Seminar - Fall 2023

Medical Image Segmentation for Realizing Human Digital Twin

Approaches and Technical Methods

Syed Hasnain Raza Shah

Outline

1. **Recap**

2. **Approaches & Technical Methods**

3. **Implementation**

Outline

1. **Recap**
2. Approaches & Technical Methods
3. Implementation

Recap

Problem

- **To identify precise structure of bones and muscles from CT/MRI images in order to build human digital twin**

Motivation

- **Lack of research focusing on Segmentation of Bones and Muscles**
- **Advancing Healthcare**
- **Personalized treatment strategies, optimizing outcomes and reducing risks.**
- **Assisment in surgical planning, leading to safer and more efficient surgical procedures.**
- **Segmentation-driven digital twins are at the forefront of biomedical research and technology.**

Recap

Research Goals

- Understanding State-of-the-art techniques for medical image segmentation
- Aquiring the latest Vision Transformer-based expertise
- Development of segmentation model for bones and muscles on Visible Korean

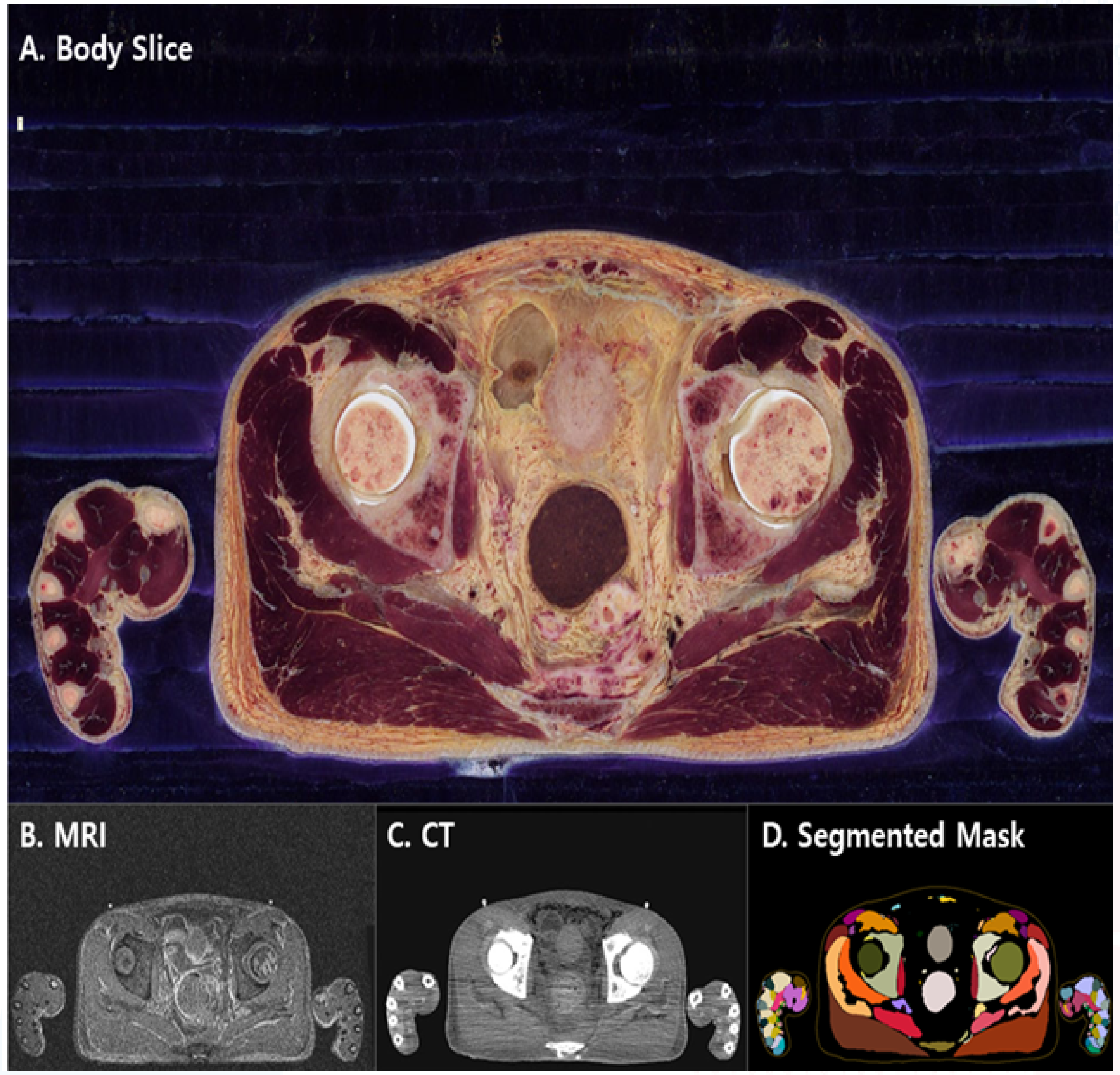
Recap

Visible Korean Dataset

- Full-body scans of a Korean man and a woman, and the corresponding CT, MRI, and segmentation masks

- The segmentation mask consists of a total of 13 major categories

    - MRI: 4mm (256x256, 12bit)
    - CT : 1mm (512,512, 12bit, 1702 in total)
    - BodySlice: 1mm (2048x1216, 24bit, 1702 in total)



Outline

1. Recap
2. **Approaches & Technical Methods**
3. Implementation

Approaches & Technical Methods

Medical Image Segmentation on Automatic Cardiac Diagnosis Challenge (ACDC)

