

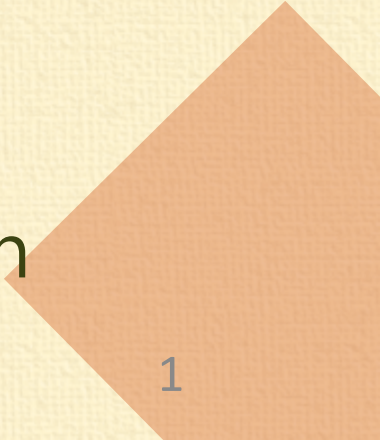


Automated Convolutional Neural Network and Transformer for Multi-Class Classification of Three-Dimensional Brain Images

Cheng-Chun Wu and Guan-Hua Huang

Institute of Statistics

National Yang Ming Chiao Tung University, Taiwan



Parkinson's Disease (PD)

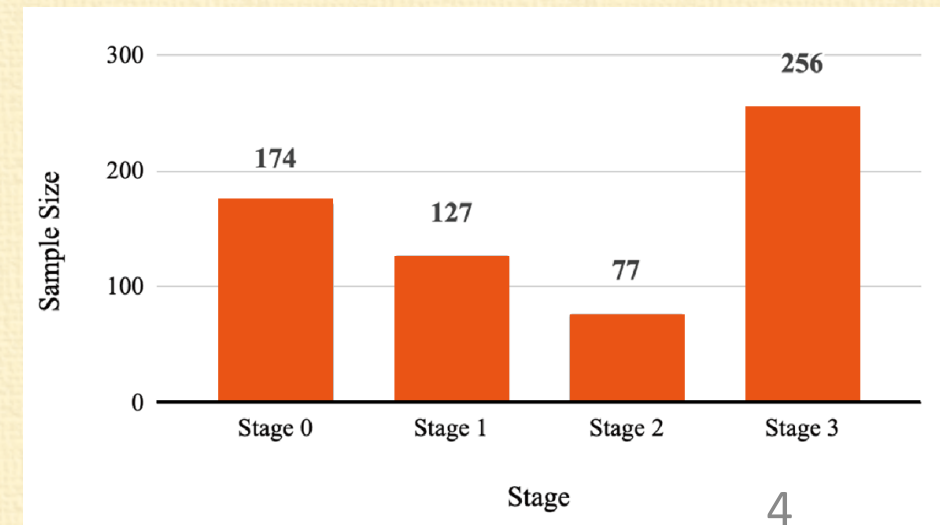
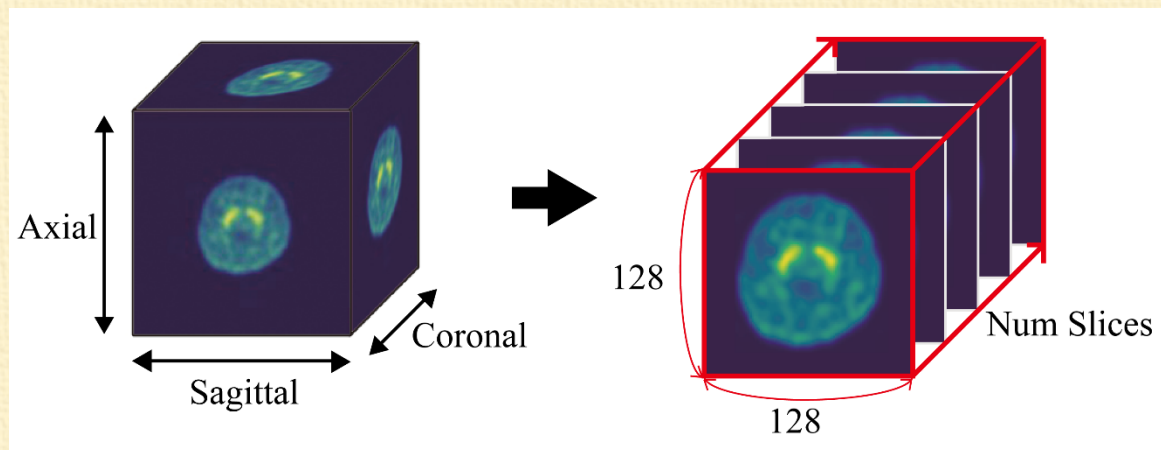
- Parkinson's disease is a progressive nervous system disorder that affects movement. Symptoms start gradually, sometimes starting with a barely noticeable tremor in just one hand. Tremors are common, but the disorder also commonly causes stiffness or slowing of movement.
- Diagnosis:
 - Clinical disabilities or symptoms
 - Functional imaging
 - ✓ Positron Emission Tomography (PET) 正子造影
 - ✓ Single Photon Emission Computed Tomography (SPECT) 單光子電腦斷層掃描

Objectives

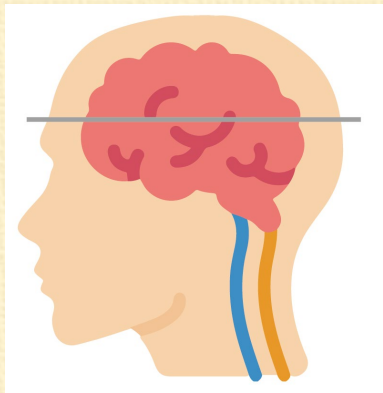
- Develop an appropriate model to deal with multi-class classification task in predicting stages of Parkinson's disease
- Use the whole 3D information as model input
- Take age and gender into consideration
 - The incidence and prevalence in males are higher than in females
 - The incidence rates rise rapidly after the age of 60
- Use 2D CNN or Transformer plus Attention Mechanism model

Dataset

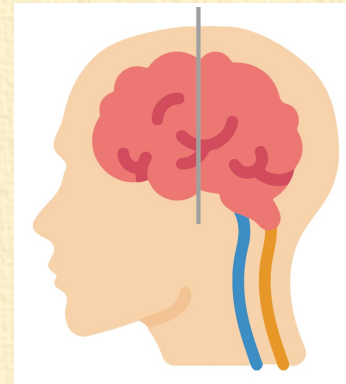
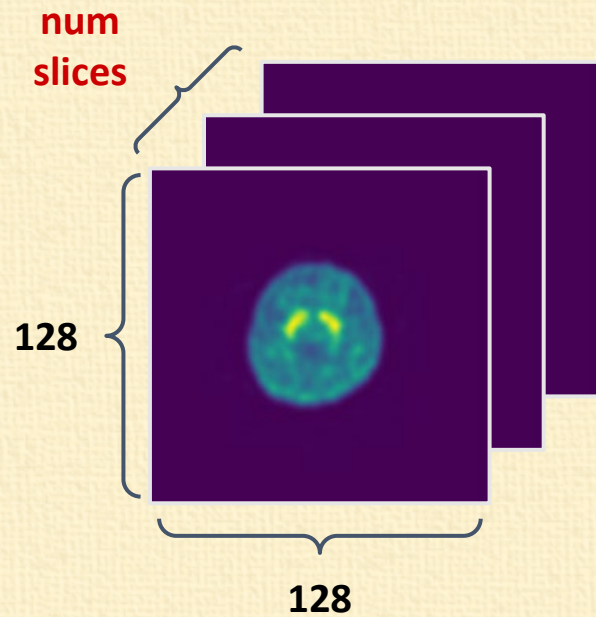
- Source: Kaohsiung Chang Gung Memorial Hospital
- Type: 99mTc-TRODAT-1 SPECT imaging
- Format: DICOM (Digital Imaging and COmmunications in Medicine)
- Shape: (Num Slices X 128 pixel X 128 pixel)
- Total Number: 634



