CS-726 Medical Image Computing

Parallelized TransCNN Encoder for Medical Image Segmentation

Shruthi Akkala - 23M1074
Shivam - 23M1189

Motivation

- Image segmentation plays a pivotal role in medical image by enabling accurate analysis, diagnosis and treatment planning etc.,
- To improve the efficiency of segmentation by utilizing local and global information of images at different resolutions and also handling long term dependencies.
- 2D feature extraction by using CNN and Transformer

- UNet only models local information, lacking the capacity for global modeling.
- Swin Unet which uses only transformer block for encoding and decoding focus on only global features and lack local modeling.

Objective

- Image Segmentation using parallelized encoder structure, where one branch uses ResNet to extract local information from images, while the other branch uses Transformer to extract global information.
- **Input :** Slices of image for each scan
- Output : Segmentation of multi organ present in the slice.

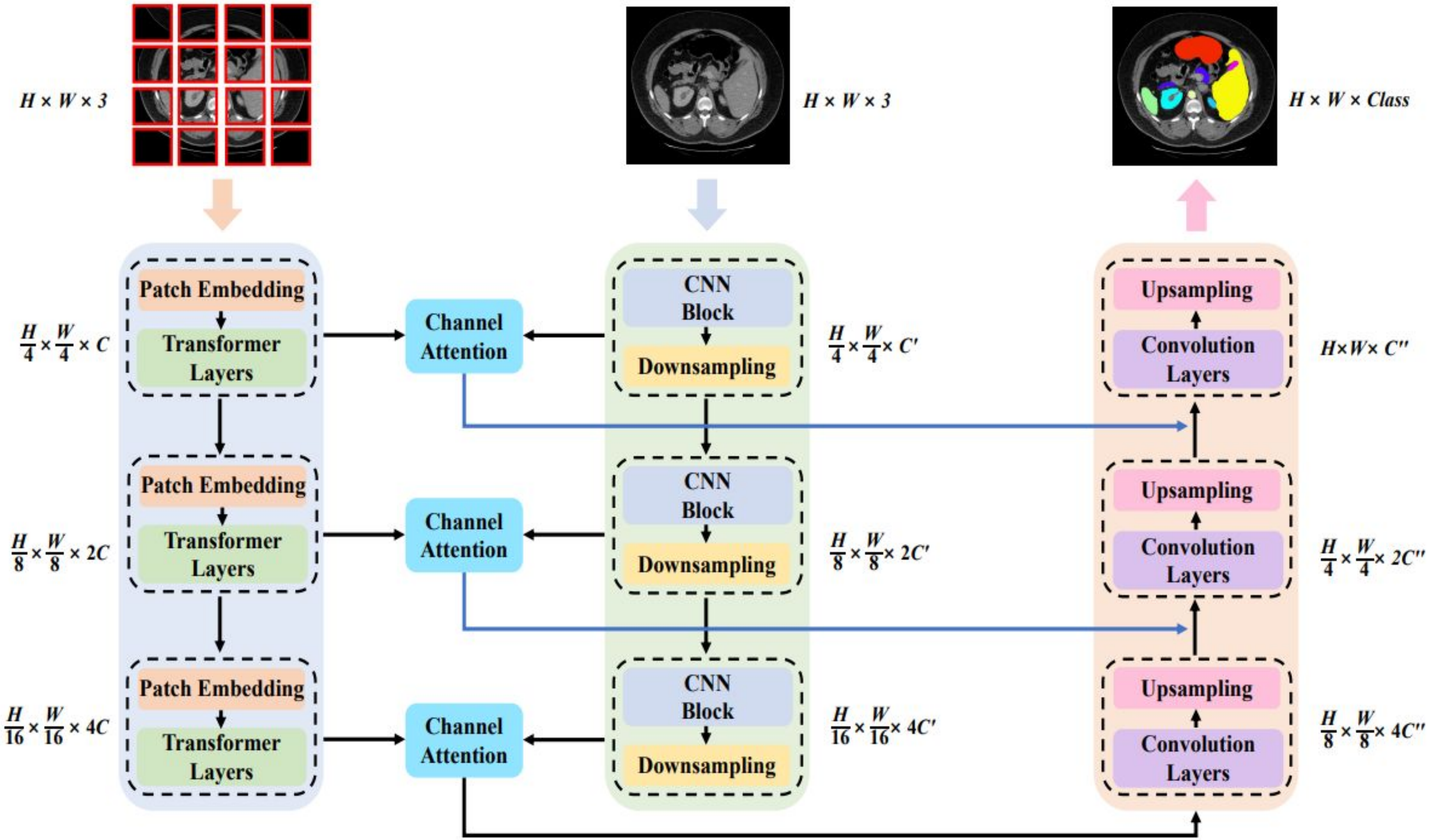
Implementation details:

