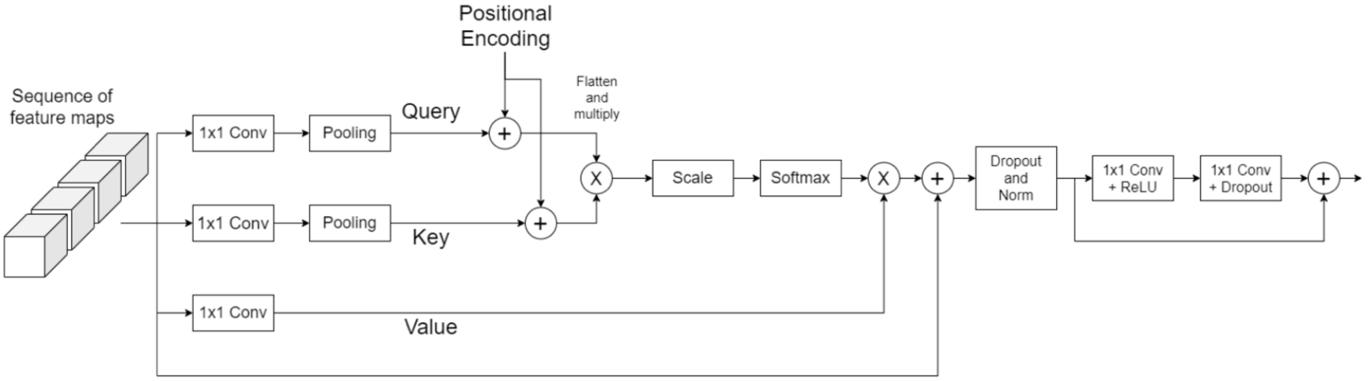
A Transformer-Based Network for Anisotropic 3D Medical Image Segmentation

Danfeng Guo[1] and Demetri Terzopoulos[1,2]
[1] University of California, Los Angeles, California, USA
[2] VoxelCloud, Inc., Los Angeles, California, USA

Abstract: Imaging anisotropy poses a critical challenge in applying deep learning models to 3D medical image analysis. Anisotropy downgrades model performance, especially when slice spacing varies significantly between training and clinical datasets. We propose a transformer-based model to tackle the anisotropy problem. It is adaptable to different levels of anisotropy and computationally efficient. Our model outperforms baseline models in 3D lung cancer segmentation experiments.

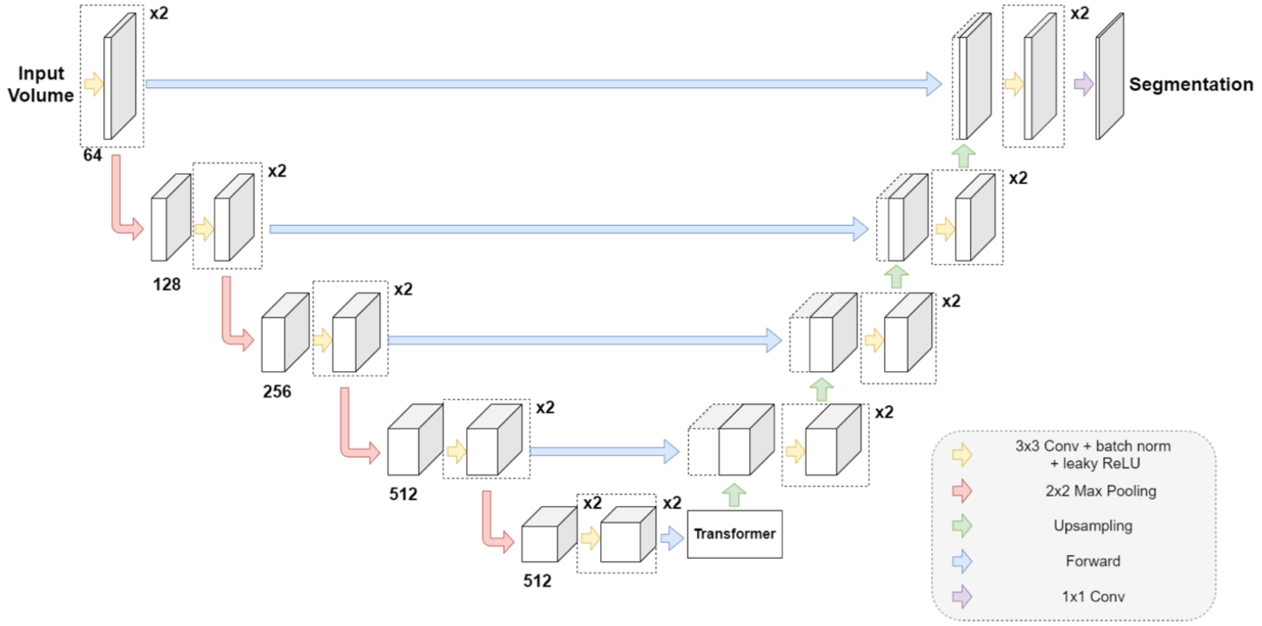


Architecture of the transformer Layer

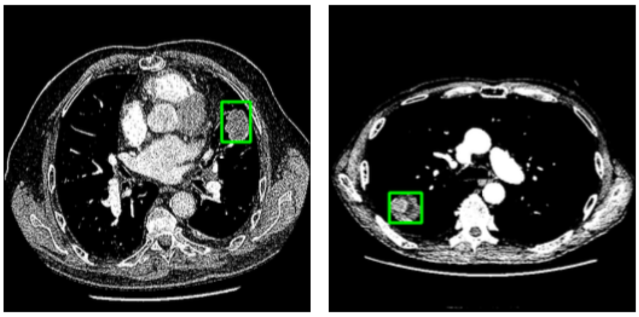
- **Developed by Vaswani *et al.* [1]**
- **Use queries and keys to capture the correlations between the slides.**
- **The new feature map of each slice is a weighted sum of its feature map and those of its neighboring slices.**
- **Use Positional Encoding (PE) to inject information about the sequence order.**

$$\text{Attention}(Q,K,V)=\text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V.$$

$$\text{PE}(i,j)=\begin{cases}\sin(j/w^{\frac{i}{d_k}}), & i\ \text{even},\\ \cos(j/w^{\frac{i}{d_k}}), & i\ \text{odd},\end{cases}$$