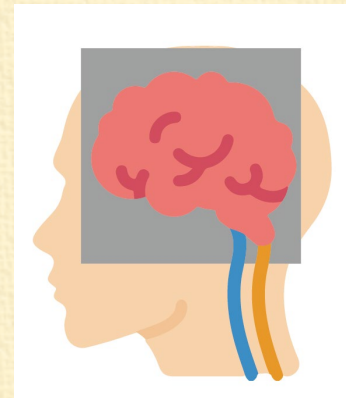
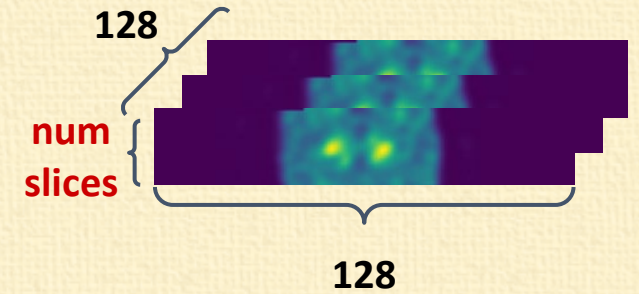
SPECT Image



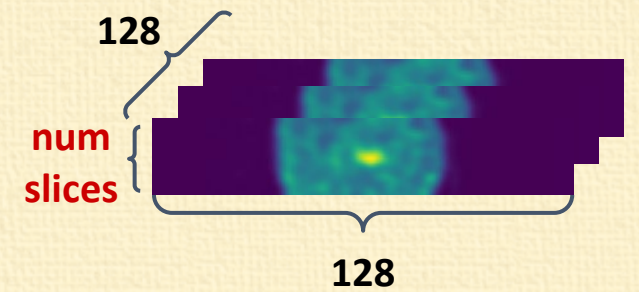
Axial View



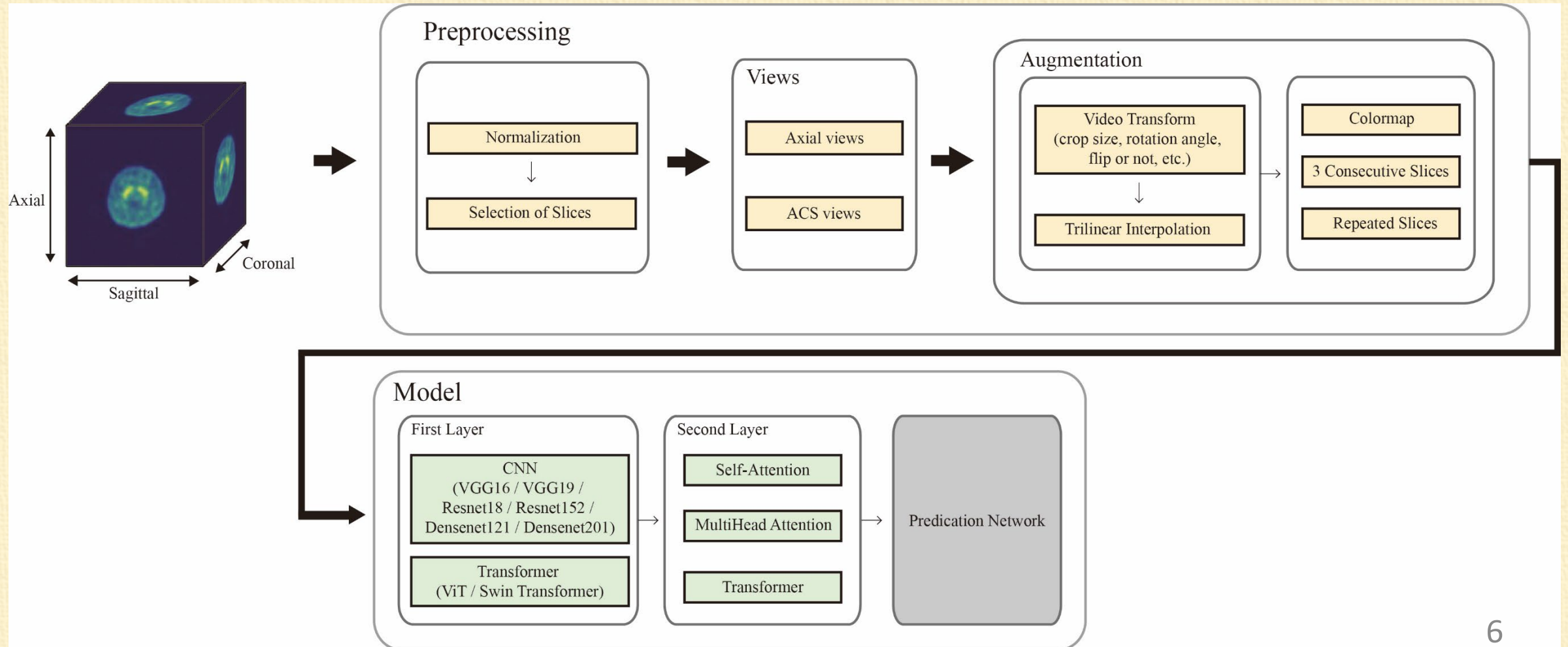
Coronal View



Sagittal View

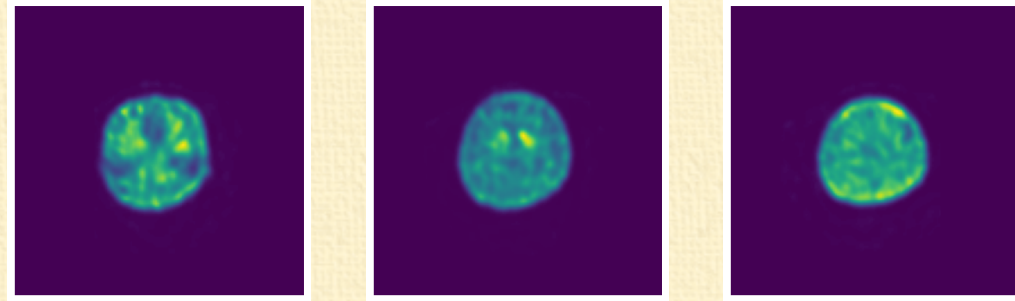


Analytic Procedure

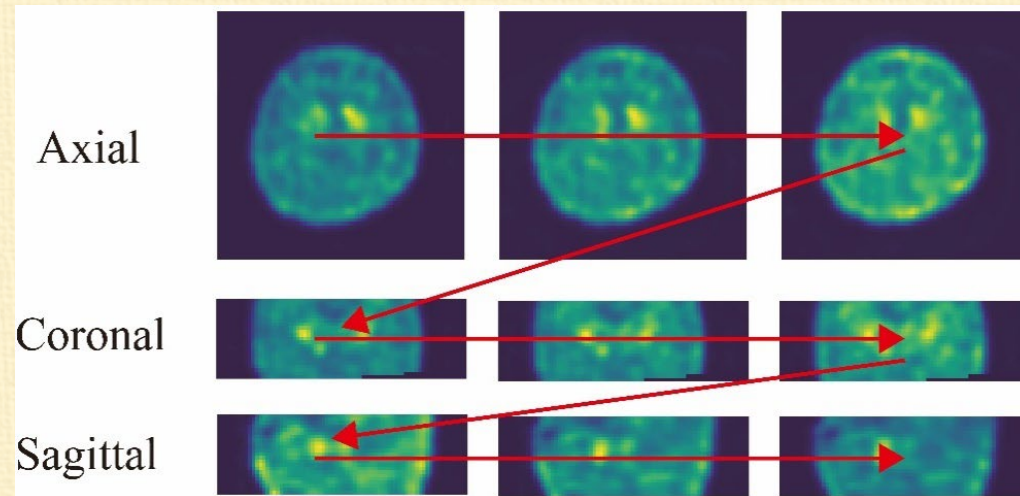


Input Data

- Axial View