- Loss function : Dice loss and cross Entropy lo $\mathcal{L} = \lambda_1 \mathcal{L}_{Dice} + \lambda_2 \mathcal{L}_{CE}$
- Image size – 224*224*3
- Learning rate – 0.01
- Epoch – 50
- Batch size – 2

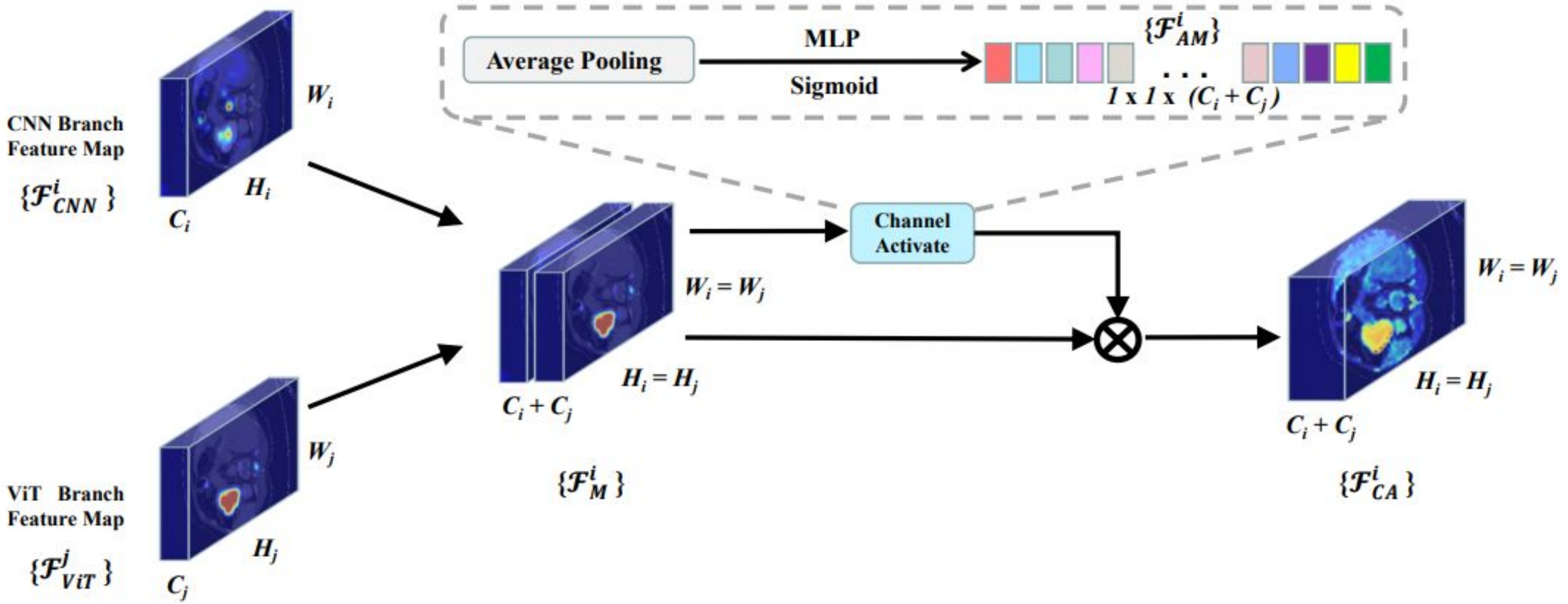
Proposed method

1. Our **U-shaped** medical image segmentation architecture involving a parallelized encoder consisting of the CNN and Transformer.
2. Where encoder can extract local and global features of different dimensions and efficiently integrate them to provide the decoder with rich pixel-level semantic information.
3. The Transformer branch introduces a pyramid structure to learn global information at different scales, while the CNN branch uses the same downsampling strategy to learn local information.
4. To achieve effective **multi-scale feature fusion,** a **channel attention** module is used to activate useful channel features and suppress unnecessary features.
5. Conducted experiments to find Dice Similarity Coefficient (DSC) and Hausdorff Distance (HD) for each organ in test dataset.

