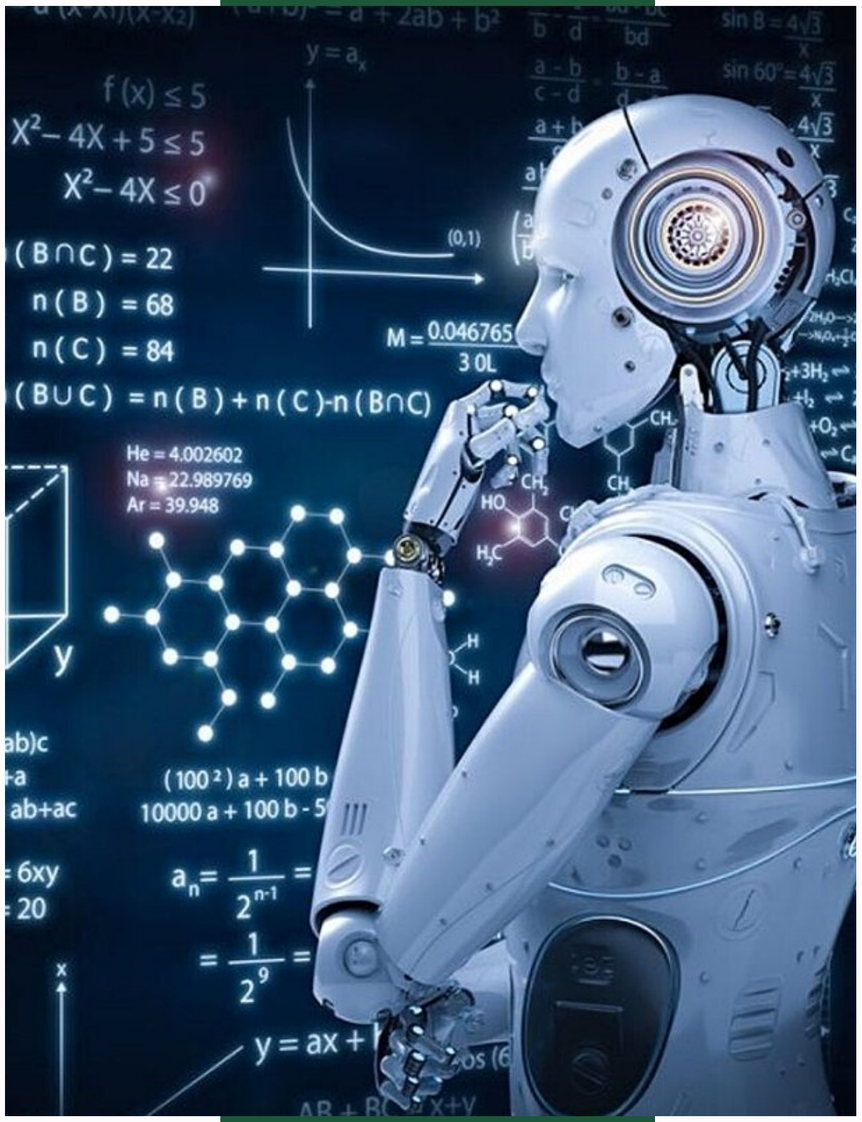
MEDICAL IMAGE SEGMENTATION THROUGH VISION TRANSFORMERS



PROBLEM STATEMENT

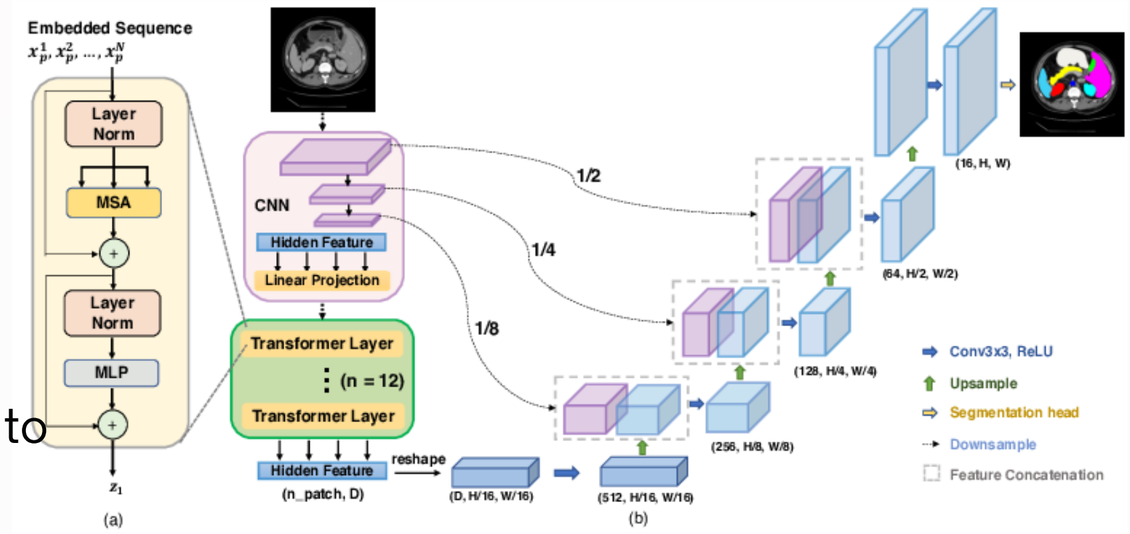
- To perform Segmentaion of 3D Biomedical images using TransUNet

BASE PAPER

TransUNet: Advancing Medical Image Segmentation through Vision Transformers.

KEY FEATURE IN OUR PAPER :

- Model uses Synapse multi-organ segmentation dataset..
- A hybrid CNN - Transformer architecture.
- For the encoder , UNET architecture is being used to process the images and to extract the features
- For the Decoder , Transformer is being combined through cross-attention with CNN features



ADVANTAGES

1) It allows us to leverage the intermediate high-resolution CNN **feature maps** in the decoding part.

2) The **hybrid CNN-Transformer** encoder performs better than simply using a pure Transformer as the encoder

3) Though built upon the 3D nnU-Net, can be easily modified to fit 2D tasks by simply switching the backbone model and reducing all operations back to 2D

LIMITATIONS

While U-Net has been successful, its **limitations** in handling **long-range dependencies** have prompted the **exploration of Transformer** as an alternative architecture, exemplified by our previously developed TransUNet, harnessing the combined strengths of U-Net and Transformers

1.Data Efficiency: Transformer-based models often require large amounts of data for training to achieve optimal performance.

2.Training Time: Training Transformer-based models like TransUNet can be time-consuming, especially for large datasets and complex architectures.

LITERATURE SURVEY

PUBLISHED BY	YEAR	AUTHOR	TITLE OF PAPER	ADVANTAGE	DISADVANTAGE
Xiaojun Xu et al.	2022	Xiaojun Xu et al.	Transformer-Based Biomedical Image Segmentation Using Patch-Level Training	- Adopts patch-level training strategies for improved memory efficiency and scalability. - Suitable for handling large-scale medical image datasets.	- Patch-level training may require careful balancing of patch sizes and overlap to avoid artifacts. - Implementation complexity may be higher compared to pixel-wise training strategies.
