MEDINFO23

8 – 12 JULY 2023 | SYDNEY, AUSTRALIA

之江实验室
ZHEJIANG LAB

Dual-Attention Model Fusing CNN and Transformer for Pancreas Segmentation

Yan Zhu

Postdoctoral,
Zhejiang Laboratory

zhuyan_me@
zhejianglab.com

AIDH IMIA

Background

- Pancreatic cancer:

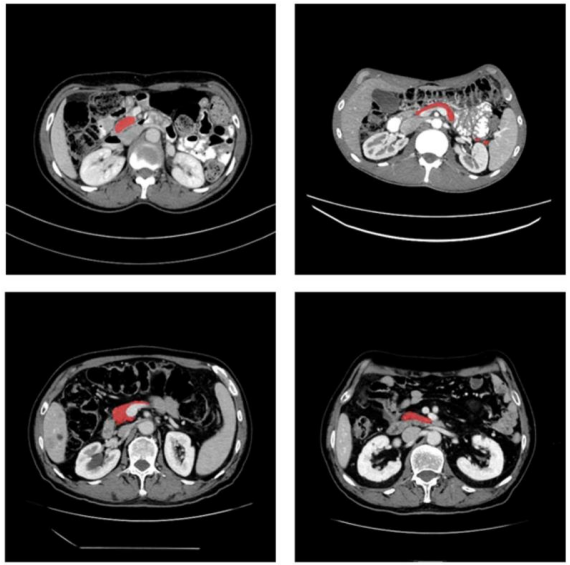
  fourth leading cause of cancer death in the U.S.A

  spread and metastasize rapidly

- pancreas characteristics:

  small size/irregular morphology/blurred borders/shaded

- **Precise segmentation is essential and challenging!!!**



Research Overview

- introduce a dual attention mechanism into the segmentation framework
- integrate CNN and Transformer structure to enhance the representation of pancreas-related features

Methods

Channel Attention Module

- 3D Squeeze and Excitation structure
- Embedded in the shallow layers
- Cascaded architecture
- Enriched the capacity of encoder

$$Z_c = F_{squeeze}(M_c) = \frac{1}{XYZ}\sum_{i=1}^{X}\sum_{j=1}^{Y}\sum_{k=1}^{Z} M_c(i,j,k)$$

