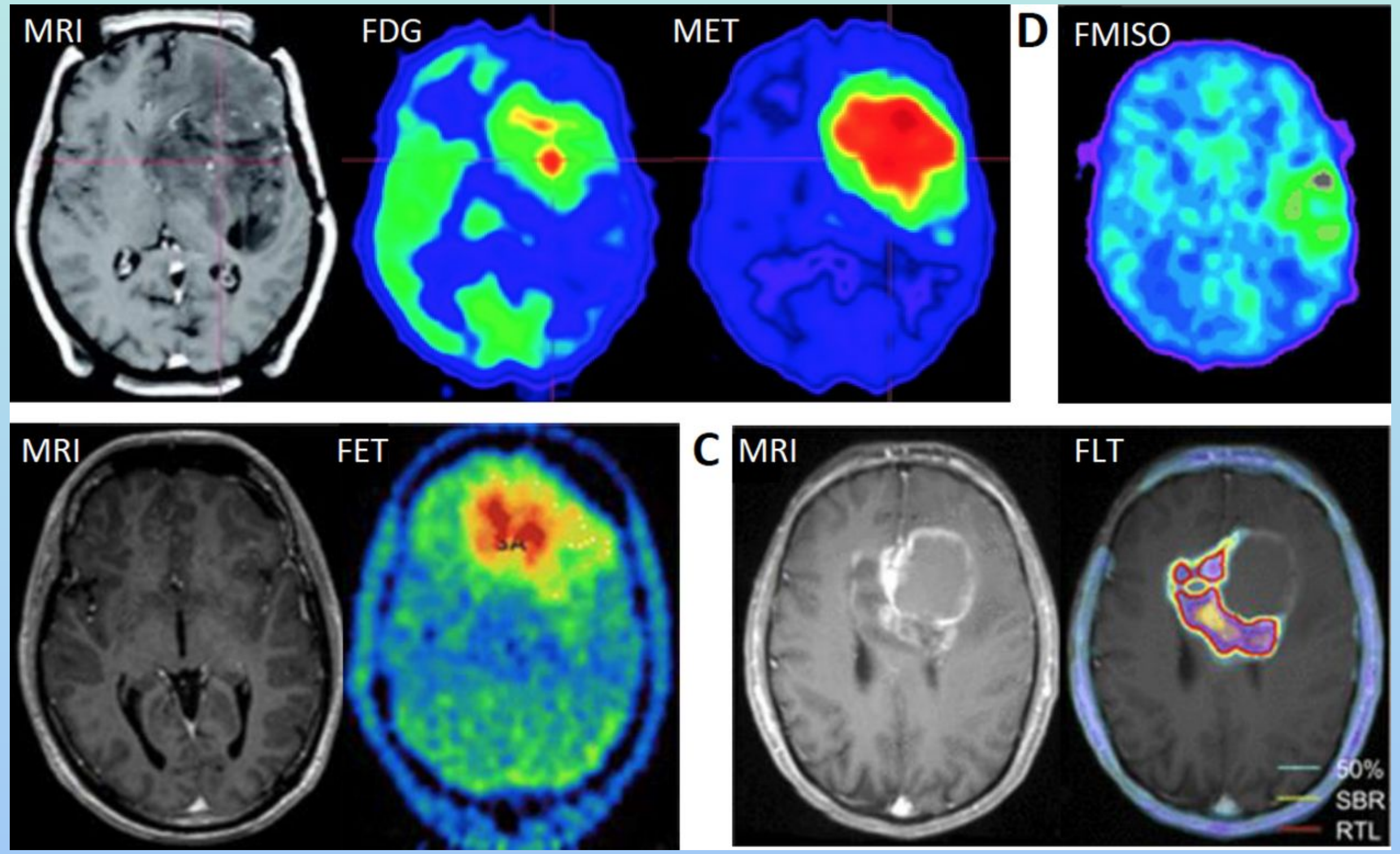
  - Label 4: GD-enhancing tumor (ET).



Literature Review

Previous Studies

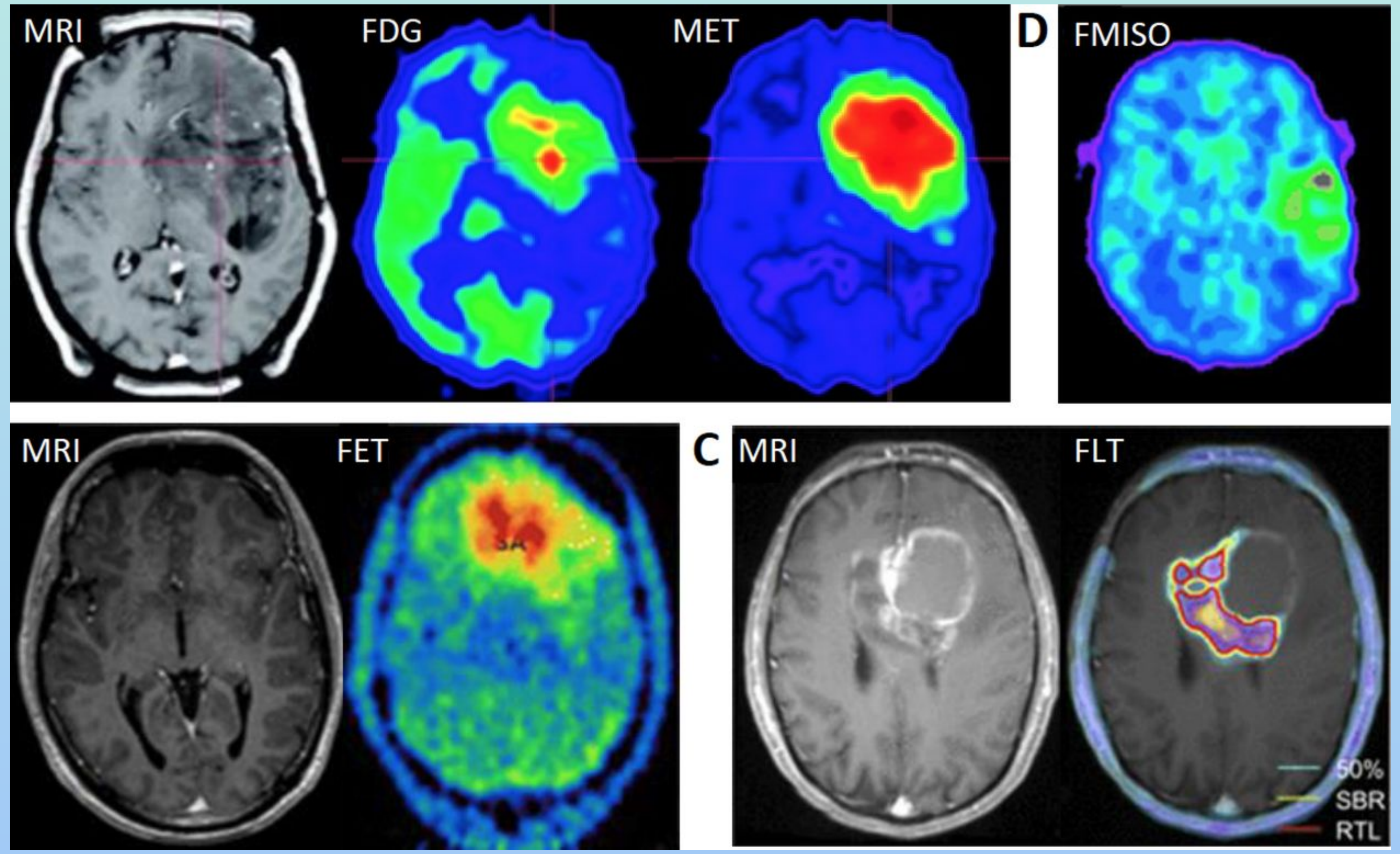
- Earlier approaches primarily used manual segmentation:
  - Time-consuming.
  - Error-prone.
- More recent studies employ various CNN architectures like: U-Net, Attention U-Net, and Residual U-Net.