Chong Wang et al.	2021	Chong Wang et al.	Dense Dual U-Net for Biomedical Image Segmentation	- Incorporates dense connections between encoder and decoder paths to facilitate information flow. - Improves segmentation performance, particularly in scenarios with limited training data.	- Dense connections may increase computational overhead and memory requirements. - Increased model complexity may lead to longer training times.
C. Thongrong et al.	2021	C. Thongrong et al.	Adaptive Residual Dense Network for Brain Tumor Segmentation	- Integrates residual connections and dense blocks for enhanced feature extraction. - Promotes accurate delineation of tumor regions in MRI images.	- Model training may require extensive hyperparameter tuning to optimize performance. - Complex architectures may be more prone to overfitting, requiring careful regularization techniques.
Muhammad Usama et al.	2021	Muhammad Usama et al.	Combining Transformer and UNet Architecture for Brain Tumor Segmentation	- Integrates spatial information through UNet backbone. - Leverages self-attention mechanism of transformers for feature extraction.	- Hybrid architectures may introduce additional complexity. - Model training may require extensive hyperparameter tuning.

OVERVIEW

- **A hybrid CNN - Transformer architecture.**

1. The **Transformer encoder** where a **CNN encoder** is firstly used for local image feature extraction, followed by a pure Transformer encoder for global information interaction

2. **Transformer decoder** that reframes per-pixel segmentation as mask classification using learnable queries, which are refined through cross-attention with CNN features, and employs a coarse-to-fine attention refinement approach for enhanced segmentation accuracy.