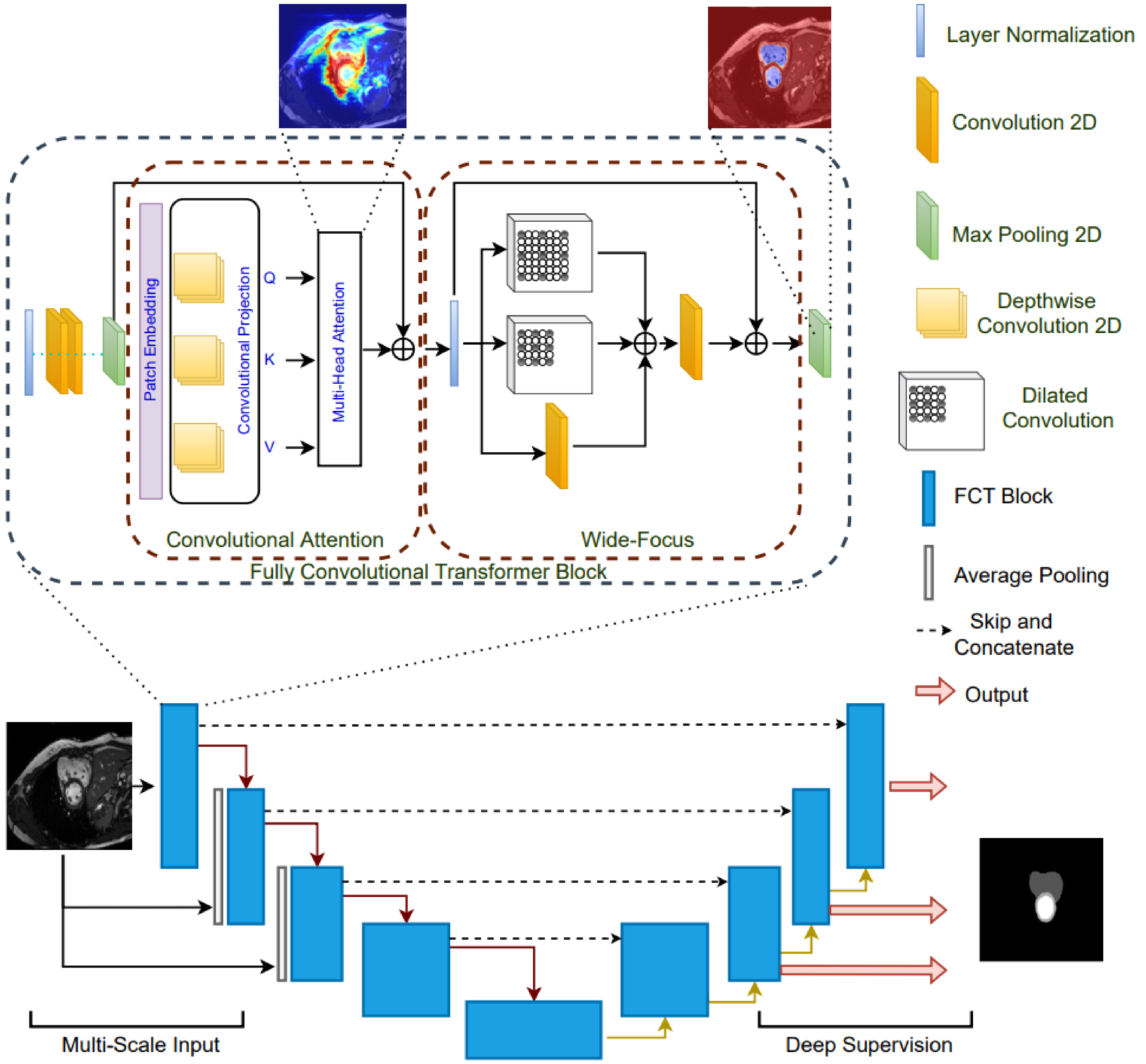
Rank	Model	Avg DSC	Paper	Code	Result	Year	Tags
1	**FCT**	93.02	The Fully Convolutional Transformer for Medical Image Segmentation			2022	
2	**MERIT**	92.32	Multi-scale Hierarchical Vision Transformer with Cascaded Attention Decoding for Medical Image Segmentation			2023	
3	**nnFormer**	92.06	nnFormer: Interleaved Transformer for Volumetric Segmentation			2022	
4	**TransCASCADE**	91.63	Medical Image Segmentation via Cascaded Attention Decoding			2022	
5	**PVT-CASCADE**	91.46	Medical Image Segmentation via Cascaded Attention Decoding			2022	
6	**SwinUnet**	90.00	Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation			2021	
7	**TransUNet**	89.71	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021	
8	**MISSFormer**	87.9	MISSFormer: An Effective Medical Image Segmentation Transformer			2021	
9	**R50-ViT-CUP**	87.57	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021	
10	**R50-AttnUNet**	86.75	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021	