Block Diagram (ParaTransCNN)



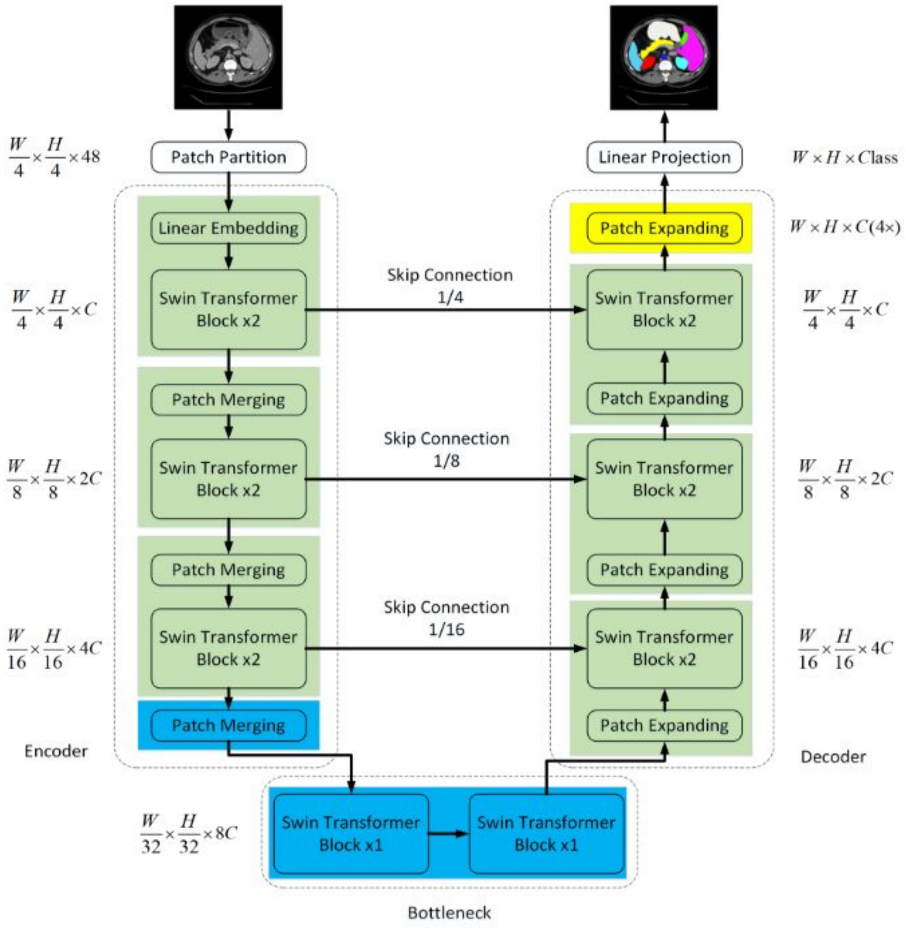
Channel Attention module



Related Work

Swin Transformer for Medical Image Segmentation:

- This model used Unet-like pure Transformer for medical image segmentation.
- The tokenized image patches are fed into the Transformer-based U Shaped Encoder-Decoder architecture with skip-connections for local global semantic feature learning.
- A symmetric Swin Transformer-based decoder with patch expanding layer is designed to perform the up-sampling operation to restore the spatial resolution of the feature maps.
- Disadvantage : No proper edge segmentation.



Implementation Details

- We collected the pre-processed data of **synapse multi organ dataset.**
- For the pyramid structure of Transformer we use the ViT model with patch size 4 for first transformer and 2 for next two. For each pyramid layer we used 3 transformer layers with 8 attention heads in each.
- For CNN block we used pretrained weights of **ResNet 32.**
- In channel attention to dual fusion of features from CNN and Transformer at each level we followed 3 steps i.e., **squeeze -> excitation -> fuse**
- In upsampling in Decoder blocks we used series of conv2d,batchnorm and relu functions

- Unlike Swin-Unet which focus only on global feature, ParaTransCNN focuses on both local and global features at different resolutions.

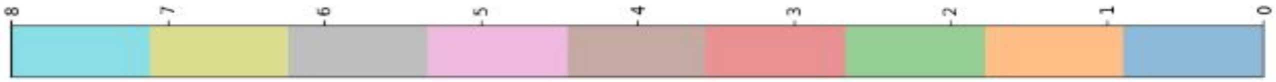
Dataset

- Synapse multi-organ datasets consists of **30 abdominal CT scans** with **3779 axial** contrast enhanced clinical CT images .
- Each volumetric sample is composed of **85 ~ 198 slices** of the same size of 512 × 512 pixels in the entire 3D data. All slices were resampled to 1mm x 1mm, with the HU value of [0,1].