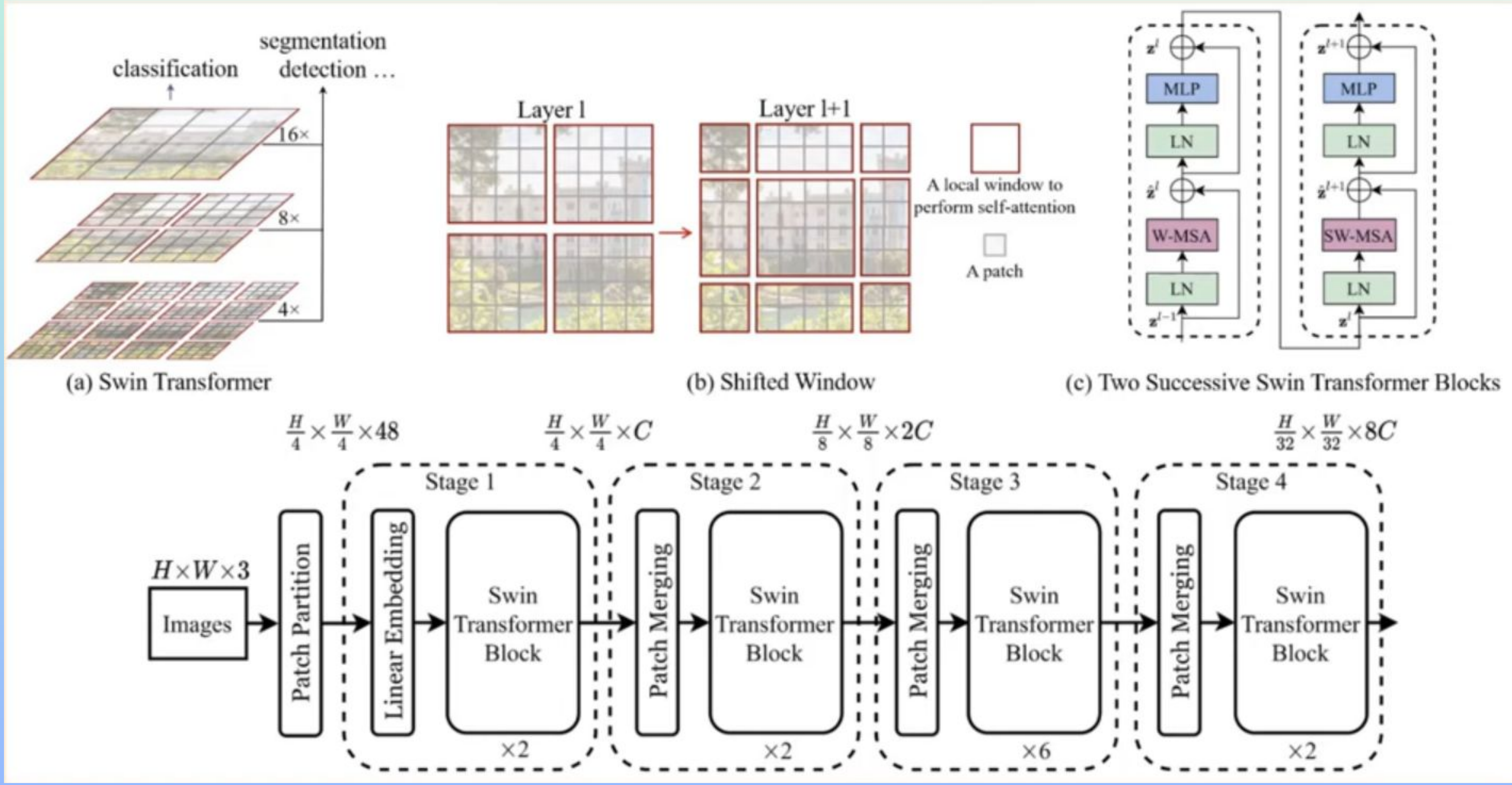
Literature Review

Our Approach

- Comparing different CNN architectures and assess their performance against a Swin Transformer model.
- Identify the most effective method for brain tumor segmentation.

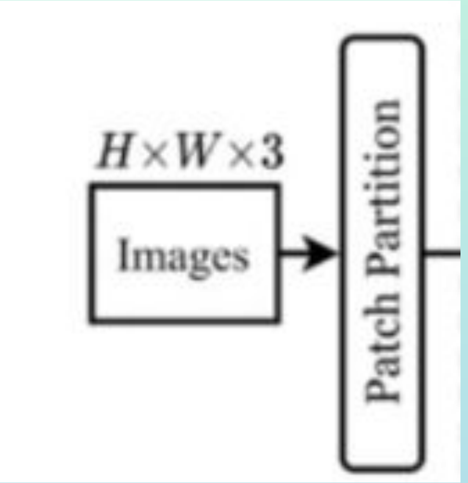


Our Approach



Our Approach

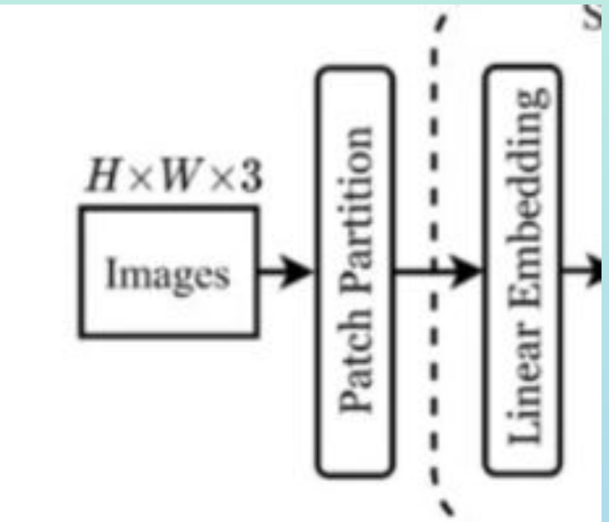
1- Patification



- Image size 224 x 224
- Patch size: 4 x 4
- Number of patches per row: 56
- Number of patches per column: 56
- Total patches: 3136
- Every patch have 48 entries or values.