Fully Convolutional Transformer

Neither Transformer-CNN hybrid nor Transformer-UNet

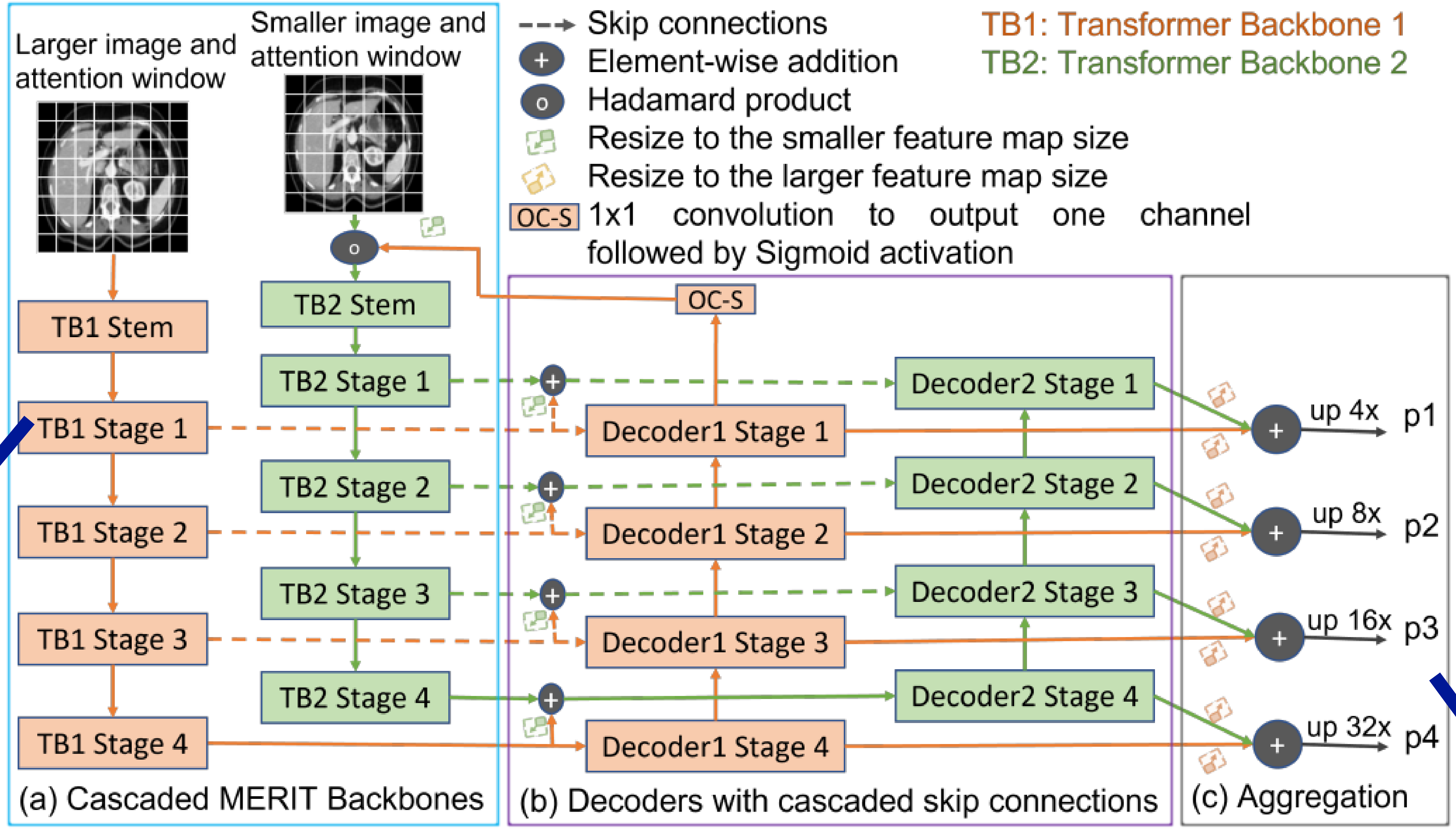
Uses *FCT layer* as building block

- consists of convolutional layers followed by Gelu activation function
- convolutional attention module replacing linear projection with Depthwise-Convolutions, removes positional encoding
- wide focus module contains dilated convolutions and convolutional layer for feature aggregation



"The fully convolutional transformer for medical image segmentation." Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. 2023.

Multi-scale Heirarchical Vision Transformer (MERIT)



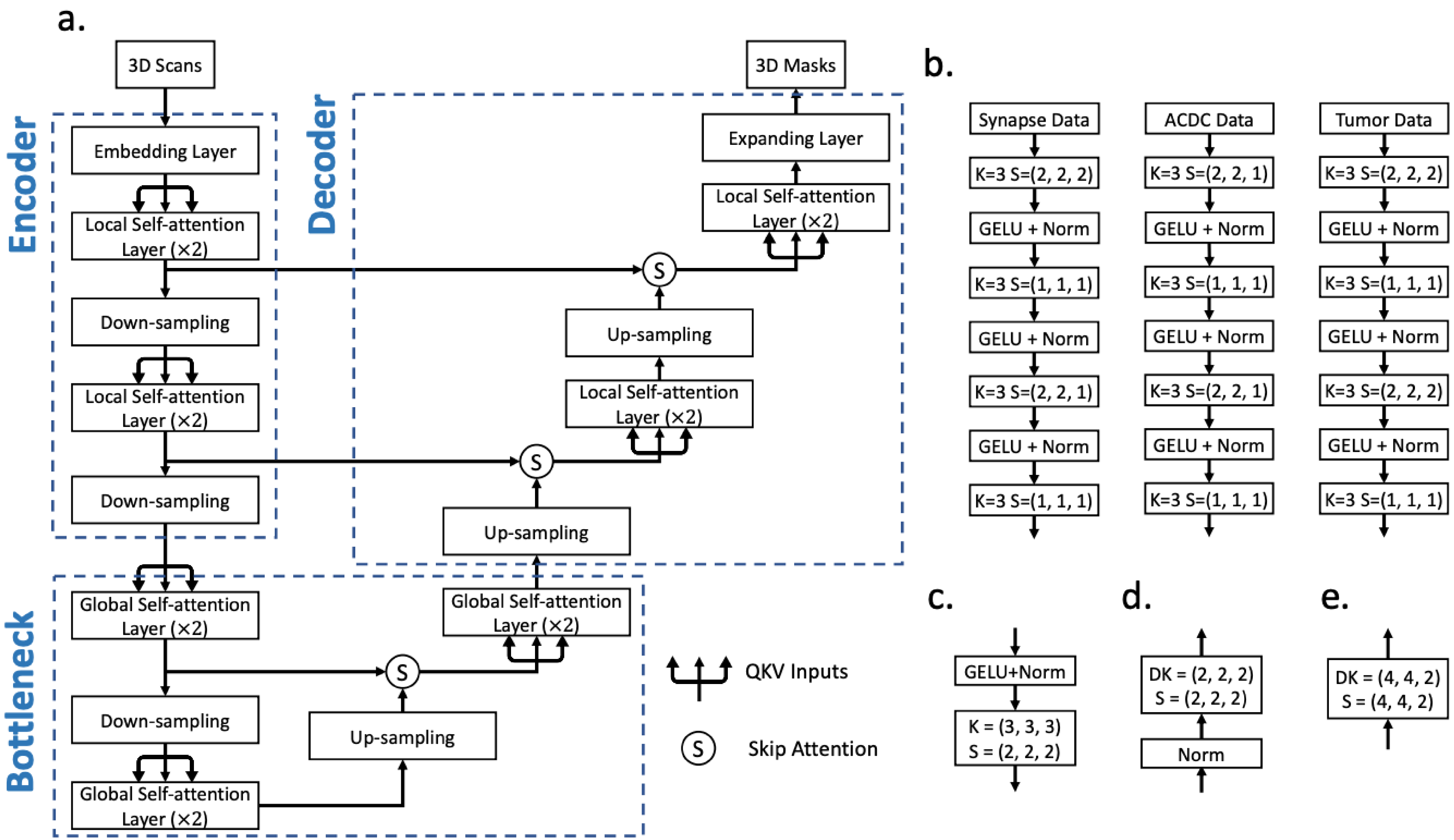
MaxViT blocks

$$\hat{y} = \alpha \times p1 + \beta \times p2 + \gamma \times p3 + \psi \times p4$$

- **combines MaxViT and Cascaded Decoders**
- **captures both multi-scale and multi-resolution features**