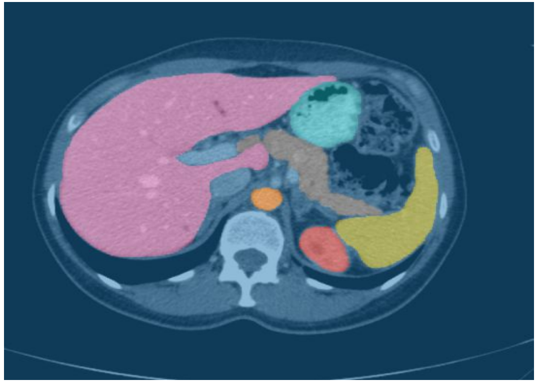
- Train dataset – 18 scans
- Test dataset – 12 scans
- Organs to segment :

No.	Label
1	Aorta
2	Gallbladder
3	Kidney(L)
4	Kidney(R)
5	Liver
6	Pancreas
7	Spleen
8	Stomach

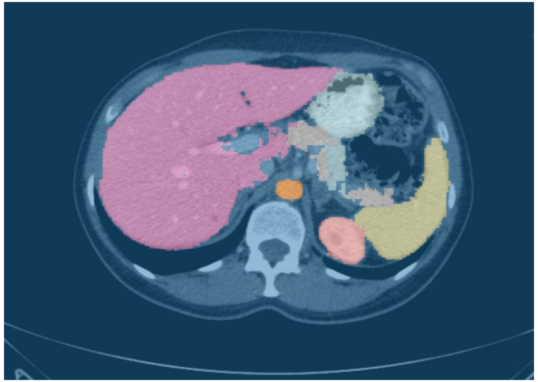
Output Segmentation and Comparing ParaTransCNN with Ground Truth and Swin U-Net



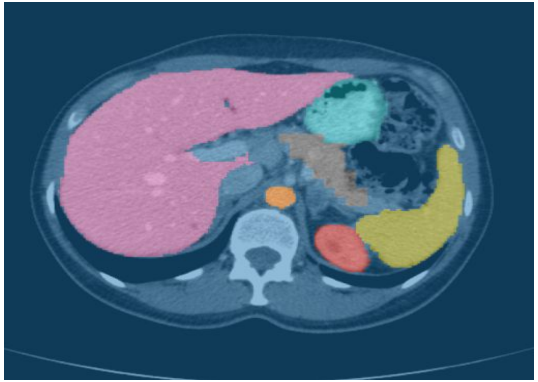
Case02 110 slice



Ground Truth

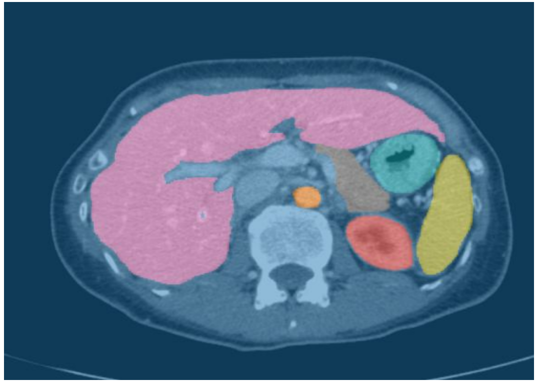


Swin Unet

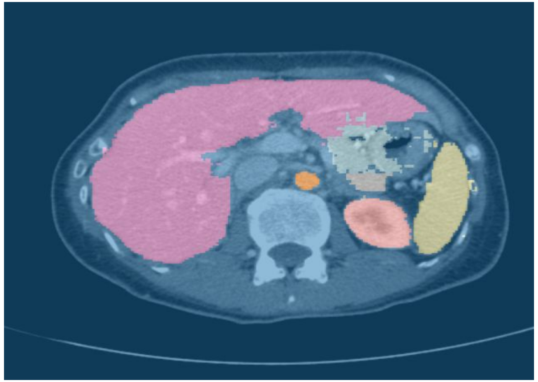


ParaTransCNN(proposed)