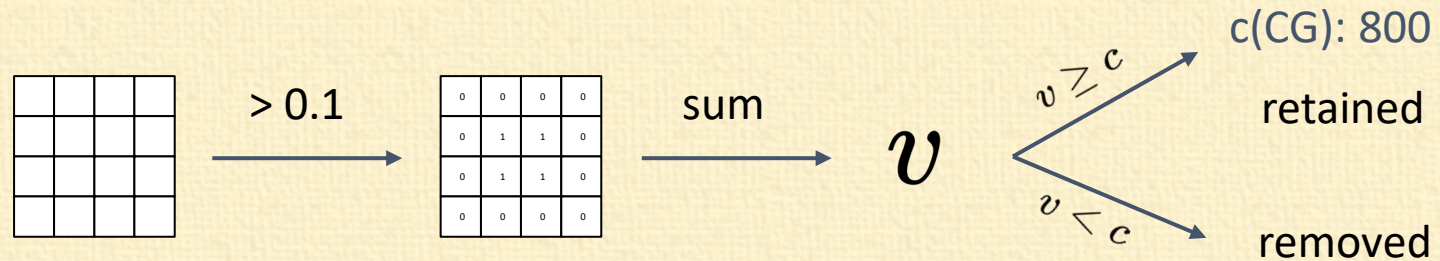
- Axial, Coronal, and Sagittal Planes



Preprocessing

- Min-Max Normalization:
$$\mathbf{X}_i(\text{norm}) = \frac{\mathbf{X}_i - \min(\mathbf{X}_i)}{\max(\mathbf{X}_i) - \min(\mathbf{X}_i)}$$

- Selection of Slices:



- Other Information (Age & Gender):

➤ Age:

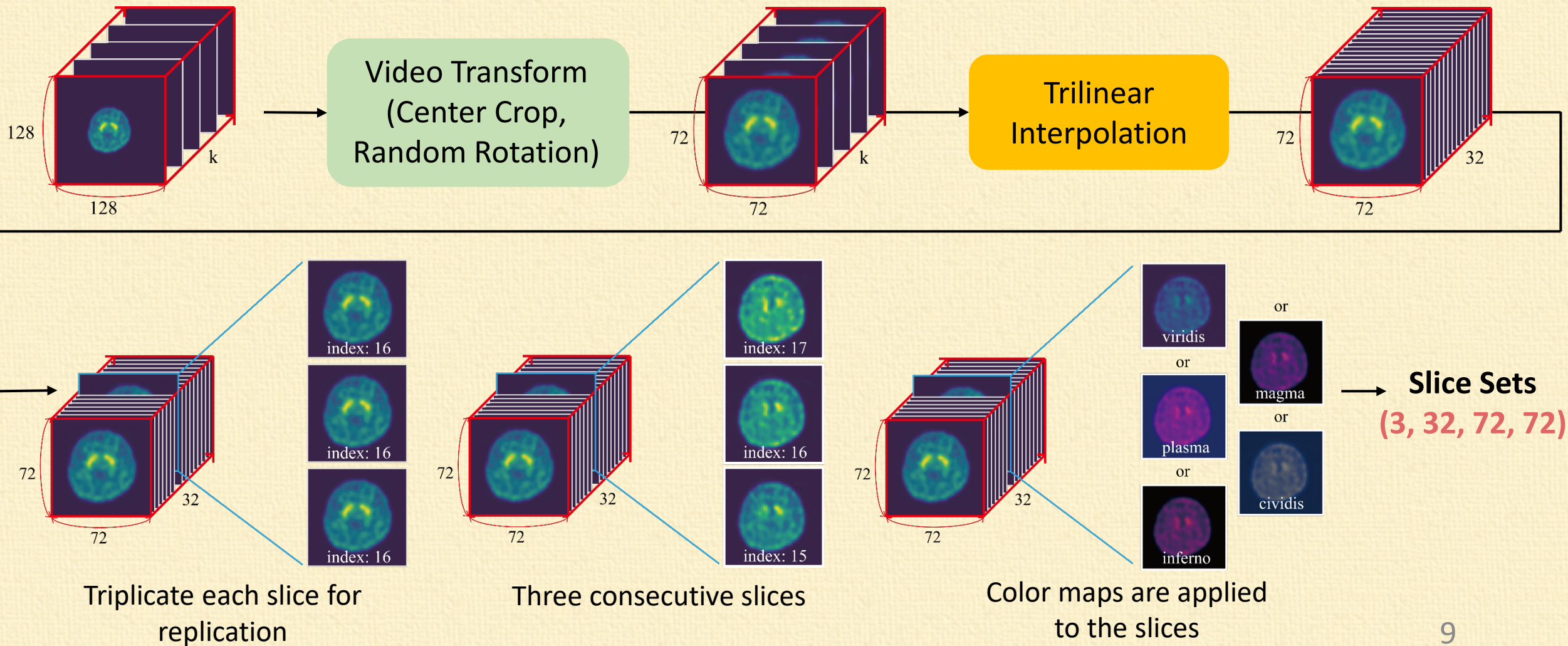
$$y'_i = y_i/100, \quad y_i \in [0, 100]$$