"Multi-scale Hierarchical Vision Transformer with Cascaded Attention Decoding for Medical Image Segmentation", MIDL 2023

nnFormer

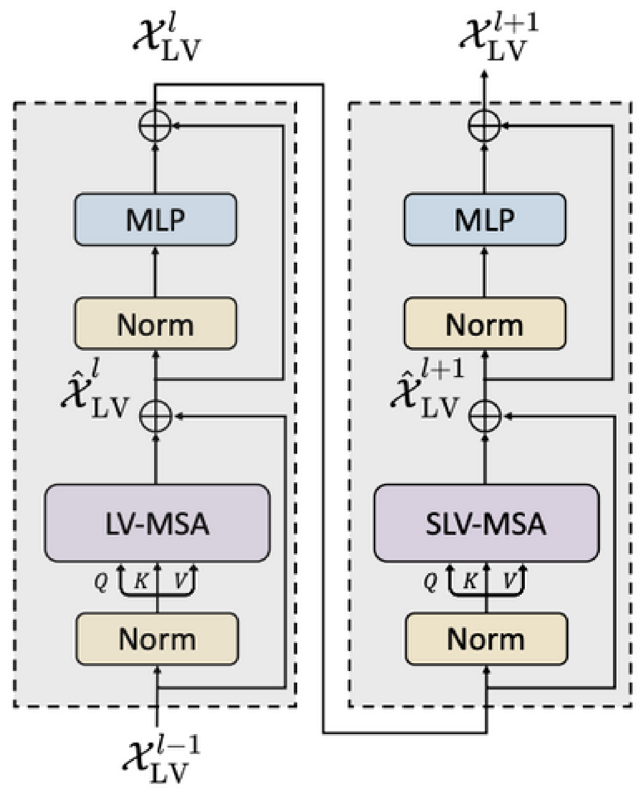


"nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer," in IEEE Transactions on Image Processing, vol. 32, pp. 4036-4045, 2023