DATASETS

The uwmgi-25d-tfrecord-dataset is a dataset that contains 2.5D TFRecord data for image segmentation tasks.

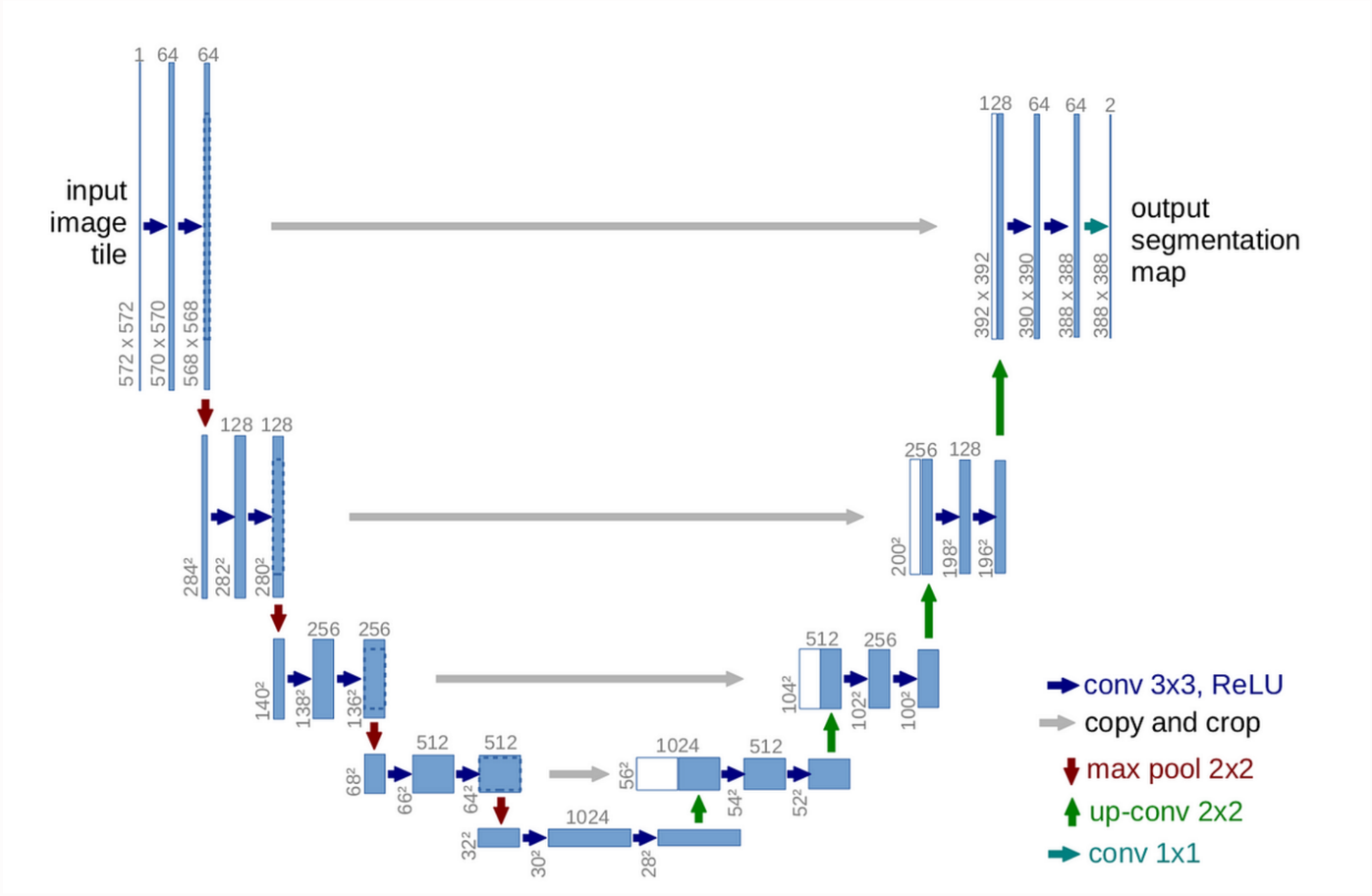
1. Name: UWMGI stands for the University of Wisconsin-Madison Gastrointestinal Tract Image Segmentation dataset.