➤ Gender:

$$I = \begin{cases} 1 & \text{if gender is male} \\ 0 & \text{otherwise.} \end{cases}$$

Methods

Augmentation



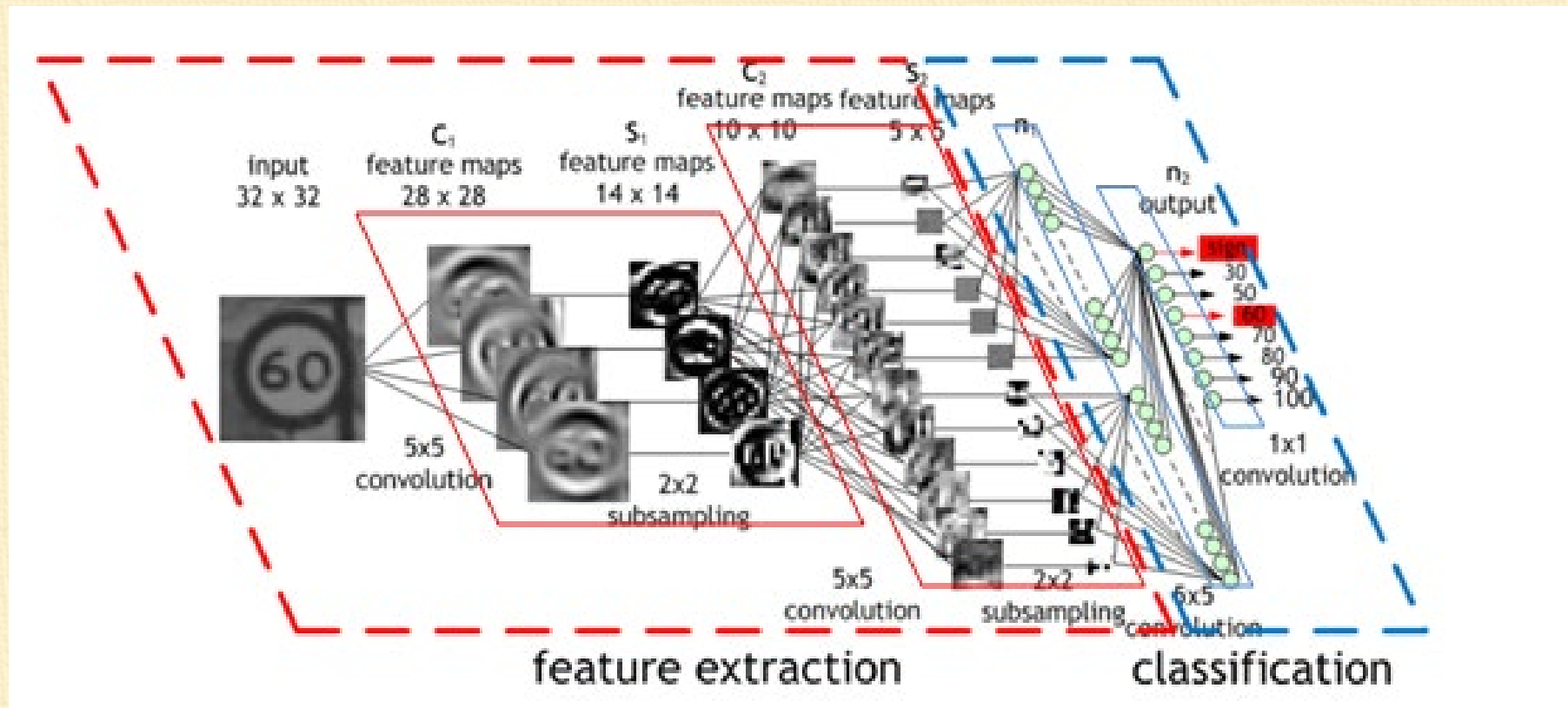
Models

- We employ two different models, namely CNN and Transformer
- The first layer of the model is dedicated to extracting feature values
- The second layer of the model combines the feature values from different slicesets to generate summaries

	CNN	Transformer
First Layer	VGG16, VGG19, ResNet18, ResNet101, Densenet121, Densenet201	Vision Transformer(ViT), Swin Transformer
Second Layer	Self-Attention, MultiHead Attention, Transformer	Self-Attention, MultiHead Attention, Transformer