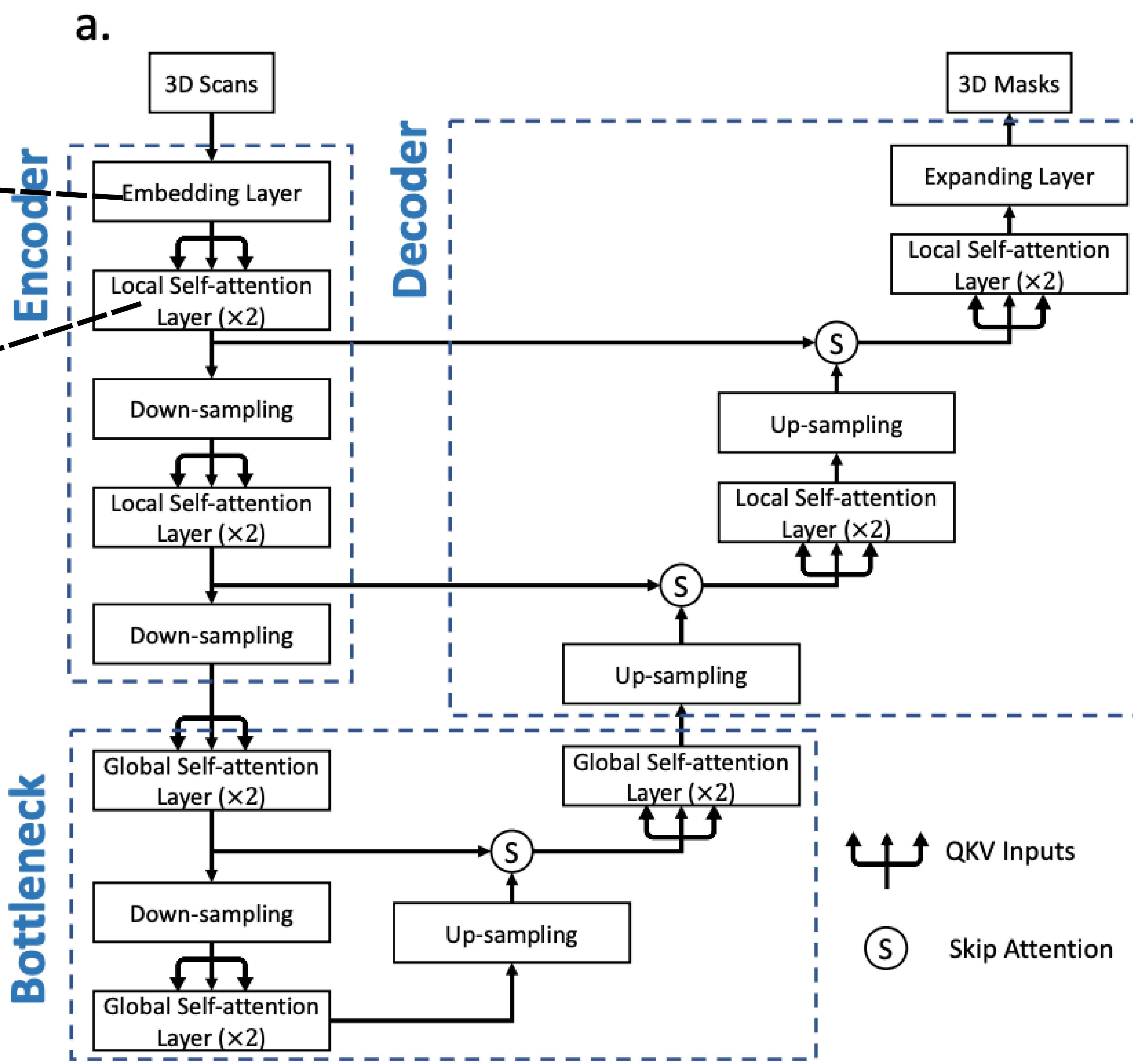
nnFormer

consists of convolution layers

1. precise spatial information
2. high resolution low level features

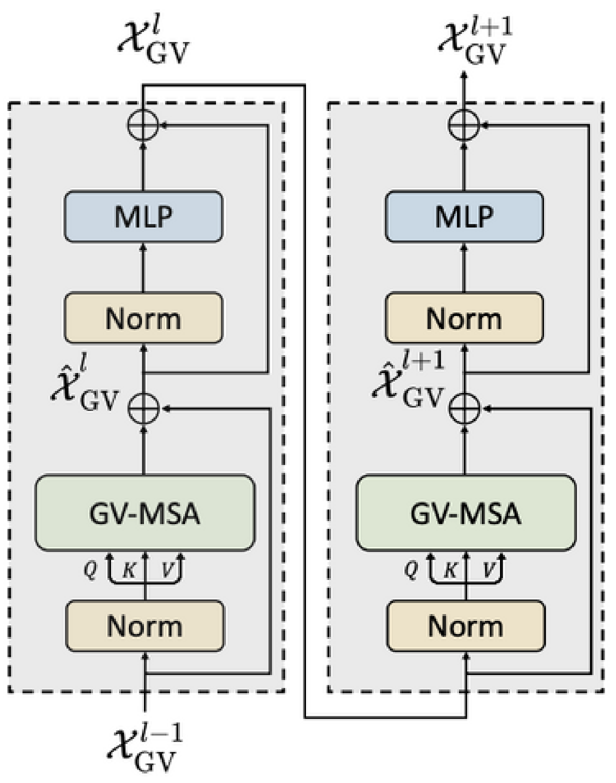


computes long range dependencies within 3D local volumes

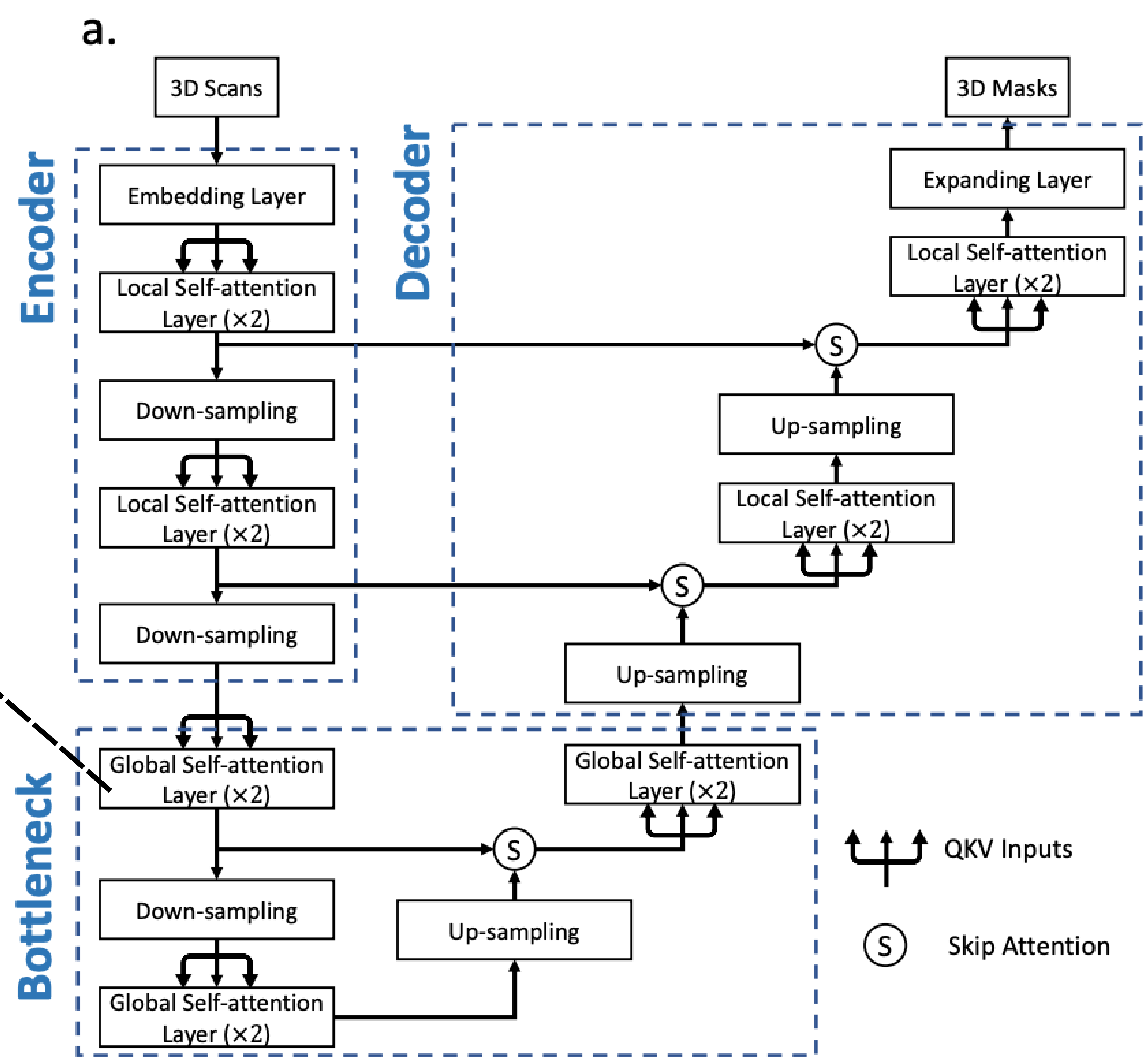


"nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer," in IEEE Transactions on Image Processing, vol. 32, pp. 4036-4045, 2023