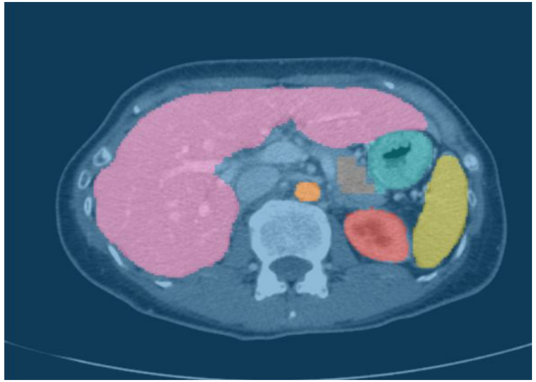
Case04 110 slice



Ground Truth

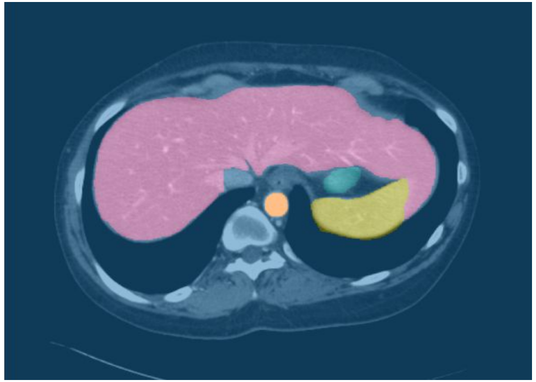


Swin Unet

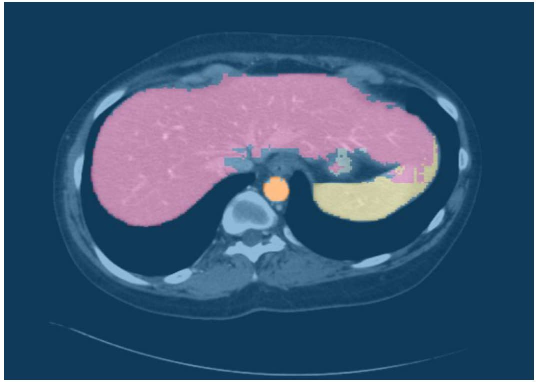


ParaTransCNN(proposed)

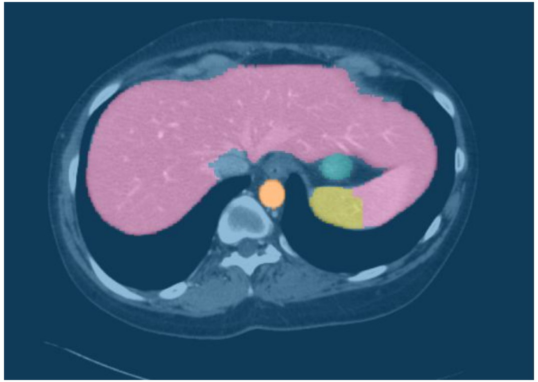
Case-08 130 slice



Ground Truth

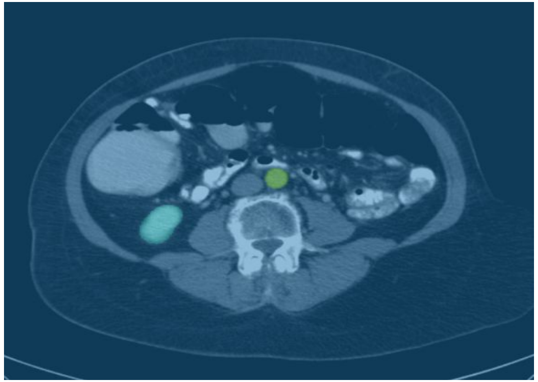


Swin Unet

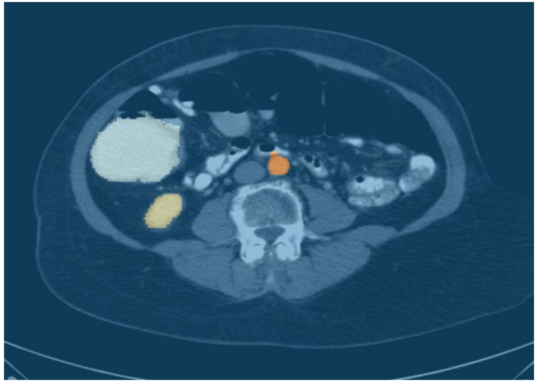


ParaTransCNN(proposed)