Our Approach

2- Linear Embedding



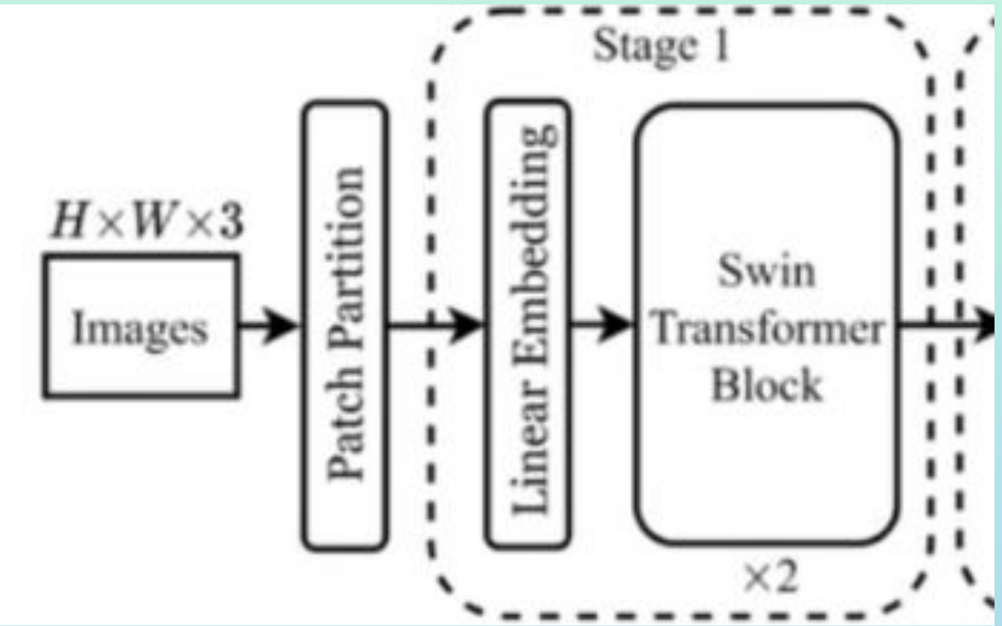
- A technique used to convert the input data into the format that can be processed by a transformer.
- Images are made of pixels.
- Transformers work with a sequence of tokens.
- Linear Embedding is a process for converting image pixels into numerical representation in a vector form.

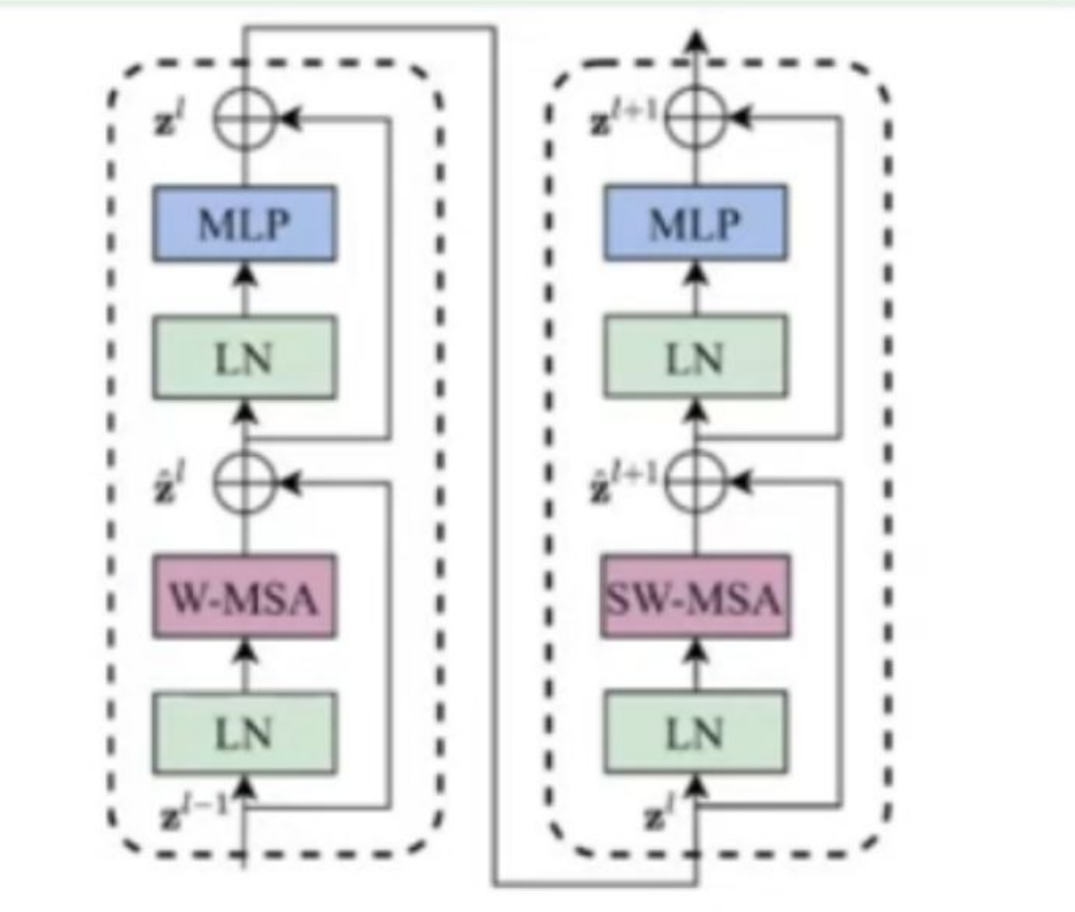
```
 Patches shape: torch.Size([1, 3136, 96])
First patch embedding vector: tensor([ 0.0872,  0.0503, -0.1719,  0.1464,  0.1546, -0.0358,  0.0016, -0.3824,
        -0.2313, -0.1955,  0.1172, -0.0174, -0.1295, -0.0316,  0.0104, -0.0479,
         0.0975, -0.0499, -0.0125, -0.1404, -0.2292,  0.1697, -0.2276, -0.0672,
         0.0015, -0.0895,  0.1411, -0.0210,  0.0128,  0.1669,  0.0921, -0.1980,
        -0.2184,  0.0075,  0.1579,  0.0842,  0.0635, -0.0160,  0.2636,  0.1215,
         0.1866, -0.2069,  0.0335,  0.0802,  0.1717, -0.1648,  0.2139,  0.1257,
        -0.1655, -0.0129, -0.1318, -0.1114,  0.1215, -0.0470,  0.0056, -0.2260,
        -0.0460,  0.0295, -0.0118, -0.4155,  0.0999, -0.0171,  0.1431, -0.0716,
        -0.0708,  0.0945, -0.1236,  0.0800, -0.1344,  0.0319,  0.0184,  0.1027,
         0.0112, -0.0439, -0.2414,  0.0518,  0.1925, -0.0069,  0.0818,  0.0732,
         0.1824,  0.1271,  0.0092,  0.0964, -0.0937,  0.1673, -0.2281, -0.0448,
        -0.1030, -0.0190,  0.1538, -0.0324,  0.2129,  0.1216,  0.0420, -0.2114],
```

Our Approach

3- Swin Transformer Block

- Consists of two sub-units
- Each unit consists of a normalization layer (LN) followed by an attention module which is different in each unit.
- after the attention modules there is another normalization layer followed by a multi-layer perceptron layer.