nnFormer

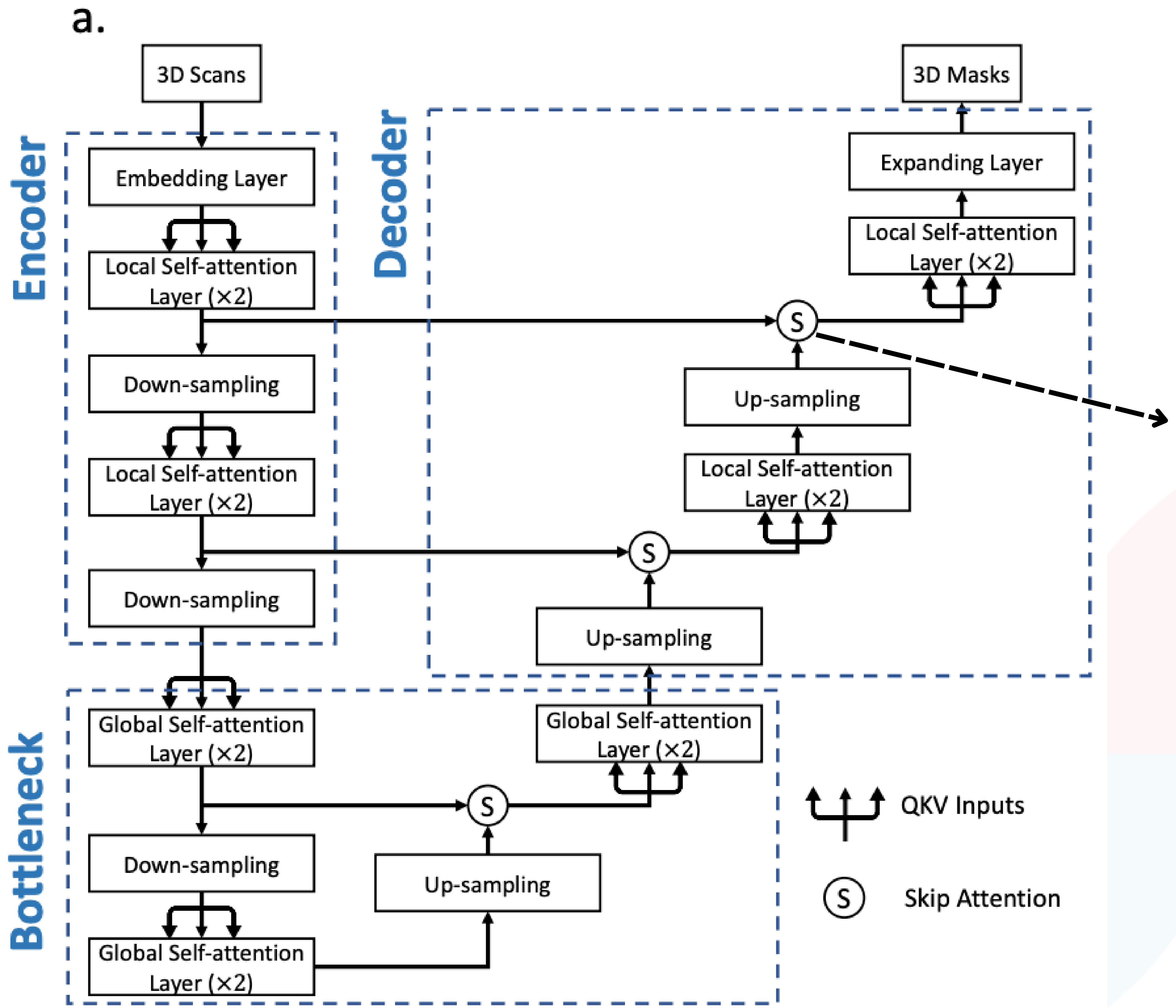


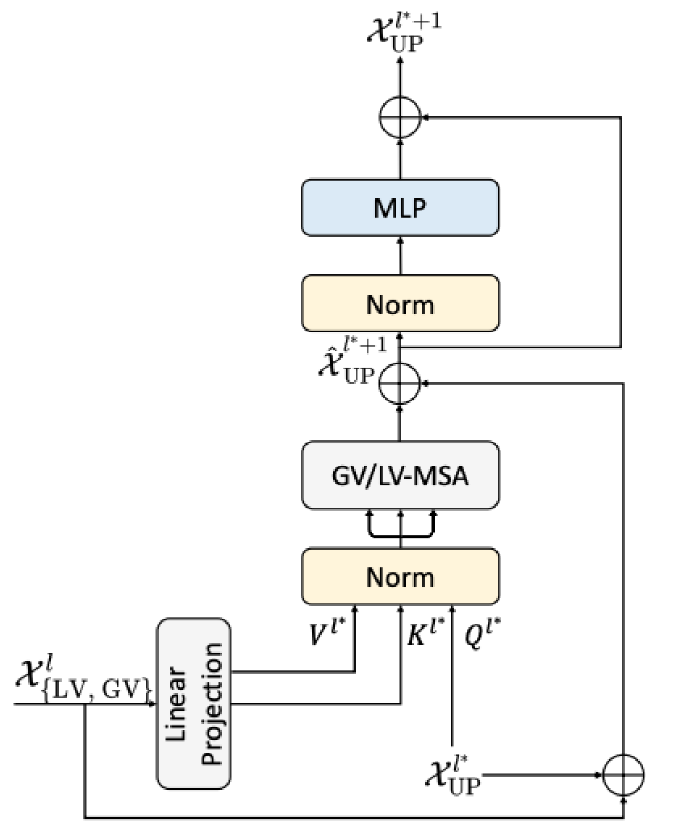
computes long range dependencies within 3D global volumes



"nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer," in IEEE Transactions on Image Processing, vol. 32, pp. 4036-4045, 2023

nnFormer



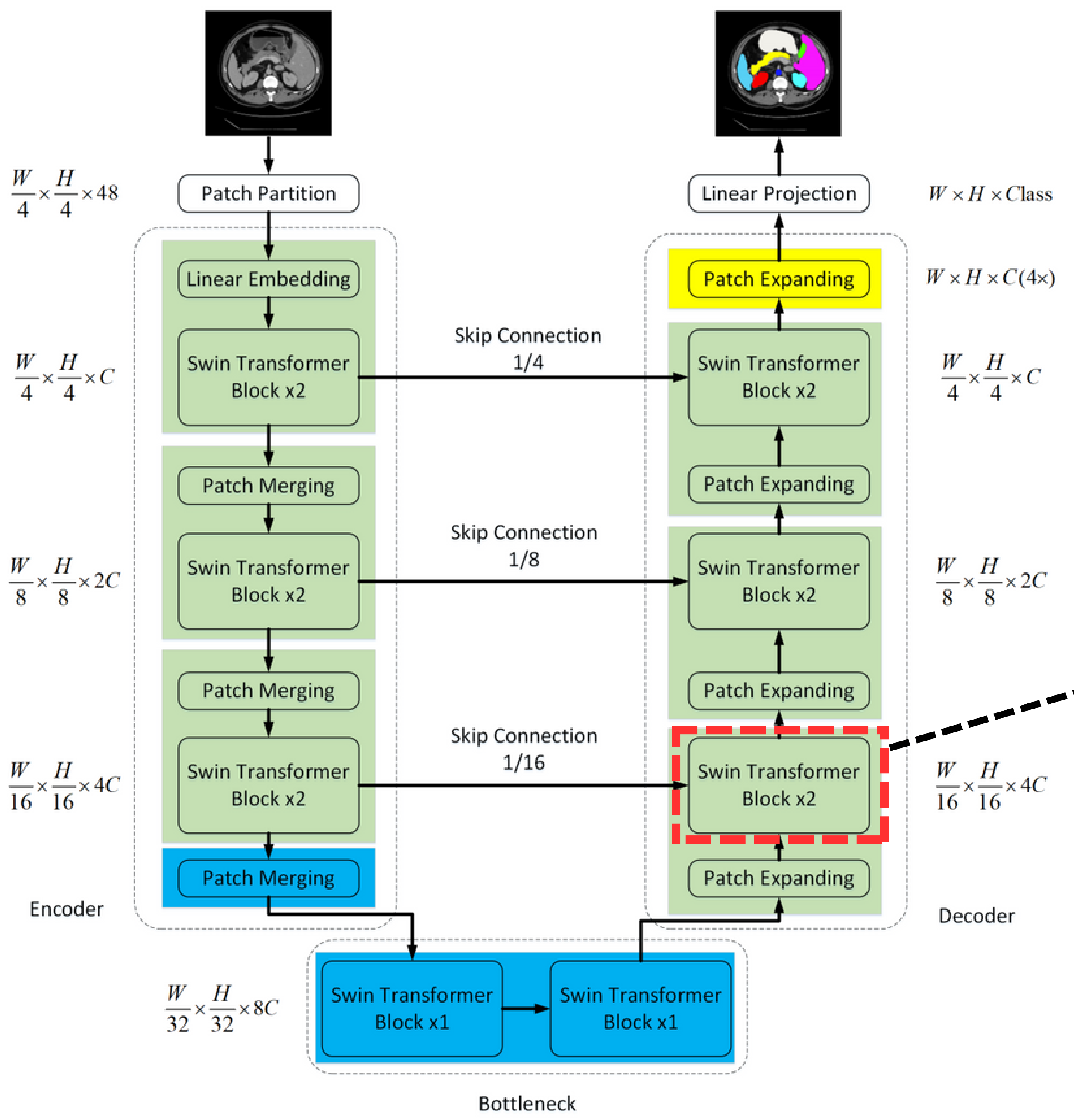


replaces traditional skip connection with skip attention

"nnFormer: Volumetric Medical Image Segmentation via a 3D Transformer," in IEEE Transactions on Image Processing, vol. 32, pp. 4036-4045, 2023