Models - CNN

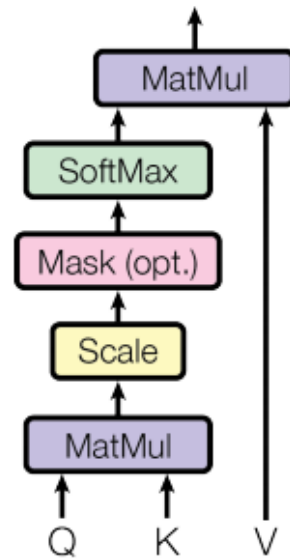
- CNN classic structure



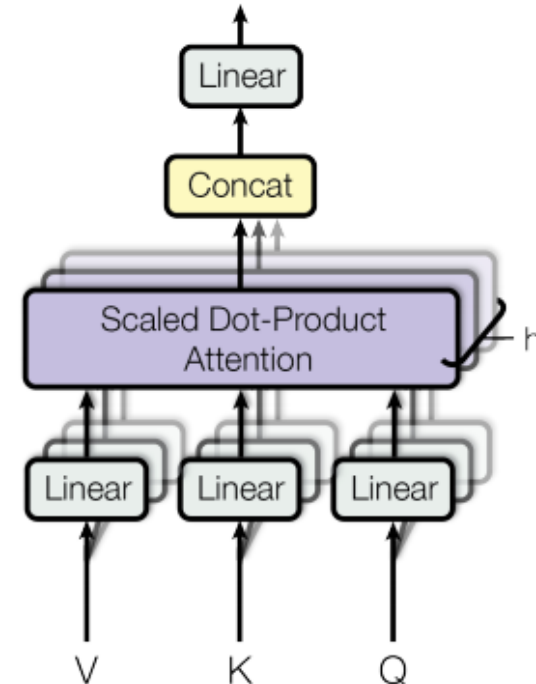
Models - Attention Mechanism

- Attention Mechanism classic structure

Scaled Dot-Product Attention

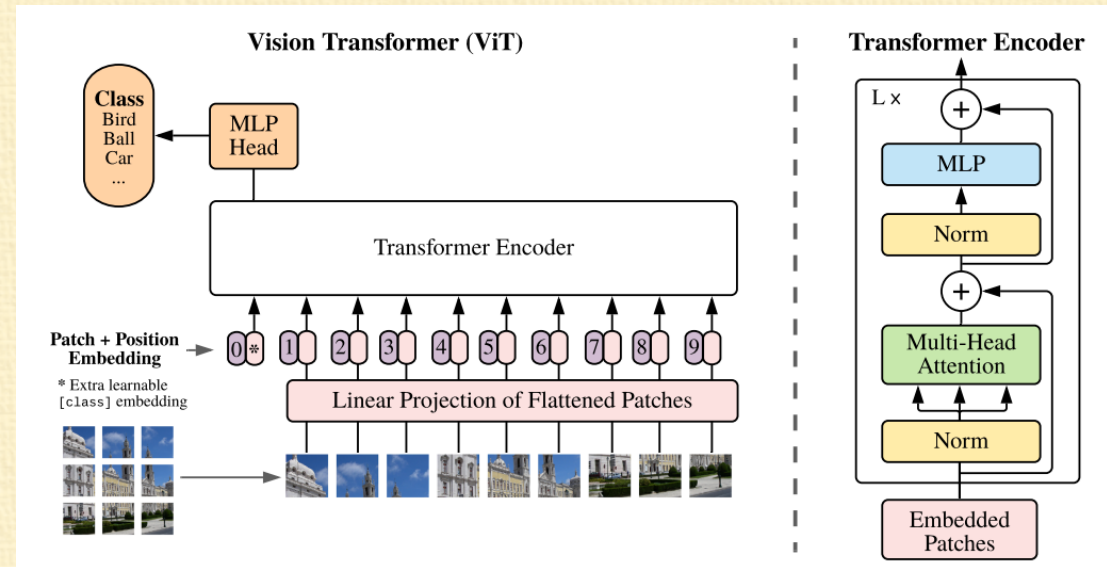
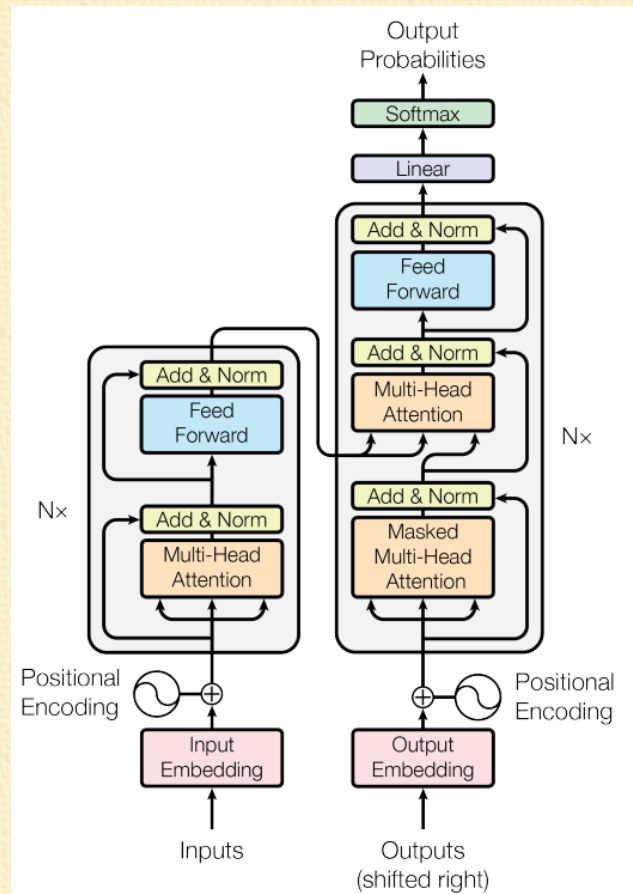