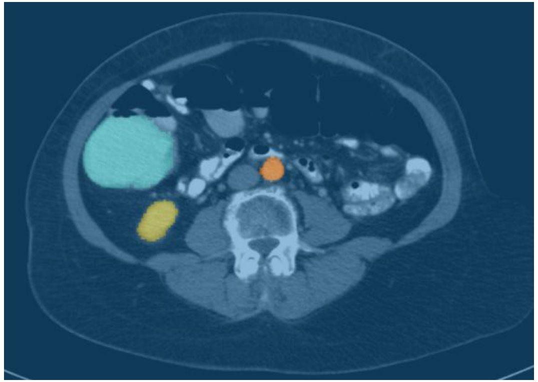
Case-32 75 slice



Ground Truth



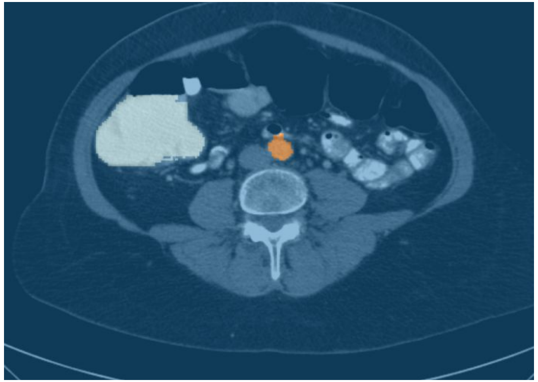
Swin Unet



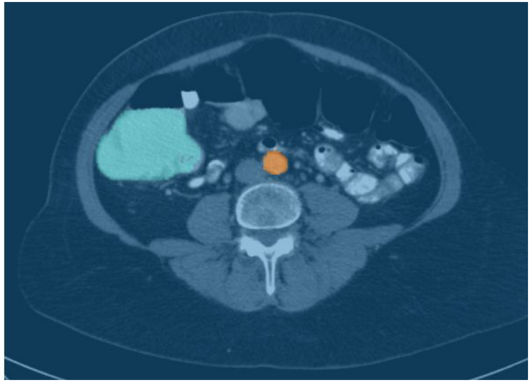
ParaTransCNN(proposed)

Case-32 70 slice

Ground Truth

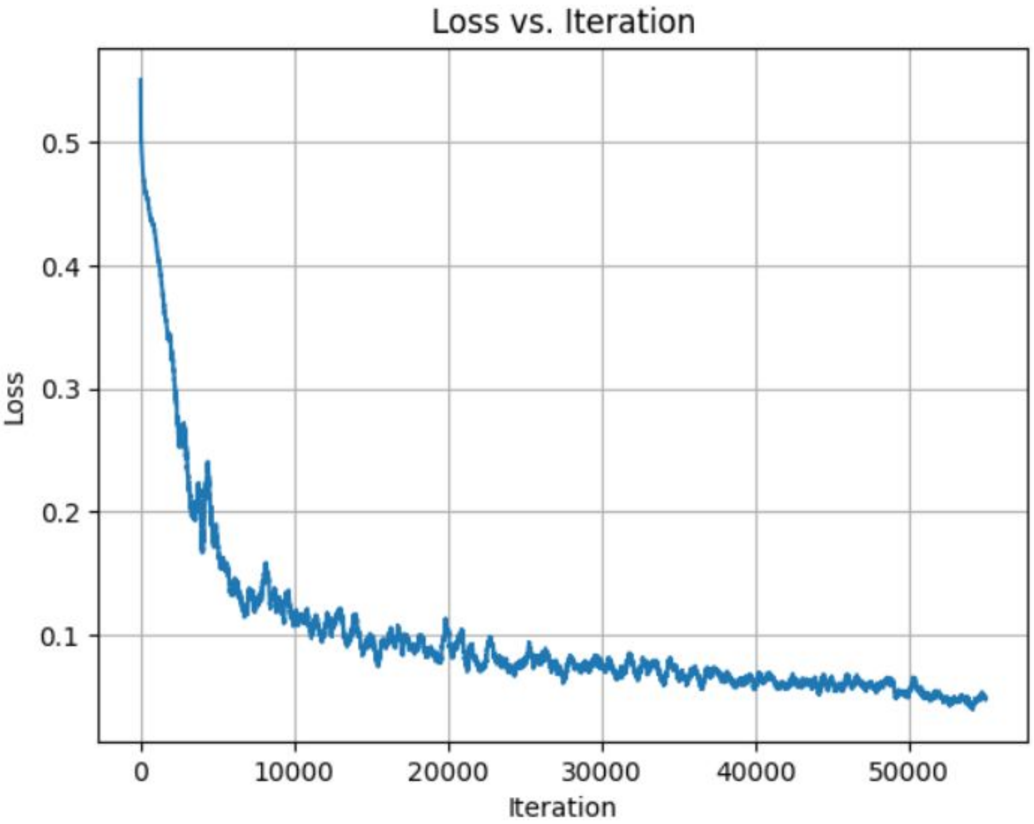


Swin Unet



ParaTransCNN(proposed)