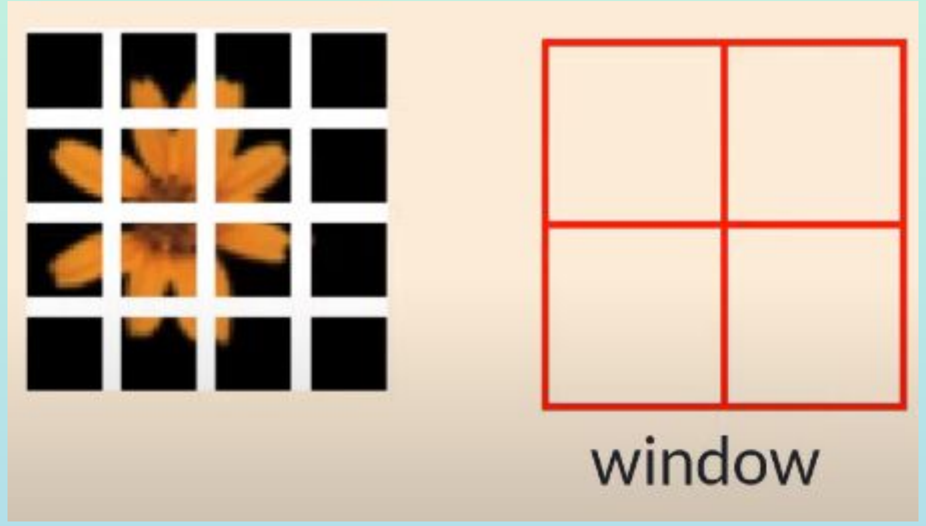


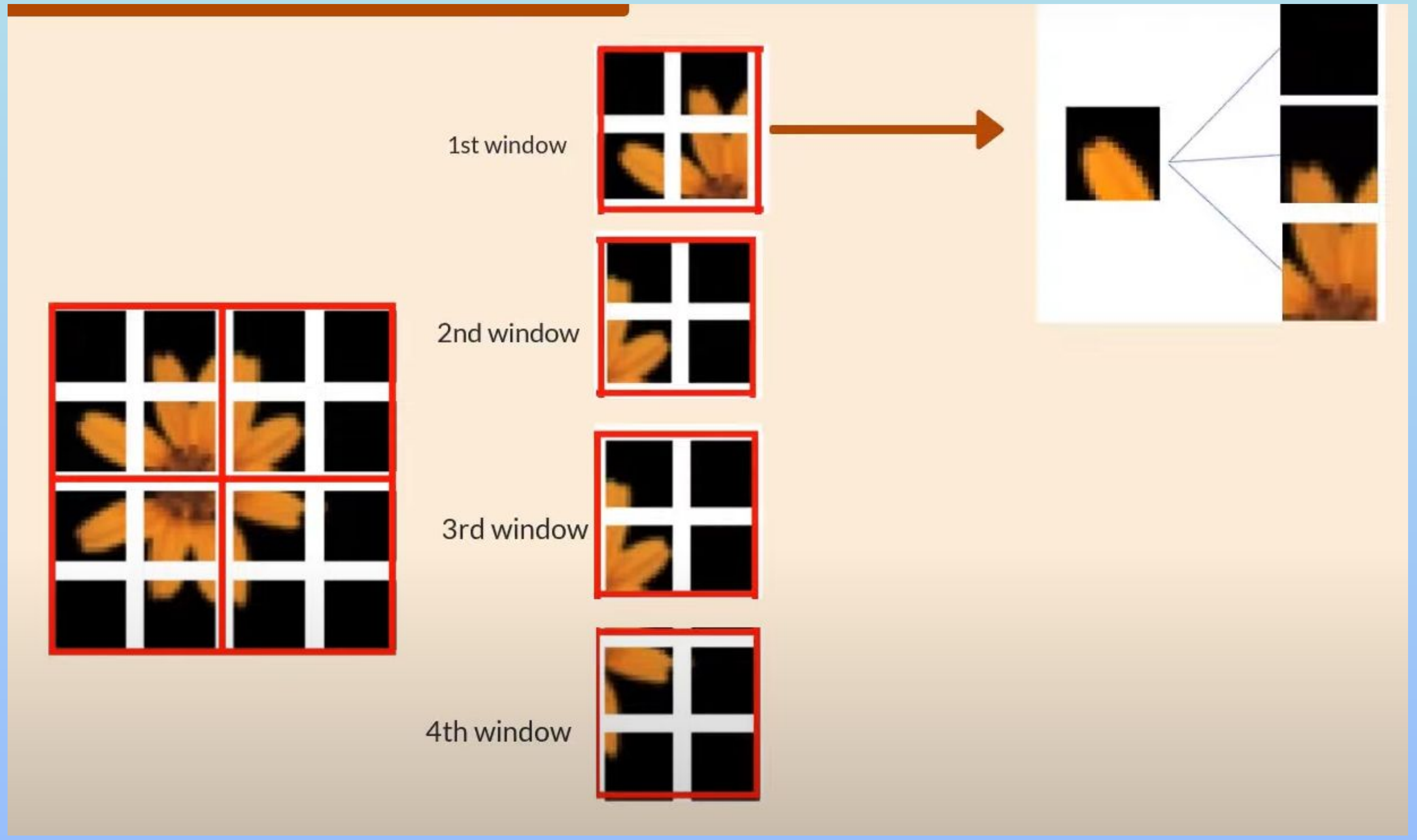
) Two Successive Swin Transformer Blocks

Our Approach

3- Swin Transformer Block (Window-based MSA)

- The attention is done on each one of the batches of each window.
- How will we see the connection between the batches?