SwinUnet

Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation 5



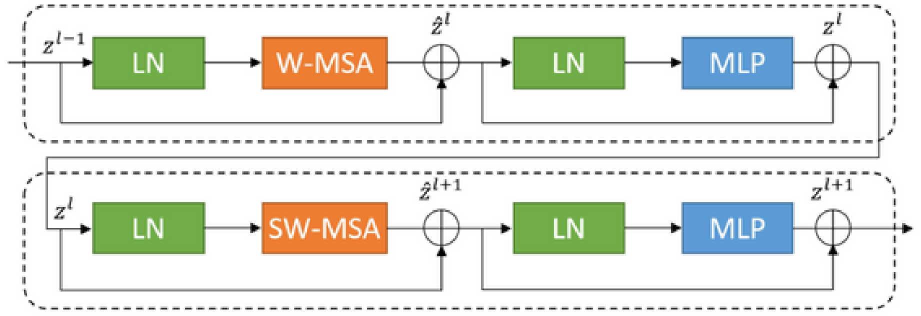
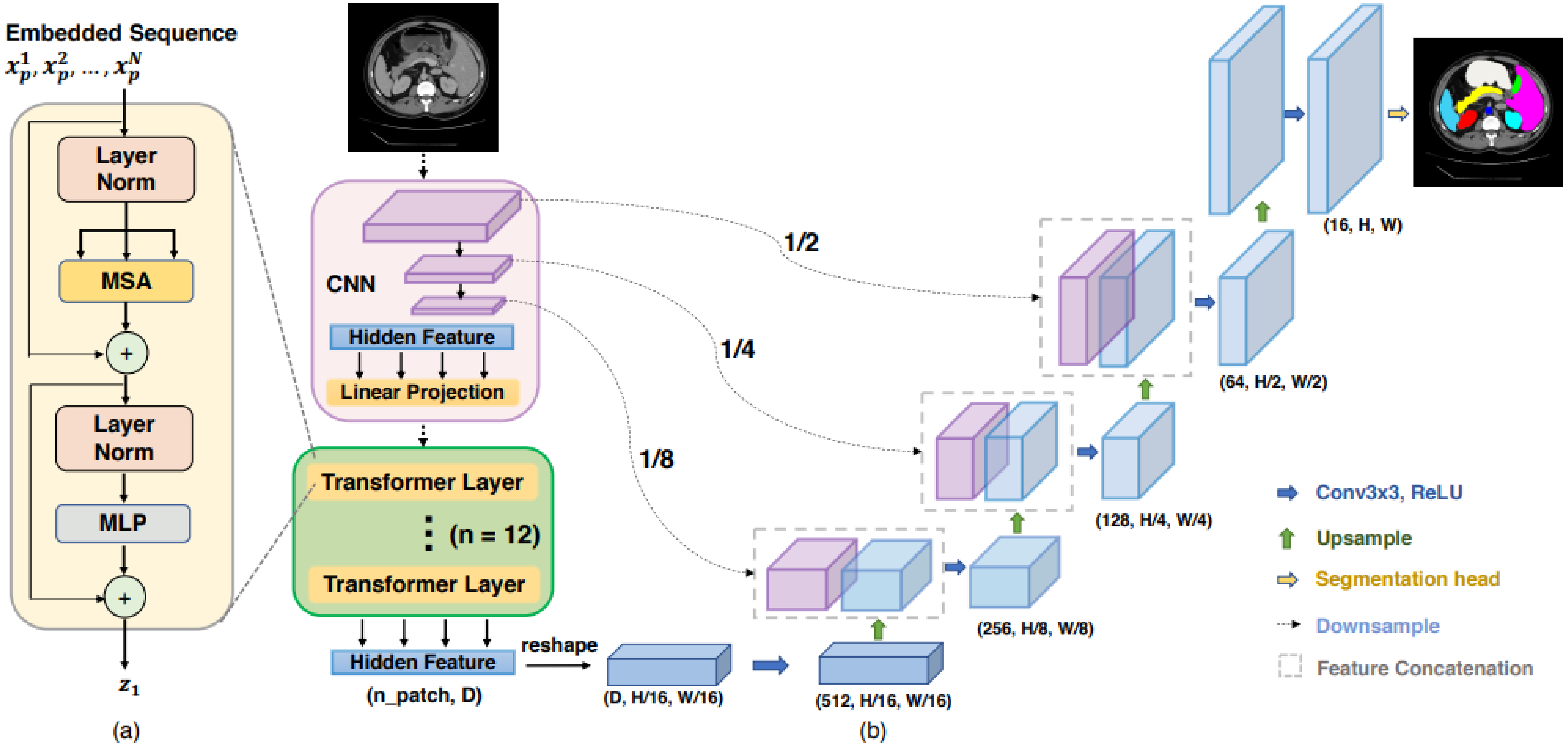


Fig. 2. Swin transformer block.