2. Content: The dataset likely includes images of the gastrointestinal (GI) tract, that includes the stomach, small intestine, and large intestine (colon). These images are annotated for segmentation tasks, which involve labeling each pixel in the image with the corresponding class (such as stomach, small bowel, or large bowel).

3. Purpose: The dataset is intended for use in training and evaluating deep learning models for medical image segmentation tasks. By segmenting the GI tract images, researchers and practitioners can analyze and understand the structures and features of the digestive system.

.

U-Net Architecture



U-Net is a convolutional neural network (CNN) architecture

1.Contracting Path (Encoder):

- The left side of the U-Net consists of a series of convolutional and pooling layers.
- Each convolutional layer is typically followed by a rectified linear unit (ReLU) activation function.
- Max-pooling operations are used to downsample the feature maps.

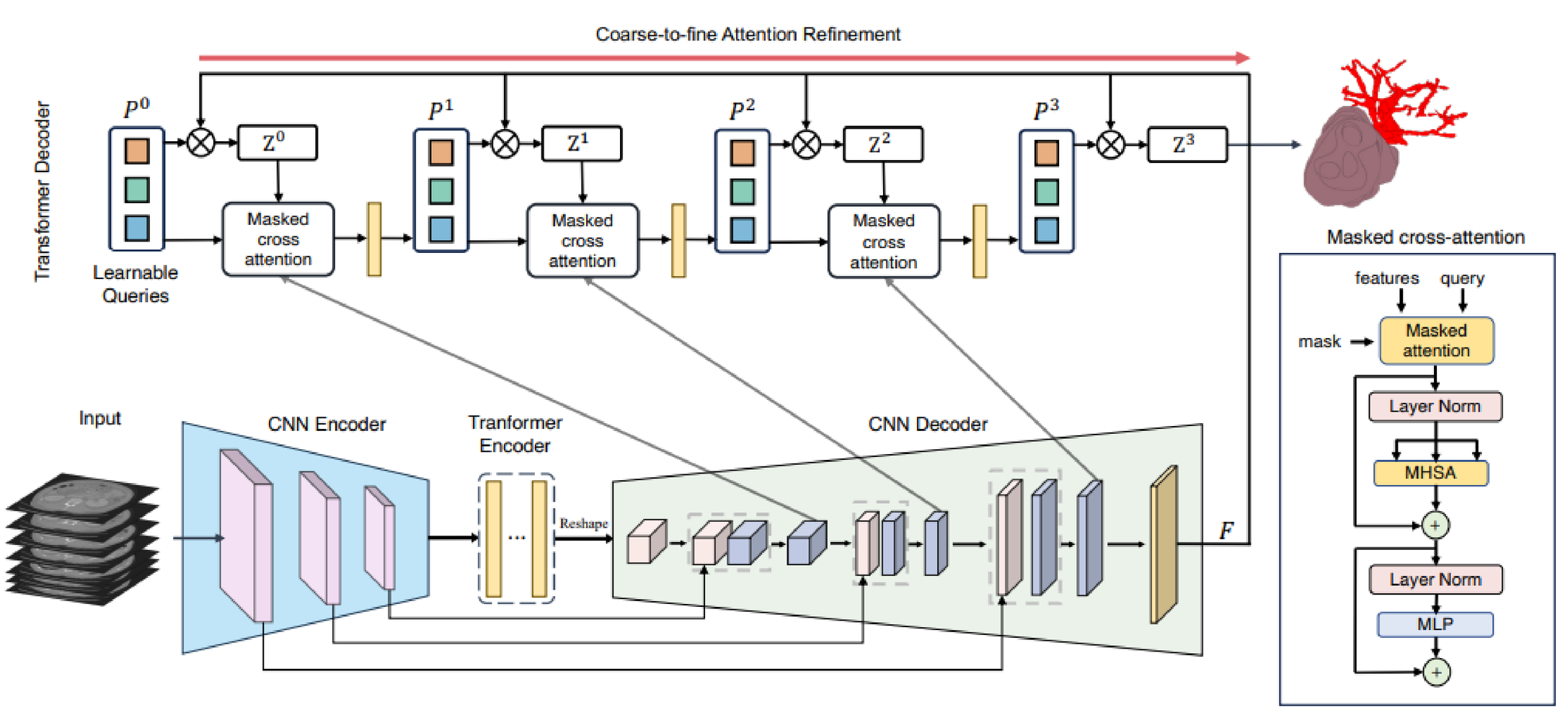
2.Expansive Path (Decoder):

- The right side of the U-Net consists of a series of upsampling and convolutional layers that gradually recover the spatial resolution of the feature maps while reducing the number of channels.
- Skip connections are established between corresponding layers in the contracting and expansive paths.