Multi-Head Attention

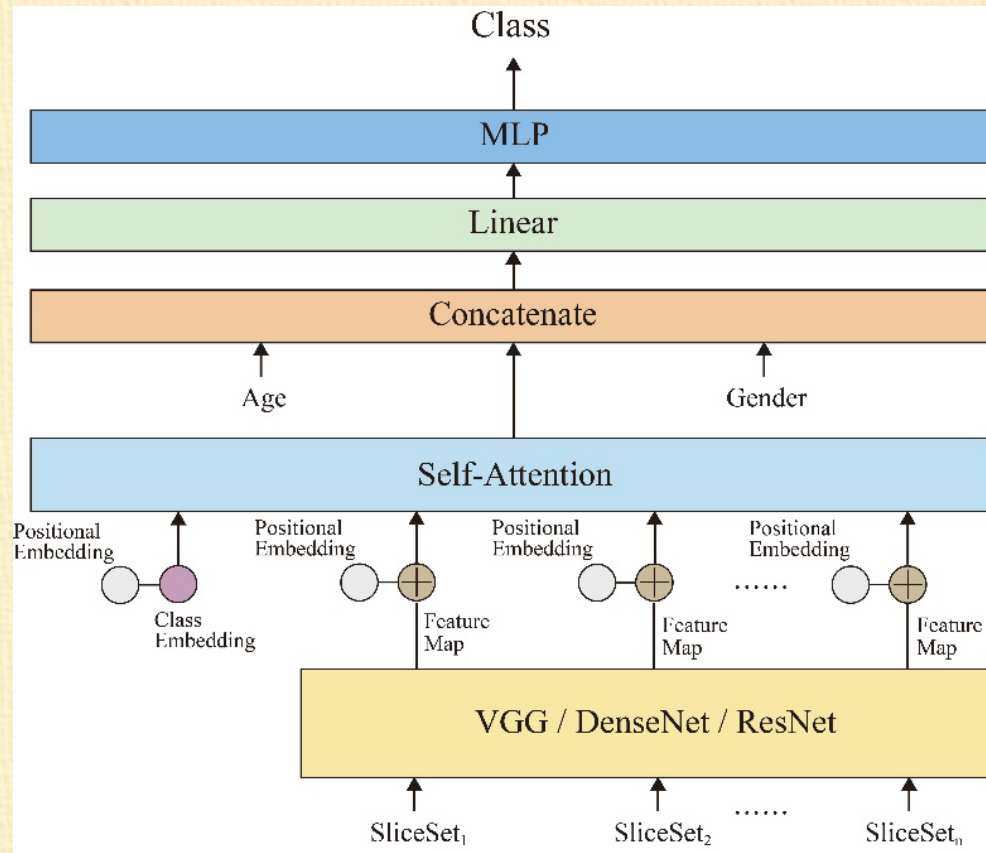


Models - Transformer

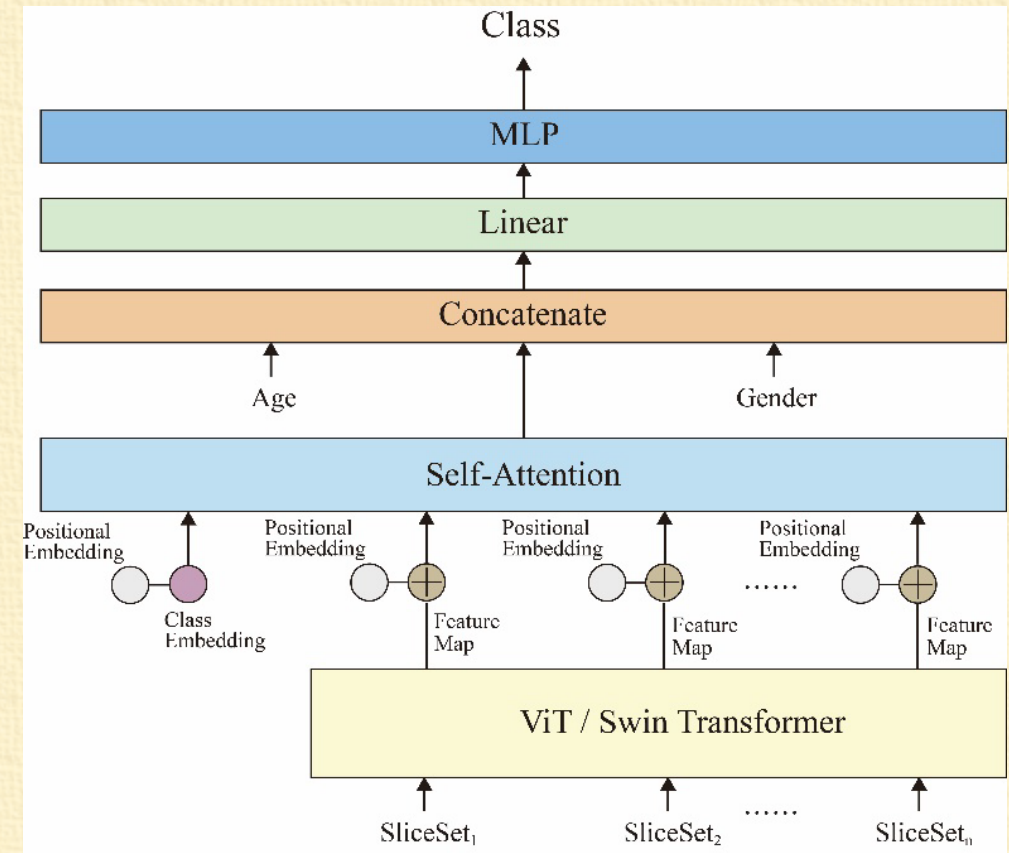
- Transformer classic structure



CNN / Transformer plus Self-Attention

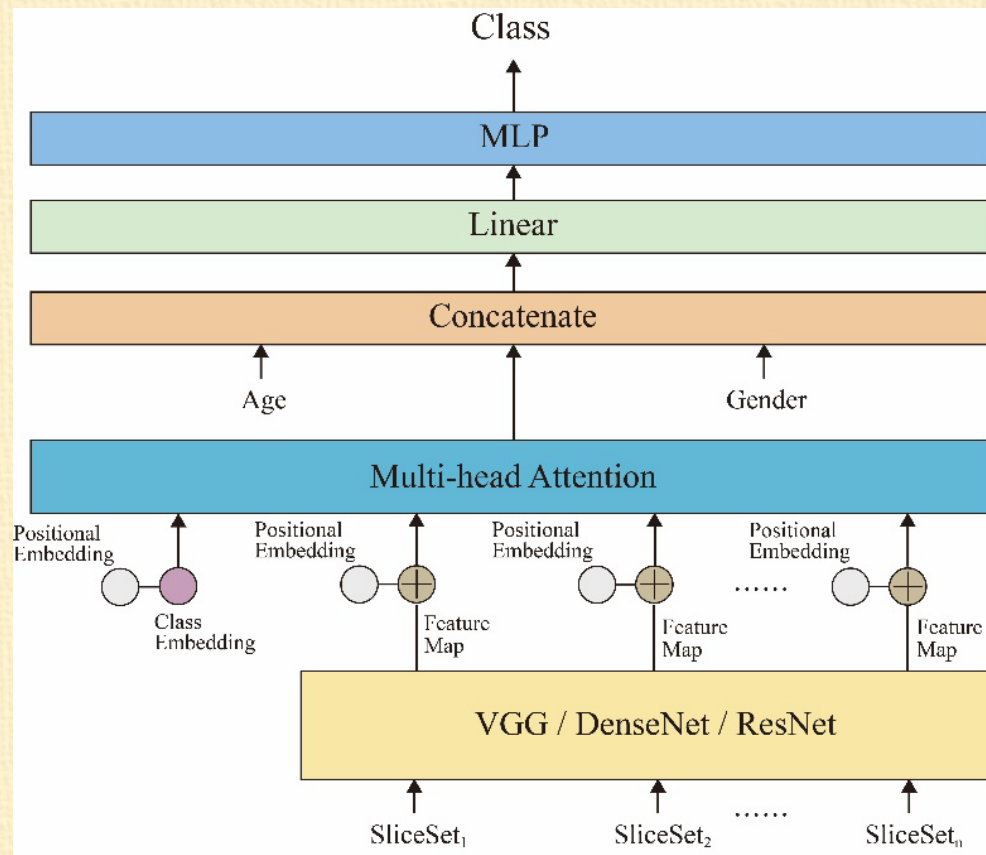


CNN plus Self-Attention

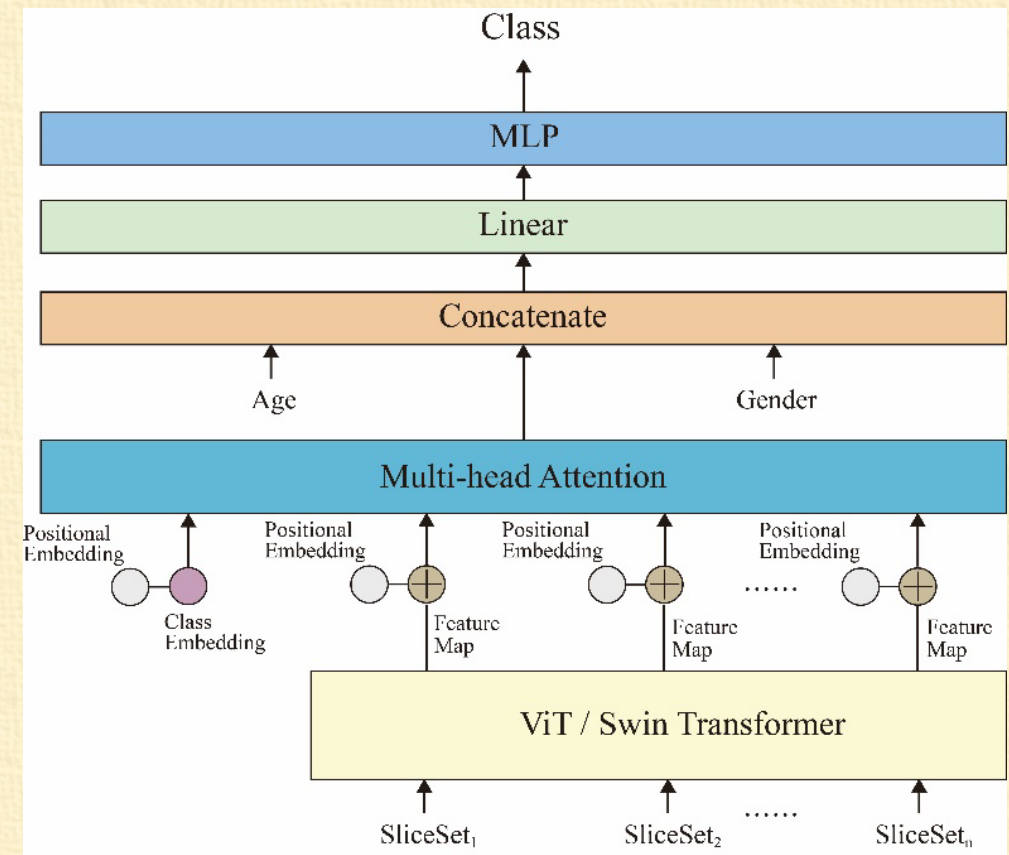


Transformer plus Self-Attention