Loss function vs iterations



Observations

We did the analysis by calculating
- Overall DSC for test dataset
- HD for test dataset
- Avg DSC for individual class(8)

Dice Similarity Coefficient (DSC)

$$DSC = \frac{2|A \cap B|}{|A| + |B|}$$

Hausdorff Distance (HD)

$$d_H(X, Y) = max\{d_{XY}, d_{YX}\}$$

$$= max\left\{\max_{x \in X} \min_{y \in Y} d(x, y), \max_{y \in Y} \min_{x \in X} d(x, y)\right\}$$

Model	Swin-Unet	Paralized TransCNN
Dim (C)	320	320
Epoch	50	50
Layer	1	[3,3,3]
DSC(%)↑	79.46	72.23
HD↓	13.31	16.82
Aorta	83.79	88.38
Gallbladder	50.0	0.00
Kidney(L)	82.83	83.71
Kidney(R)	79.96	76.29
Liver	92.88	93.58
Pancreas	58.05	44.25
Spleen	89.21	90.37
Stomach	71.84	73.84

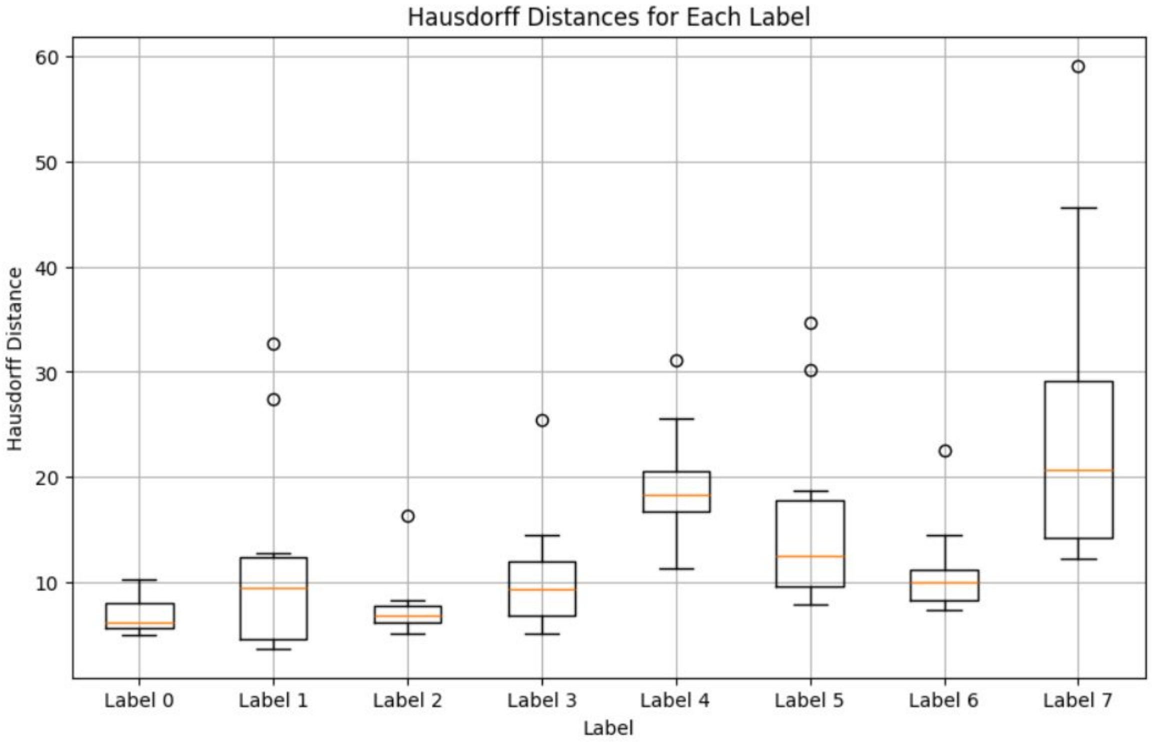
DSC–
test dataset 12 3-D slices

Test case:	Swin-Unet	Paralized TransCNN
08	65.34	61.16
22	89.22	78.13
38	82.17	72.70
36	84.75	76.31
32	86.68	77.22
02	85.78	77.08
29	76.47	66.53
03	67.78	59.35
01	78.72	80.34
04	74.14	74.15
25	87.26	67.55
35	75.71	76.23

Hausdorff distance For 12 abdominal CT test scans.