3.Final Layer:

- The final layer of the U-Net typically consists of a 1x1 convolutional layer followed by an appropriate activation function.

TransUNet Architecture



TransUNet Architecture

1.Transformer as Encoder

- Image sequentialization.
- Patch embedding.

  Map the vectorized patches into a latent denc-dimensional embedding space using a trainable Linear projection.

  Each Transformer layer consists of Multihead Self-Attention (MSA) and Multi-Layer Perceptron (MLP) blocks

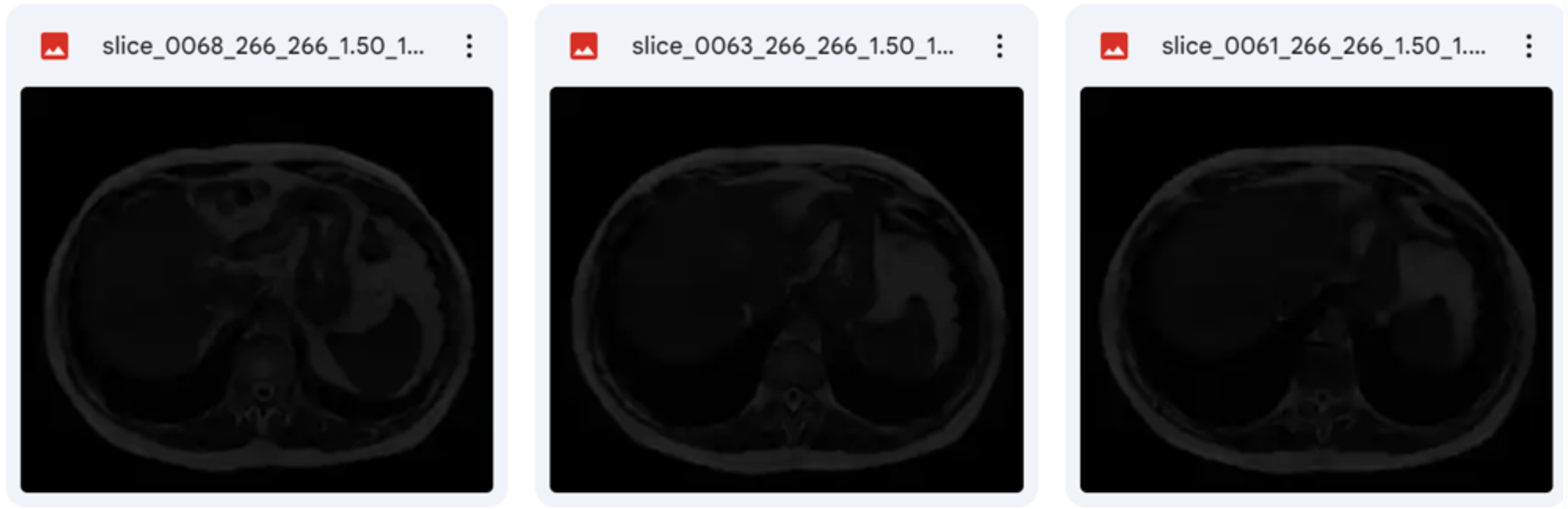
2.Transformer as Decoder

- Coarse candidate estimation:
- Transformer decoder:
- Coarse-to-fine attention refinement:

3.Encoder + Decoder

- We integrate both the Transformer encoder and the Transformer decoder into the 3D nnUNet model.

Input images



Validation of model using DICE and JACCARD

1.Jaccard Index (IoU):

The Jaccard Index measures the similarity between two sets by calculating the size of their intersection divided by the size of their union.