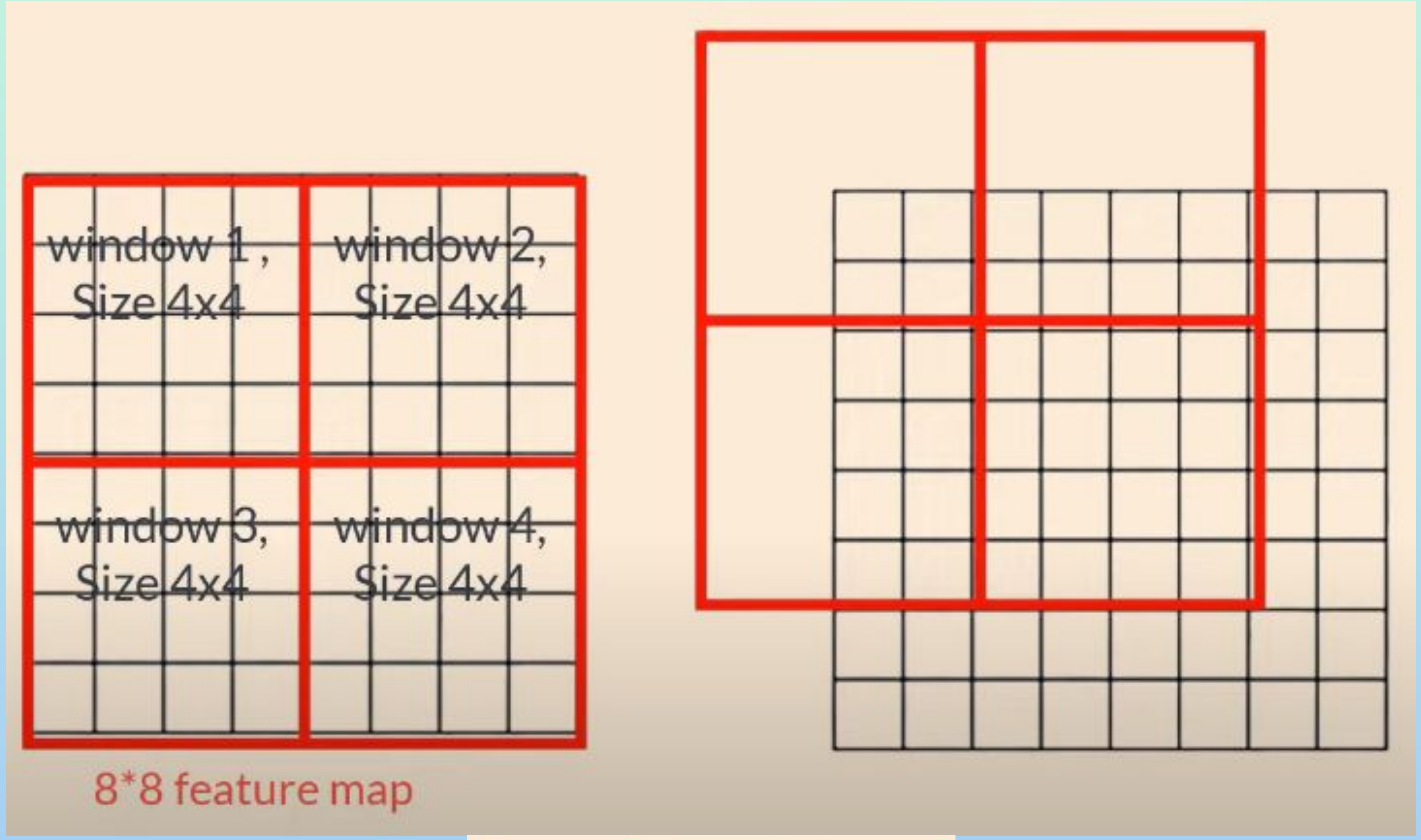


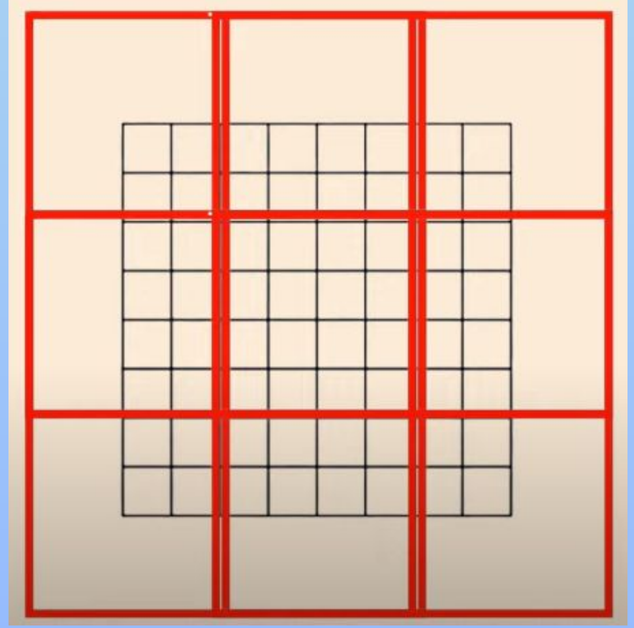


Our Approach

3- Swin Transformer Block
(Shifted Window-based MSA)

- Provides the connection between widows while maintaining its computational efficiency.
- In this approach, we displace the window from the regularly partitioned window.
- Here the shift size is 2
- This will result in 9 windows
- Then Cyclic shift will be used to reassemble these 9 windows back to 4 windows