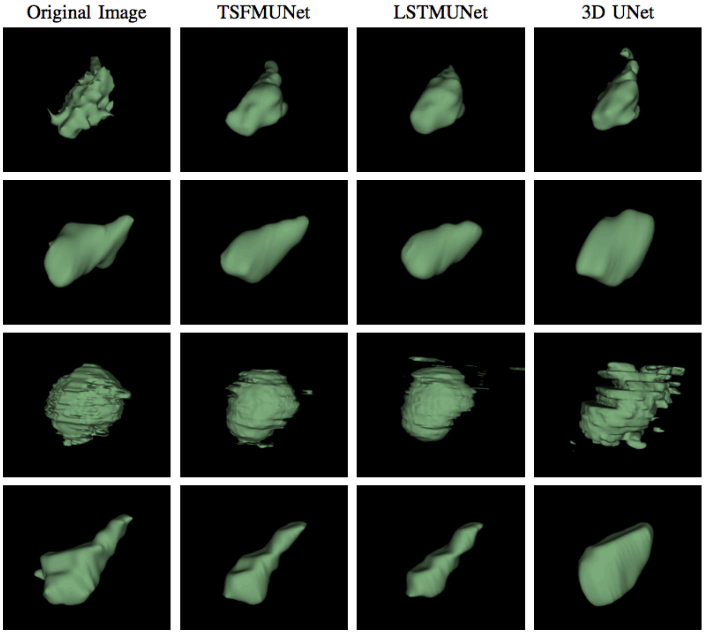
- **Experiment dataset is the lung cancer segmentation dataset from the Medical Segmentation Decathlon [2], including 20,707 lung CT slices from 63 subjects.**
- **Besides the original dataset, we created another dataset by re-slicing the original dataset into the same voxel spacing. Hence, we have two dataset of different distribution of voxel spacing.**
- **By testing the models on the original test data, we can compare the ability to adapt to variable spacing.**



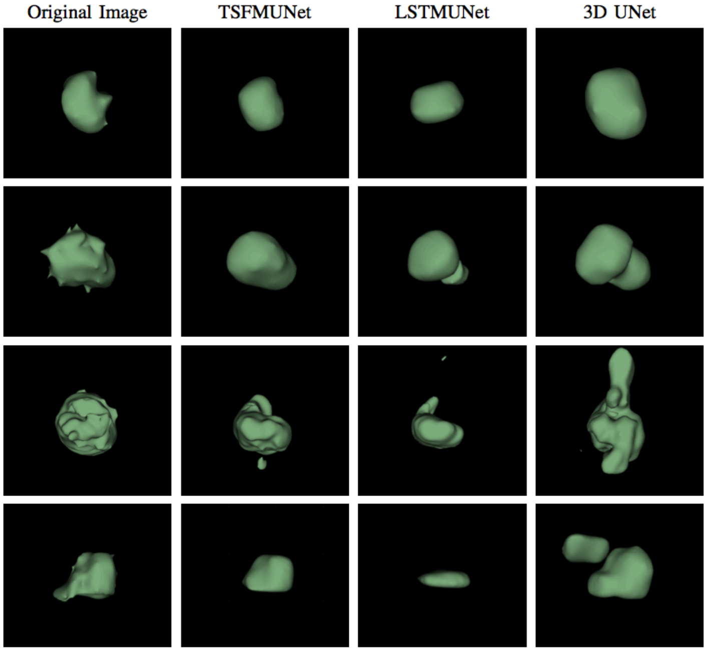
- **The result confirms that our transformer-based model outperforms the baseline models on both datasets.**
- **Comparing the results in the two tables, one sees performance drop: 3D U-Net>LSTMUNet>TSFMUNet**
- **The model trained on images with smaller inter-slice spacing will assume a tighter relationship between slices and the 3D kernels will have greater interaction between features from different slices.**

DICE SCORE COMPARISON (ORIGINAL DATA).

Model	Dice Score
TSFMUNet	**0.8717**
LSTMUNet	0.8573
3D U-Net	0.7744
2D U-Net	0.7309



Segmentation results on models trained on the original dataset



Segmentation results on models trained on re-sliced dataset

DICE SCORE COMPARISON (RE-SLICED DATA).

Model	Dice Score	Performance Drop
TSFMUNet	**0.8674**	**0.0043**
LSTMUNet	0.8217	0.0356
3D U-Net	0.7261	0.0483

References:

[1] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Advances in Neural Information Processing Systems*, 2017, pp. 5998–6008. 1, 2, 3

[2] A. L. Simpson, M. Antonelli *et al.*, “A large annotated medical image dataset for the development and evaluation of segmentation algorithms,” *CoRR*, vol. abs/1902.09063, 2019. 3