IoU= TP/(TP+FP+FN)

2.Dice Coefficient (F1 Score):

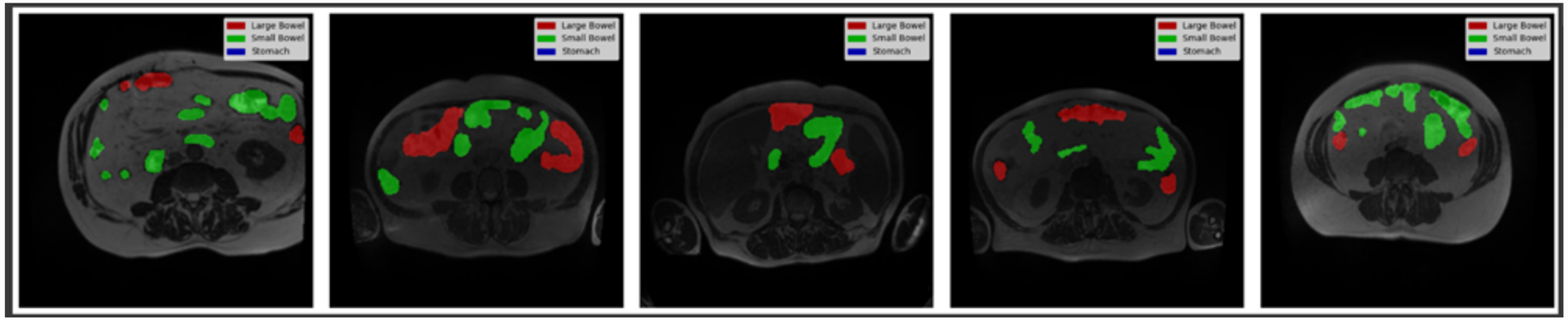
The Dice Coefficient measures the similarity between two sets by calculating twice the size of their intersection divided by the sum of their sizes.

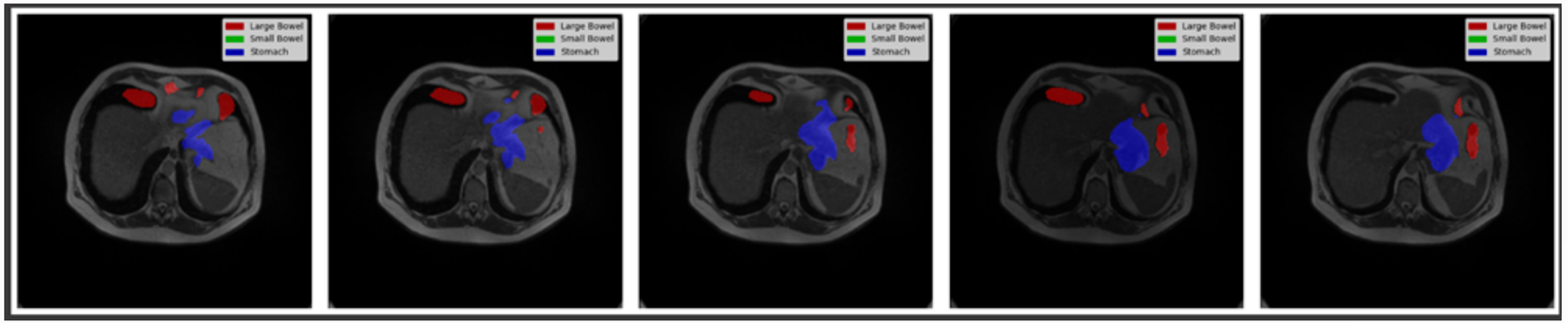
DICE = 2*TP/(2*TP+FP+FN)