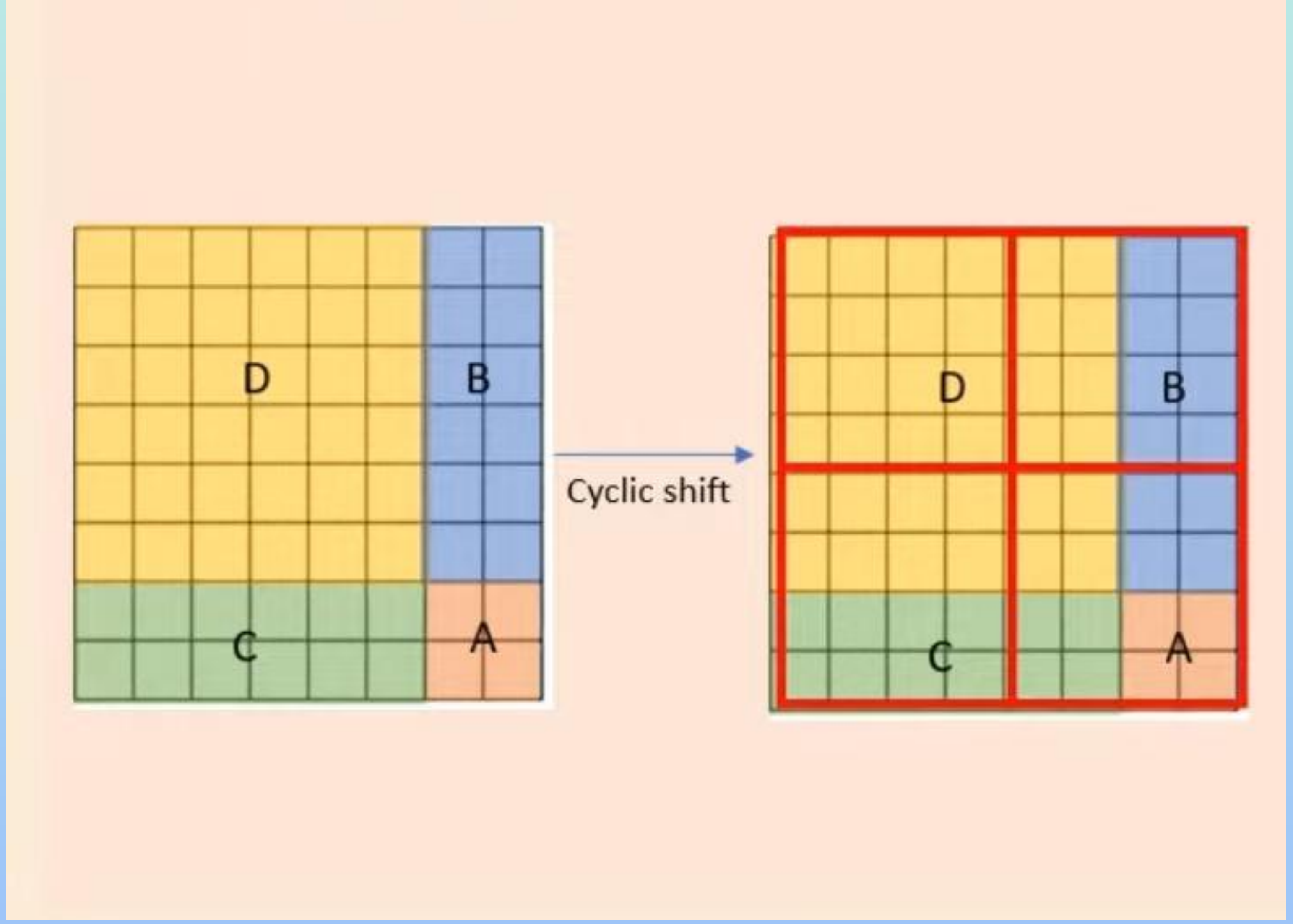




Our Approach

Cyclic Shift

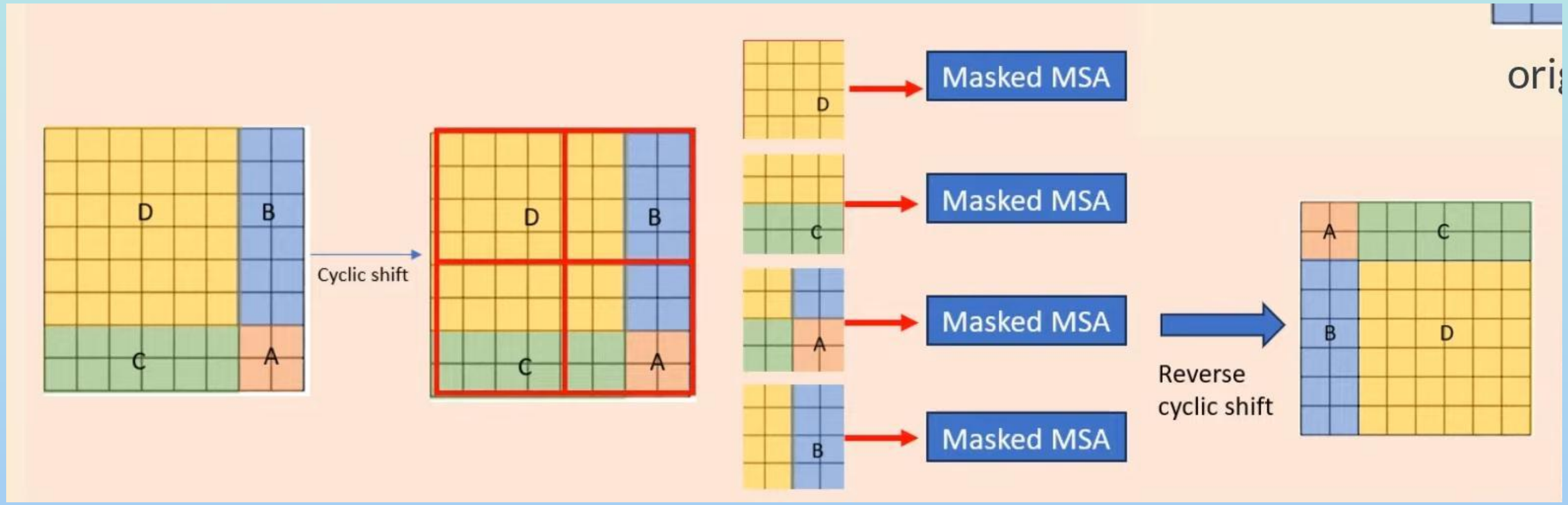
- Batches at are D are shifted 2 times up and 2 times left.
- Similarly, we need to shift the are A 2 times up and 2 times left, but there’s no space -> cyclic shifting.
- The same applies for area B and C.



Our Approach

Cyclic Shift

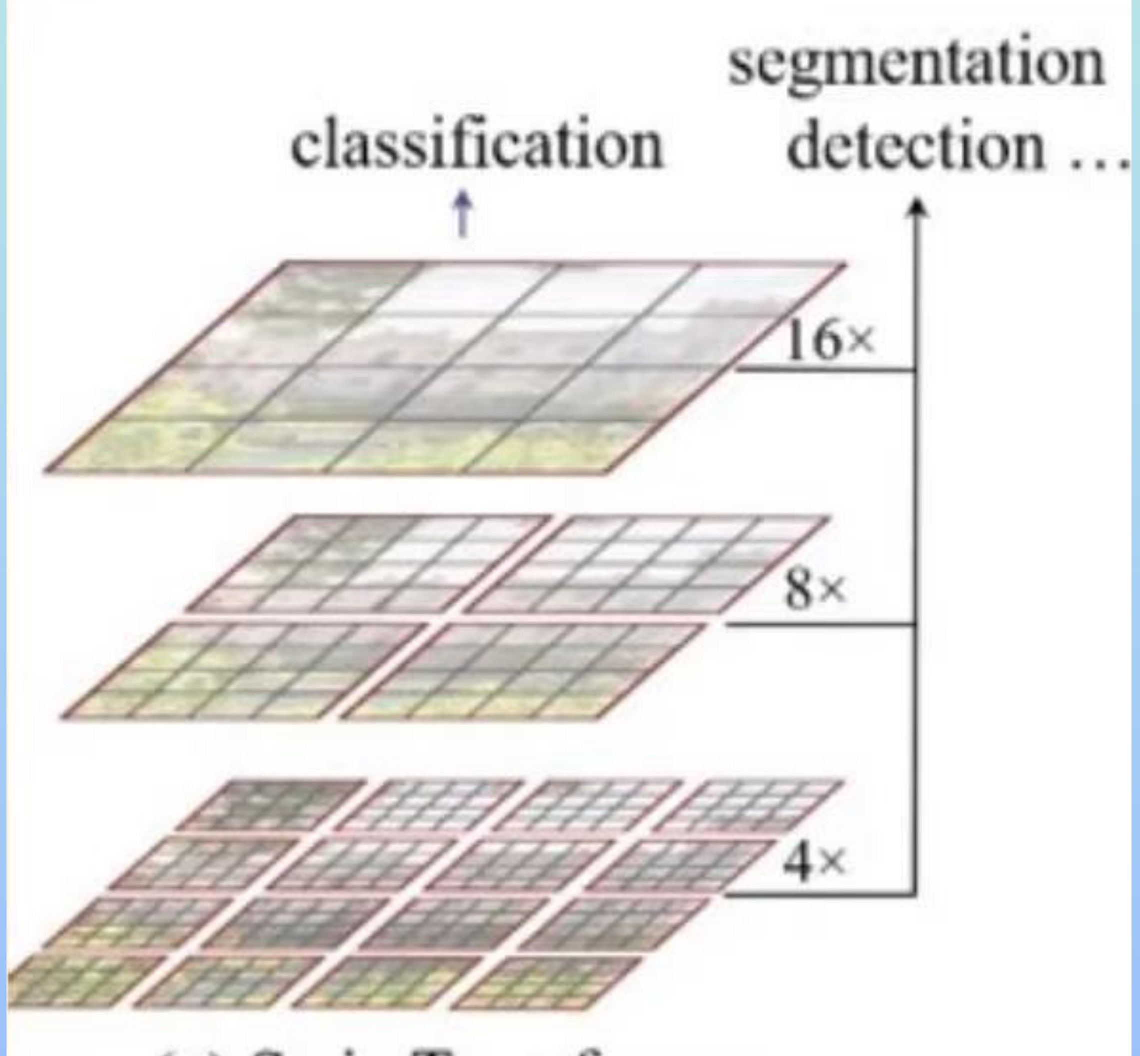
- Now, we can define the window.
- We can see different batches in the same window.