CNN / Transformer plus Multi-Head Attention

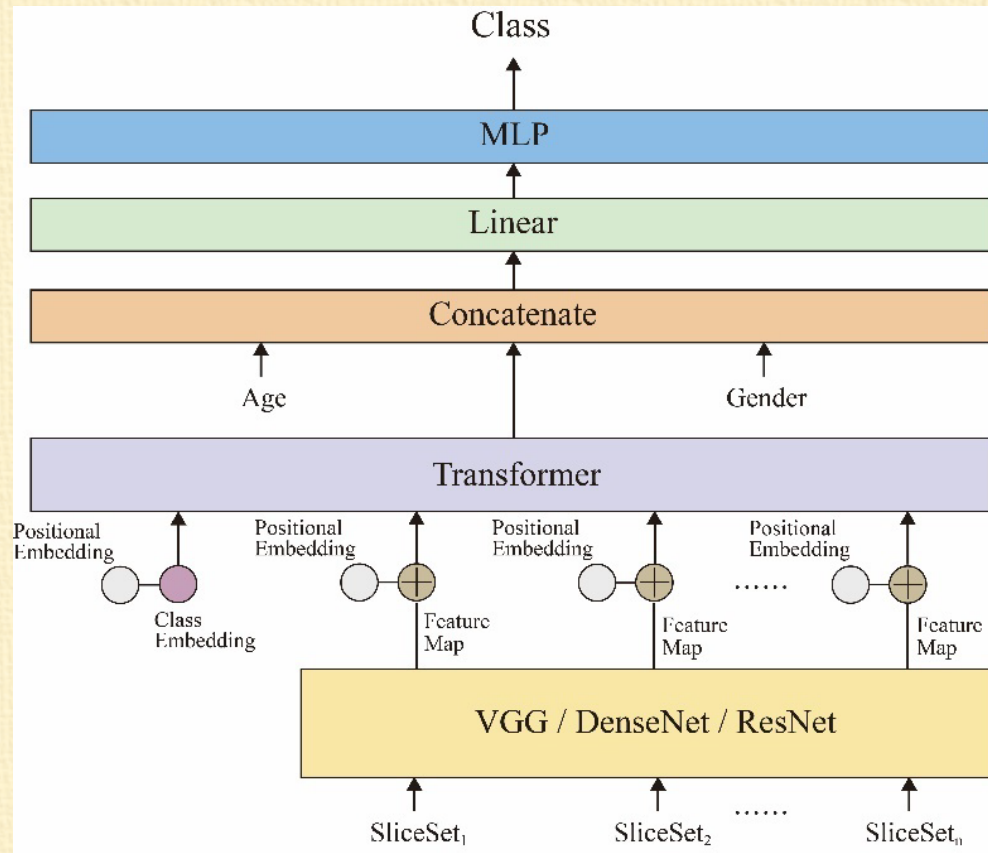


CNN plus Multi-Head Attention

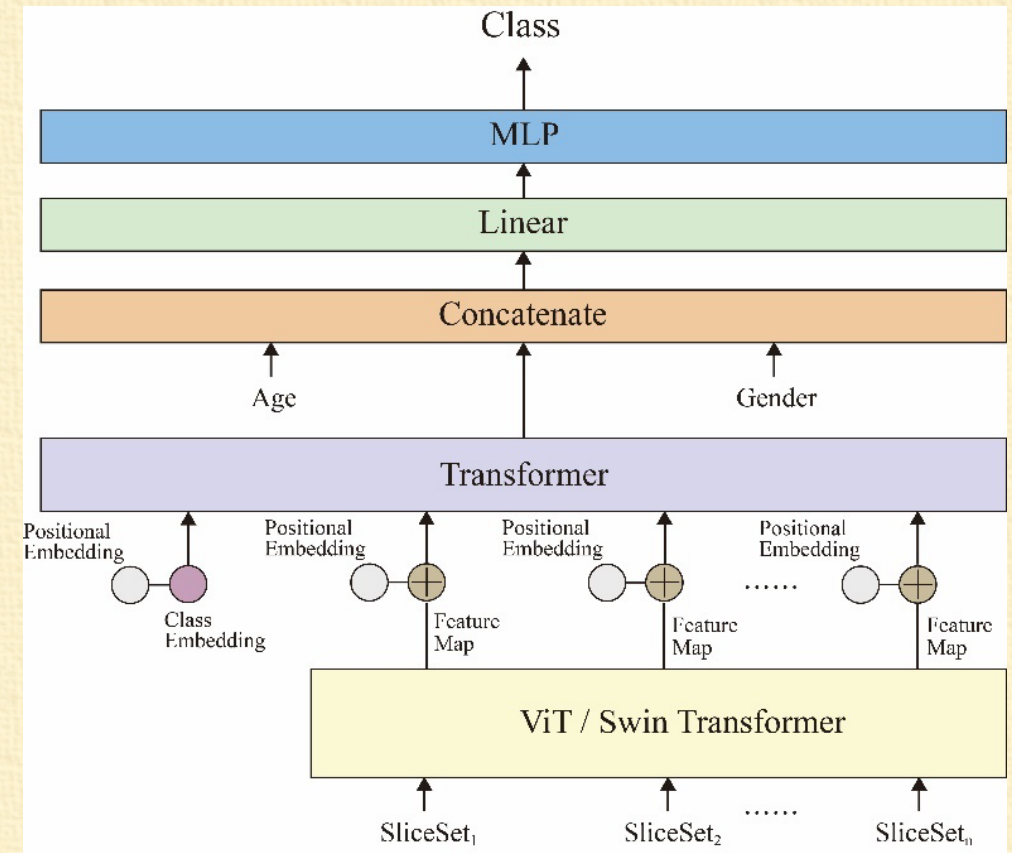


Transformer plus Multi-Head Attention

CNN / Transformer plus Transformer



CNN plus Transformer



Transformer plus Transformer

Automated Machine Learning (AutoML)