Validation of model using DICE and JACCARD

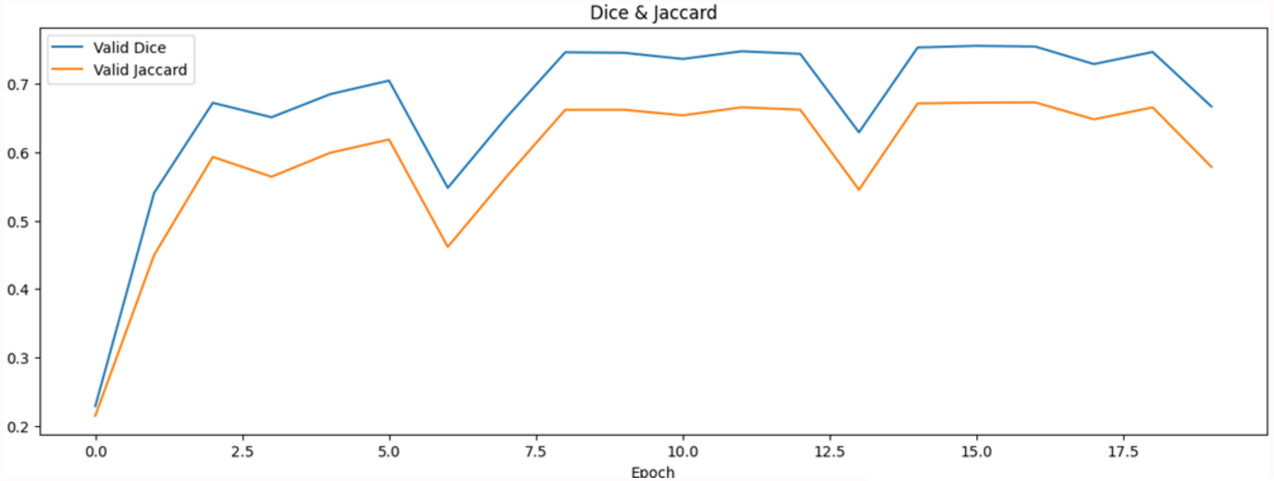
```
Valid : 100%|██████████| 18/18 [00:14<00:00,  1.25it/s, gpu_memory=10.78 GB, lr=0.00188, valid_loss=0.1163]
Valid Dice: 0.7543 | Valid Jaccard: 0.6727


Epoch 18/20Train : 100%|██████████| 176/176 [01:34<00:00,  1.87it/s, gpu_mem=5.21 GB, lr=0.00186, train_loss=0.0476]
Valid : 100%|██████████| 18/18 [00:15<00:00,  1.18it/s, gpu_memory=10.85 GB, lr=0.00186, valid_loss=0.1329]
Valid Dice: 0.7287 | Valid Jaccard: 0.6480


Epoch 19/20Train : 100%|██████████| 176/176 [01:37<00:00,  1.81it/s, gpu_mem=5.20 GB, lr=0.00185, train_loss=0.0480]
Valid : 100%|██████████| 18/18 [00:15<00:00,  1.18it/s, gpu_memory=10.78 GB, lr=0.00185, valid_loss=0.1287]
Valid Dice: 0.7463 | Valid Jaccard: 0.6654


Epoch 20/20Train : 100%|██████████| 176/176 [01:35<00:00,  1.83it/s, gpu_mem=5.21 GB, lr=0.00183, train_loss=0.0466]
Valid : 100%|██████████| 18/18 [00:15<00:00,  1.19it/s, gpu_memory=10.85 GB, lr=0.00183, valid_loss=0.1734]
Valid Dice: 0.6667 | Valid Jaccard: 0.5786


Training complete in 0h 49m 3s
```