Our Approach

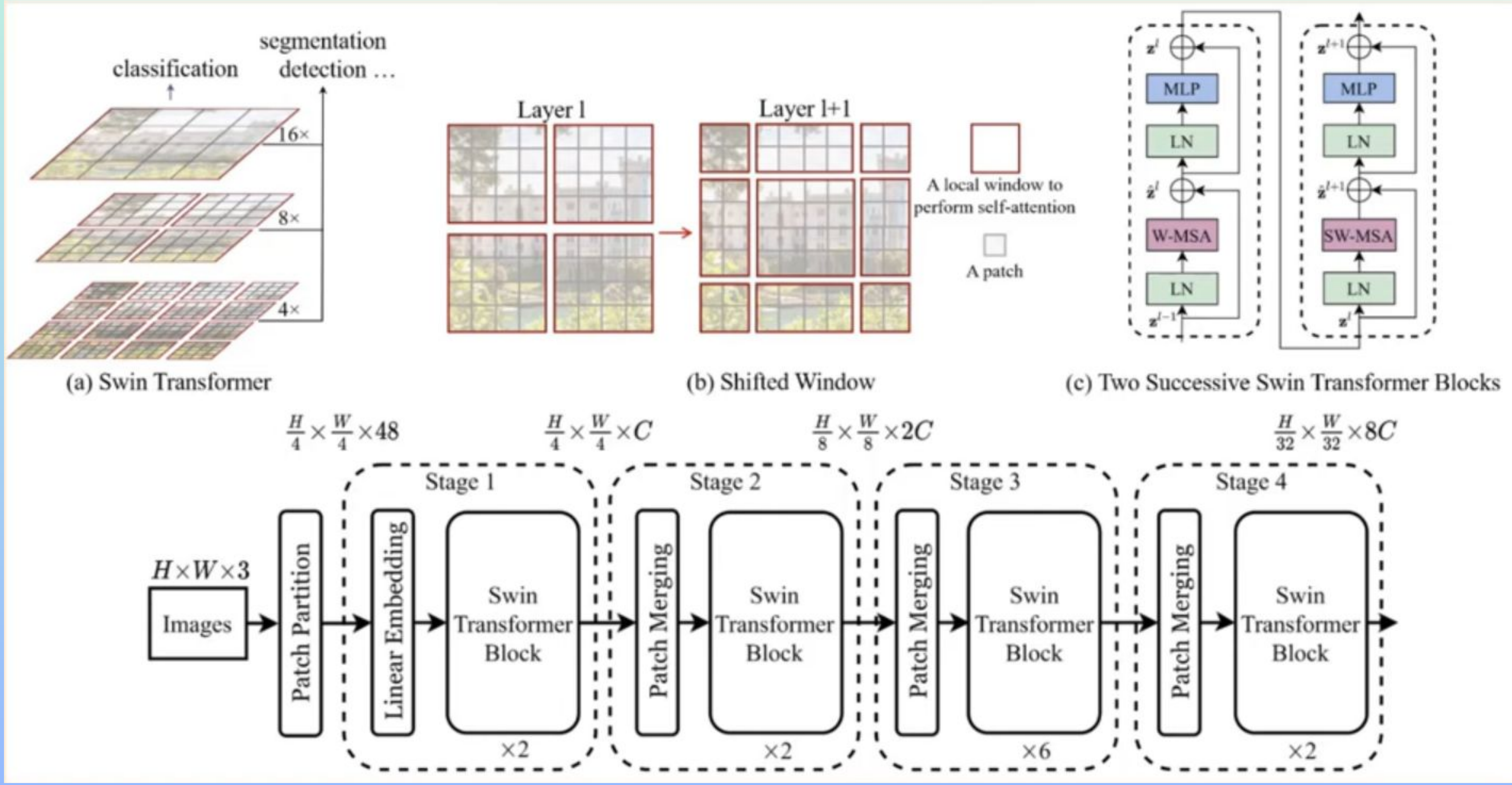
4- Patch Merging

- To capture global information effectively, swin transformer merges the adjacent batches to capture global information effectively, instead of dealing with each batch separately.
- This process occurs in multiple stages.
- Each stage consist of a transformer block followed by a patch merging operation.
- This layer merges 4 patches.
- every merge, both the hight and wedth are reduced by a factor of 2