- For AutoML, our selected pre-processing includes the following two views.
 - Axial View
 - ACS View
- The process of converting a grayscale 2D image into a 3-channel image.
 - 3 Consecutive Slices
 - Repeated slices
 - Viridis Color Map
 - Plasma Color Map
 - Inferno Color Map
 - Magma Color Map
 - Cividis Color Map

Automated Machine Learning (AutoML)

- AutoML Hyperparameter Categorical Selection

Hyperparameter	Categorical Selection
Pooling	last timestep 、 sum
Activation	linear 、 elu 、 tanh 、 softsign 、 hard sigmoid
Optimizer	SGD 、 Adadelta

- AutoML Hyperparameter Selection Interval

Hyperparameter	Interval Selection
Epoch	200~5,000
Learning Rate	0.00001~0.01
Validation per Step	1~200
Validation Split	0.1~0.5
Number of Heads	2~20
Drop of Transformer	0.0~0.8
Embbing dropout	0.0~0.8

Grayscale Slices Experiment Results

- Based on CNN, Axial views plus 3 consecutive slices have better performance
- Based on Transformer, ACS plus copy slice 3 times with better performance

Model	Accuracy / F1-score			
	Axial viewsplus 3 consecutive slices	Axial views plus copy slice3 times	ACS plus 3 consecutive slices	ACS plus copy slice 3 times
Resnet101 (CNN representative)	0.6303/0.4373	0.6299/0.4362	0.6251/0.4536	0.6173/0.4461
Swin Transformer V2 (Transformer representative)	0.3622/0.1692	0.2519/0.1702	0.4015/0.1448	0.4218/0.2436

Color Map

- Using a color map in ACS is expected to yield better performance compared to Axial views
- Using a color map does not necessarily result in better performance for CNN and Transformer models

Table: Axial views plus Color Map of Slices Experiment Results

Model	Viridis color map	Plasma color map	Inferno color map	Magma color map	Cividis color map
Resnet101 (CNN representative)	0.5118/0.3088	0.4015/0.1432	0.4015/0.1432	0.4803/0.2613	0.4015/0.1432
Swin TransformerV2 (Transformer representative)	0.4015/ 0.1432	0.2755/ 0.1080	0.4015/ 0.1432	0.2755/ 0.1080	0.4015/ 0.1432

Color Map

- ACS utilizing a color map is expected to deliver superior performance compared to Axial views
- CNN and Transformer models do not exhibit improved performance when using a color map

Table: ACS plus Color Map of Slices Experiment Results

Model	Viridis color map	Plasma color map	Inferno color map	Magma color map	Cividis color map
Resnet101 (CNN representative)	0.6015/0.3432	0.5858/0.3791	0.6015/0.3432	0.6409/0.4453	0.6015/0.3432
Swin- TransformerV2 (Transformer representative)	0.4015/0.1448	0.2381/0.1243	0.4388/0.2362	0.4015/0.1448	0.4251/0.2536

CNN in Model First Layer Experiment Results

- ResNet demonstrates outstanding performance among all the tested CNN models on our small dataset

Model	Accuracy	F1-score
Densenet201	0.6094	0.4395
Densenet121	0.5149	0.4308
Resnet152	0.6724	0.4735
Resnet18	0.7511	0.5511
VGG19	0.6015	0.3432
VGG16	0.6724	0.4833

Transformer in Model First Layer Experiment Results

- The results indicate that Swin Transformer outperforms ViT in terms of performance
- Overall, CNN still outperforms Transformer in terms of performance

Model	Accuracy	F1-score
Vision Transformer(ViT)	0.3047	0.0849
Swin Transformer V2	0.4218	0.2436

CNN / Transformer plus Attention Mechanism model

- Using Transformer in the second layer of the model performs exceptionally well, both in CNN and Transformer architectures
- The performance of the model is directly influenced by the complexity of its second layer

Model	Self-Attention	Multi-Head Attention	Transformer
Resnet18 (Best CNN)	0.6181 / 0.4528	0.6228 / 0.4735	0.7511 / 0.5511
Swin Transformer V2 (Best Transformer)	0.3936 / 0.1606	0.3968 / 0.2832	0.4218 / 0.2436

Different Training Weights with CNN / Transformer Model

- Shared weight training methods yield more efficient weights compared to individual weight training methods
- Increasing training data volume leads to more stable training results, but stability does not guarantee improved accuracy

Model	Each slices train its weights	Train the weights of the all-slices share	Model	Each slices train its weights	Train the weights of the all-slices share
Resnet18 (Best CNN)	0.6409 / 0.4453	0.7511 / 0.5511	Resnet18 (Best CNN)	0.6685 / 0.4589	0.7280 / 0.5496
Swin Transformer (Best Transformer)	0.3858 / 0.3177	0.4218 / 0.2436	Swin Transformer (Best Transformer)	0.3826 / 0.3204	0.4076 / 0.2697

1 Batch Size

8 Batch Size

AutoML

- The pre-processing techniques and hyperparameters selected by human experts in the CNN model resulted in superior outcomes
- Using AutoML as a pre-processing and hyperparameter selector in the Transformer model demonstrated competitive performance

Model	Human	AutoML
Resnet18 (Best CNN)	0.7511 / 0.5511	0.7187 / 0.5292
Swin Transformer (Best Transformer)	0.4218 / 0.2436	0.4801 / 0.2518

AutoML (cont.)

- Selection Results base on CNN model

Parameters	Optimal Value (AutoML)	Optimal Value (Human)
Views	Axial views	Axial views
Augmentation	3 consecutive slices	3 consecutive slices
Epoch	948	3000
Learning rate	0.000698556	0.0001
Val per step	20	10
Pooling	last timestep	last timestep
Activation	relu	relu
Validation split	0.2	0.2
Optimizer	SGD	Adam
CNN Dropout	0.1	0.1
Num of heads	3	3
Drop of transformer	0.1	0.1
Emb dropout	0.1	0.1

AutoML (cont.)

- Selection Results base on Transformer model

Parameters	Optimal Value (AutoML)	Optimal Value (Human)
Views	ACS	ACS
Augmentation	3 consecutive slices	3 consecutive slices
Epoch	523	3000
Learning rate	0.00138068	0.0001
Val per step	12	10
Pooling	last timestep	last timestep
Activation	relu	relu
Validation split	0.2	0.2
Optimizer	SGD	Adam
Number of heads of the first / second layer Transformer	3 / 3	3 / 3
Drop of transformer of the first / second layer Transformer	0.3 / 0.1	0.1 / 0.1
Emb dropout of the first / second layer Transformer	0.1 / 0.2	0.1 / 0.1

Conclusion

- The impact of the pre-processing method varies depending on the specific model employed
- The model's effectiveness and robustness increase as more relationships between slices are considered
- The combination of the CNN model with the Attention-Mechanism model demonstrates better performance than the combination of the Transformer model with the Attention-Mechanism model



Thank You



Q & A