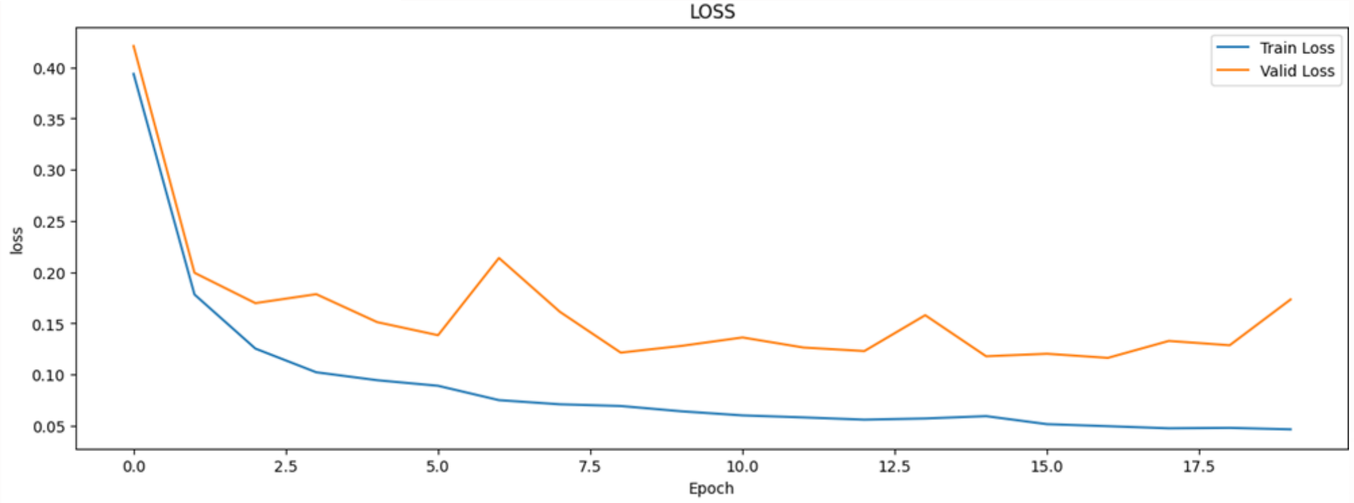
Segmented images as Output





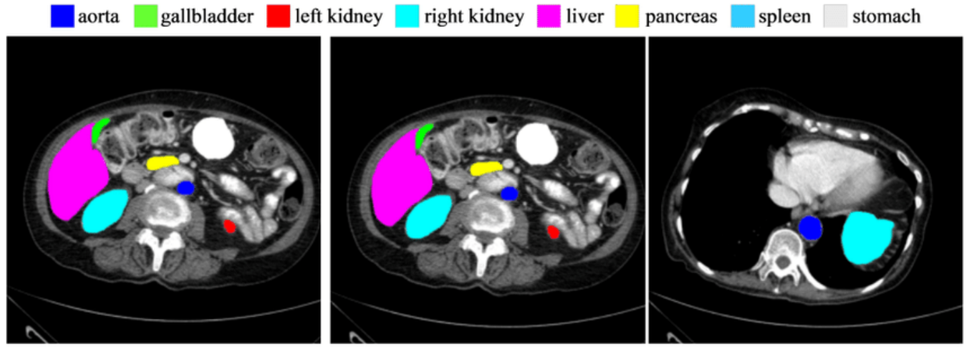
Scores and loses as graph





Applications of TransUnet

- Disease Diagnosis and Detection
- Treatment Planning
- Image-Guided Surgery
- Organ and Tissue Measurement
- 3D Reconstruction
- Neuroimaging
- Cardiac Image Analysis
- Drug Development and Research
- Automated Image Annotation:
- Population Health Studies



CONCLUSION

- TransUNet sets a new SOTA in medical image segmentation and the first to use Transformer within its architecture.

- In conclusion, medical image segmentation is a transformative technology with diverse applications that significantly enhance diagnostic accuracy, treatment precision, and overall healthcare outcomes. which provides better specification and highresolution image with fine details.

REFERENCES

TransUNet: Advancing Medical Image Segmentation through Vision Transformers. Jieneng Chen, Jieru Mei, Xianhang Li, Yongyi Lu, Qihang Yu, Qingyue Wei, Xiangde Luo, Yutong Xie, Ehsan Adeli, Yan Wang, Matthew Lungren, Lei Xing, Le Lu, Alan Yuille, Yuyin Zhou. 11 Oct 2023

1.Dosovitskiy, A., Beyer, L., Kolesnikov, A., Weissenborn, D., Zhai, X., Unterthiner, T., Dehghani, M., Minderer, M., Heigold, G., Gelly, S., et al.: An image is worth 16x16 words: Transformers for image recognition at scale. In: ICLR (2021)

2.. Fu, S., Lu, Y., Wang, Y., Zhou, Y., Shen, W., Fishman, E., Yuille, A.: Domain adaptive relational reasoning for 3d multi-organ segmentation. In: International Conference on Medical Image Computing and Computer-Assisted Intervention. pp. 656–666. Springer (2020)

3.Zheng, S., Lu, J., Zhao, H., Zhu, X., Luo, Z., Wang, Y., Fu, Y., Feng, J., Xiang, T., Torr, P.H., et al.: Rethinking semantic segmentation from a sequence-to-sequence perspective with transformers. arXiv preprint arXiv:2012.15840 (2020)

4.Schlemper, J., Oktay, O., Schaap, M., Heinrich, M., Kainz, B., Glocker, B., Rueckert, D.: Attention gated networks: Learning to leverage salient regions in medical images. Medical image analysis 53, 197–207 (2019)

THANK YOU