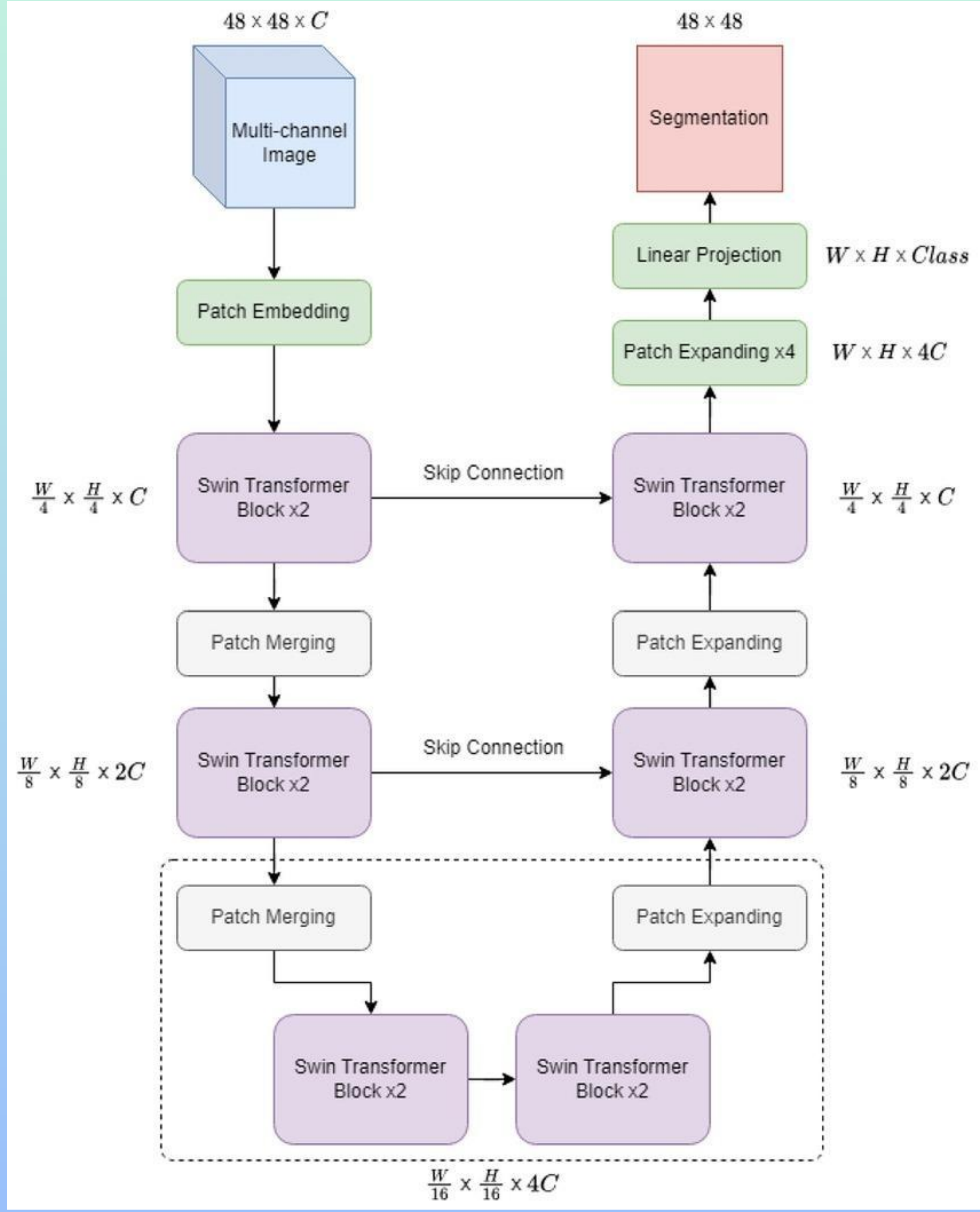
(a) Swin Transformer

Our Approach



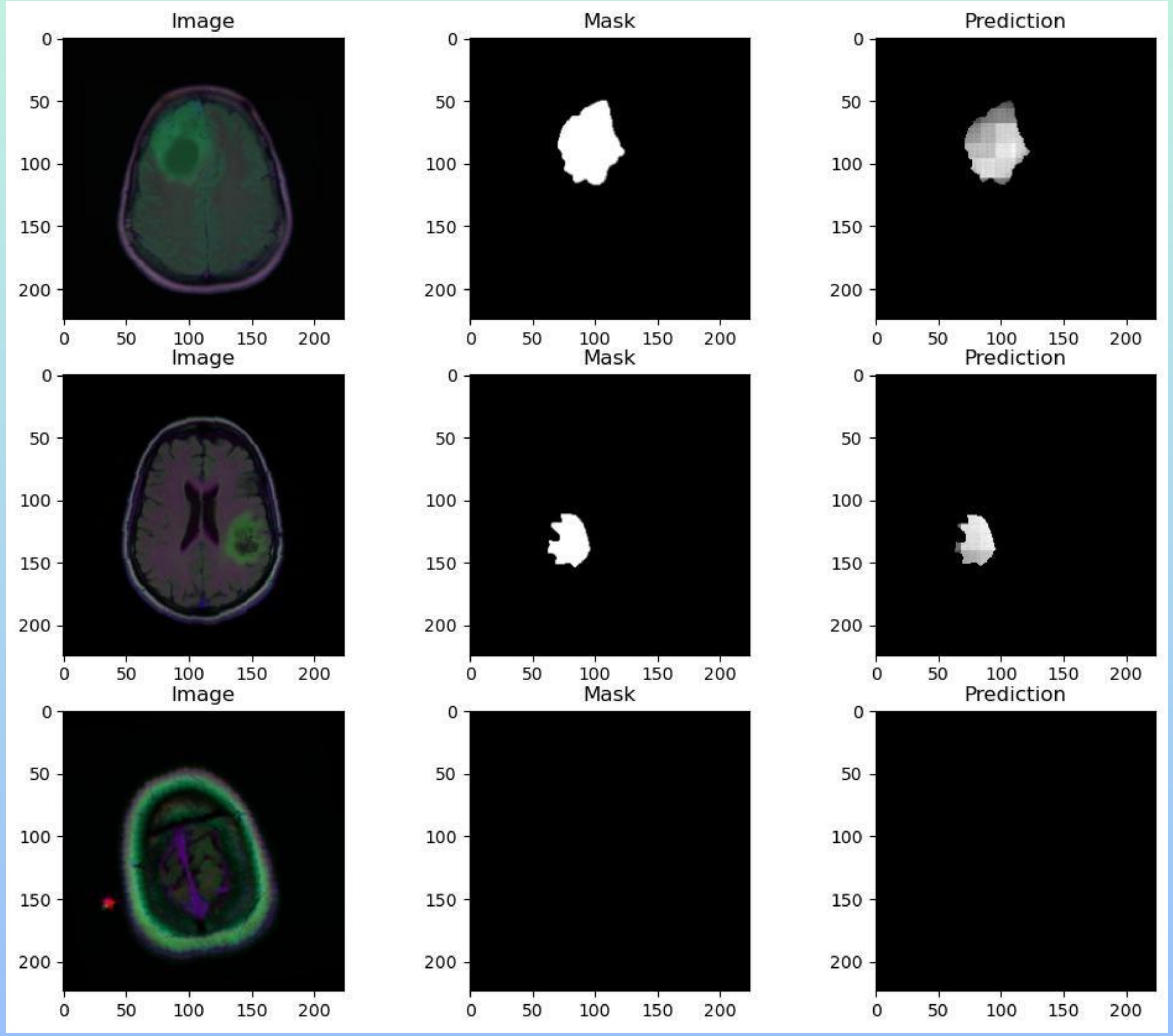
Methodology

- Data Preprocessing
  - Conversion to PNG
  - Image resizing to 224x224 pixels
  - Grayscale conversion of masks
  - Data augmentation techniques applied (flipping, rotation, and color jittering)
- Model Development
  - Swin-Unet architecture
  - Multi-scale feature representation and self-attention mechanisms
  - Training with Adam optimizer and binary cross-entropy loss function
- Loss Function: Binary cross-entropy loss function



Results

- Model Training and Validation
  - Achieved 98.9% accuracy
  - Outperformed traditional CNN architectures and Vision Transformers
  - Visual results
- Comparison of Models
  - Swin-Unet: 98.9% accuracy
  - ViT-based DNN: 97.98% accuracy
  - QFS-Net: 98.23% accuracy



Contributions

- Innovative Model
- High Accuracy

Conclusion

- Effectiveness of Swin-Unet
  - Accurate delineation of brain tumors
  - Assisting in early detection and diagnosis
- Future Work
  - Expanding dataset
  - Experimenting with Swin Transformers and different CNN architectures

Thank you!