$$M_{CA} = F_{excite}(Z_c, W) \cdot M_c = \sigma(g(Z_c, W))$$

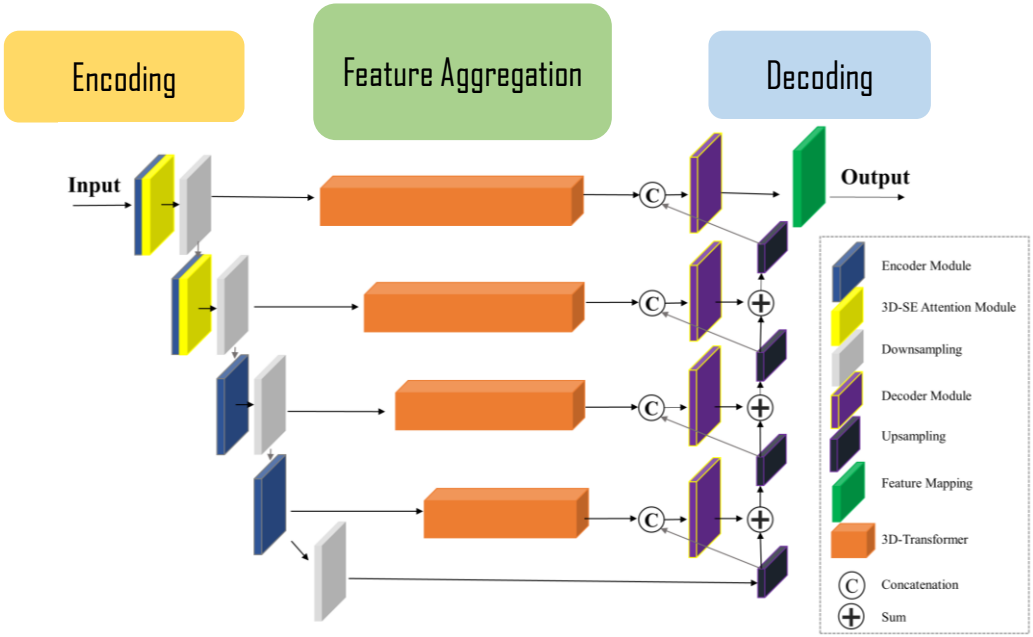
Spatial Attention Module

- Multiple vision Transformer
- Acquired global features
- Aggregated features at different scales
- Established long-term contextual sematic dependencies

Methods

Fusion Model

- Convolution for sample features
- Transformer for information interaction and fusion



$$M_{S_i} = F_{cat}(M_{SA}, M_{S_{i-1}})$$

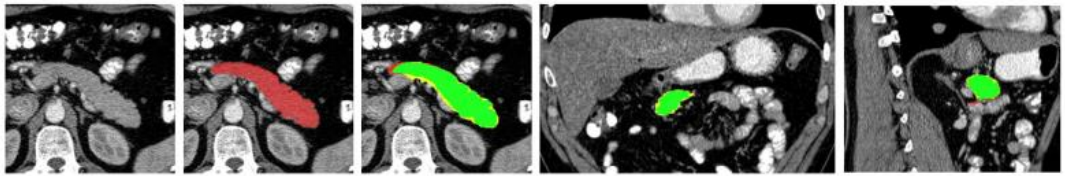
$$M_{out} = M_{S1} \cdot (1 + F_{up}(M_{S2} \cdot (1 + F_{up}(M_{S3} \cdot (1 + F_{up}(M_{S4} \cdot (1 + M_{in})))))))$$

Results

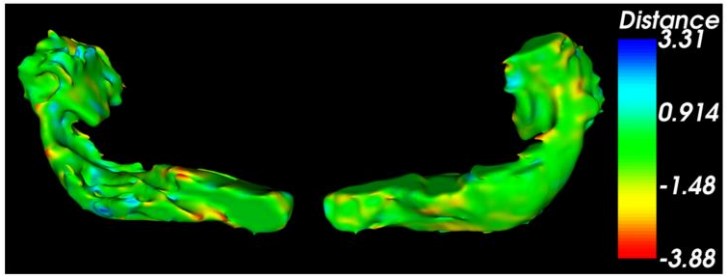
- Datasets: NIH-TCIA dataset

  82 CT images/ 512*512 resolution/ slices thickness 1.5mm~2.5mm

- Visualization segmentation results:



raw CT ground truth axial coronal sagittal



3D surface distance

Results

Table 1. Comparison of state-of-the-art methods on NIH-TCIA dataset.

Method	Mean DSC(%)	Max DSC(%)	Min DSC(%)
Zhu et al.[10]	84.50±4.86	91.45	69.62
Xia et al.[11]	84.63±5.07	91.57	61.58
Ma et al.[12]	85.32±4.19	91.47	71.04
Chen et al.[13]	85.22±4.07	91.36	**71.40**
Ours	**85.82±4.32**	**92.12**	70.54

Table 2. Performance comparison of models with different attention modules.

Model	Mean DSC(%)	mIoU(%)	Precision(%)	Recall(%)
3D ResUNet	75.98±11.16	49.5	68.35	64.26
3D ResUNet+ Channel Attention	83.13±5.88	61.26	78.63	73.50
3D ResUNet+ Dual Attention	**85.82±4.32**	**71.13**	**84.30**	**82.00**

Conclusion

- The dual attention-based fusion model outperforms state-of-the-art models with an average DSC of 85.82%± 4.52%
- Channel attention utilized the robust image comprehension capability of CNN to enhance feature representation
- Transformer structure aggregated features at different scales to establish long-term relationship among features in the spatial information level

Thank You for your listening