"Swin-unet: Unet-like pure transformer for medical image segmentation." European conference on computer vision. Cham: Springer Nature Switzerland, 2022

TransUnet



"Transunet: Transformers make strong encoders for medical image segmentation." arXiv preprint arXiv:2102.04306 (2021)

Summary

Model	Pros	Cons
FCT	• Outperforms existing models • Compact, accurate and robust • Able to reduce model parameters • five times smaller than nnFormer and three times smaller than TransUNet and LeViT-Unet	• Still relies on CNNs which may limit its ability to handle complex and diverse data distributions. • Uses fixed number of attention layers and heads which may not be optimal for different datasets
MERIT	• Captures multi-scale and multi-resolution features • Improves generalizability using self attention at multiple scales • Incorporates attention-based decoder	• Requires more computational resources and memory due to multi-scale and cascade design
nnFormer	• Able to model both local and global dependencies in 3d data • Exploits the advantage of convolutional and self attention operation in an interleaved manner	• High computation cost • Complex architecture and hard to interpret behaviour and results
SwinUnet	• Benefits of Swin Transformer and Unet • Enables multi-scale feature learning	• Requires pre-trained swin transformer model • sensitive to window size • require longer training time due to patch merging and splitting
TransUnet	• Employs benefits of both transformer and CNN	• requires pre-trained ViT models as the backbone of the Transformer encoder

Our Initial Approach

Medical Image Segmentation on Automatic Cardiac Diagnosis Challenge (ACDC)

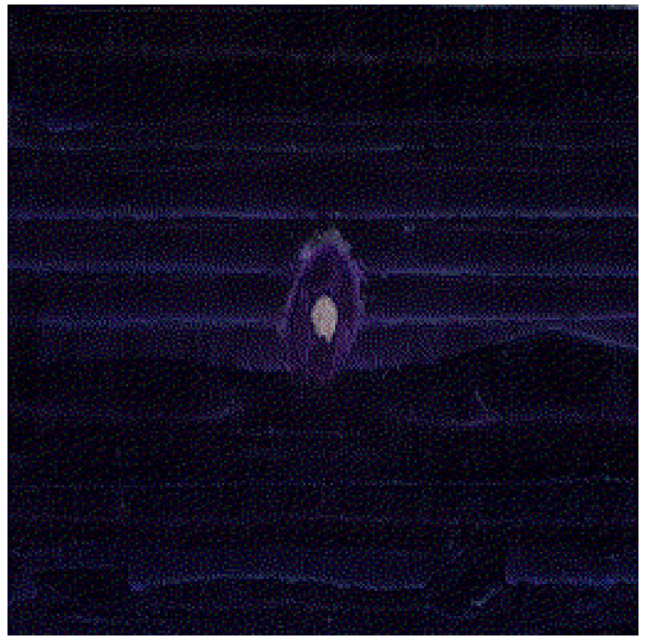
Rank	Model	Avg DSC	Paper	Code	Result	Year	Tags
1	**FCT**	93.02	The Fully Convolutional Transformer for Medical Image Segmentation			2022	
2	**MERIT**	92.32	Multi-scale Hierarchical Vision Transformer with Cascaded Attention Decoding for Medical Image Segmentation			2023	
3	**nnFormer**	92.06	nnFormer: Interleaved Transformer for Volumetric Segmentation			2022	
4	**TransCASCADE**	91.63	Medical Image Segmentation via Cascaded Attention Decoding			2022	
5	**PVT-CASCADE**	91.46	Medical Image Segmentation via Cascaded Attention Decoding			2022	
6	**SwinUnet**	90.00	Swin-Unet: Unet-like Pure Transformer for Medical Image Segmentation			2021	
7	**TransUNet**	89.71	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021	
8	**MISSFormer**	87.9	MISSFormer: An Effective Medical Image Segmentation Transformer			2021	
9	**R50-ViT-CUP**	87.57	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021	
10	**R50-AttnUNet**	86.75	TransUNet: Transformers Make Strong Encoders for Medical Image Segmentation			2021	

Outline

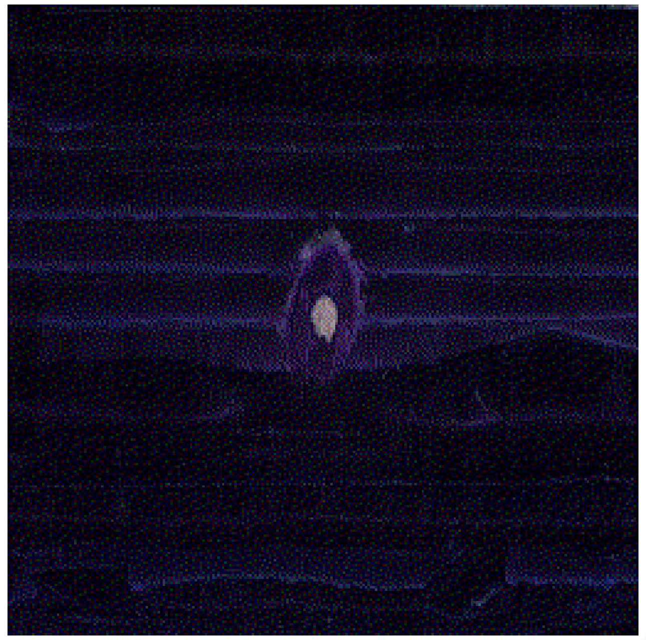
1. Recap
2. Approaches & Technical Methods
3. **Implementation**

Implementation

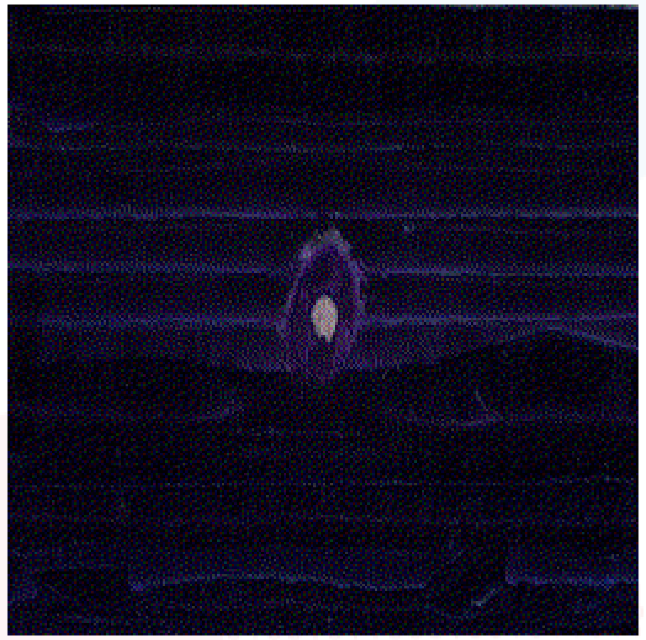
Data Preprocessing



VK-Male



VK-Male with Musclular Mask



VK-Male with Skeletal Mask

- Using body slice images of a full-body male (total of 1,702 images) and bone/muscle segmentation masks.
- Bone and muscle segmentation masks consist of 191 and 265 subcategories, respectively.
- For each consecutive slice, a musculoskeletal segmentation mask was created by integrating bone subcategories and muscle subcategories respectively.

Implementation

Dataset

- *Visible Korean Body Slice (1702 in total)*
- *Preprocessed GT Masks*
- *6:2:2 (train:validation:test)*

Performance Measure

Dice coefficient

Training Setting

RTX 4090

Tensorflow 2.0

Pytorch

Thank You!