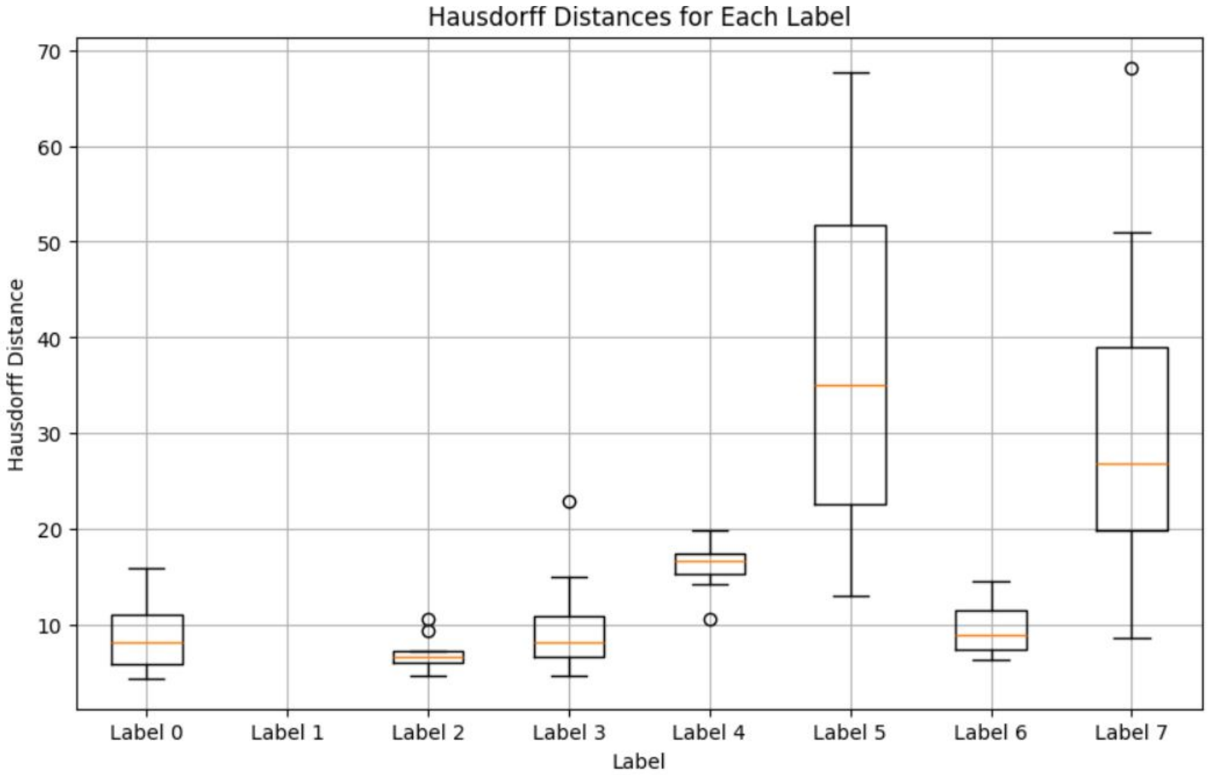
test.no:	Swin-Unet	Paralized TransCNN
08	16.58	15.00
22	12.08	13.45
38	11.79	31.21
36	08.72	13.34
32	13.30	17.06
02	7.07	15.30
29	15.39	24.52
03	12.37	23.43
01	17.12	13.00
04	23.00	16.40
25	13.08	9.43
35	9.22	9.70

Box Plot for HD distance of each label in Test cases



For Swin U-net

Box Plot for HD distance of each label in Test cases



For ParaTransCNN

Conclusion

- Less segmentation accuracy for Gallbladder.
- Time for training 11 hours.
- Segmentation for some organs are not good as we training for less epoch 50, we would get the result if we increase the epoch to 150.

Comparison:

- Compared to Swin U-net, ParaTransCNN segment more clearly and preserves the edges.
- DSC is slightly more than Swin U-net for most of organs
- HD distance is less in case of Swin U-net than ParaTransCNN

Thank you