Transformers for Classification of Knee Diseases in MRI

Wang, Tianci; Brößner, Peter[1]; Khader, Firas[2]; Truhn, Daniel[2]; Radermacher, Klaus[1]

[1]Chair of Medical Engineering, Helmholtz Institute for Biomedical Engineering, RWTH Aachen University, Germany
[2]Department of Machine Learning and Musculoskeletal Imaging, Uniklinik RWTH Aachen, Germany

Introduction

Knee joint is the largest joint in the human body. Due to its complexity and importance, it is necessary to accurately and timely find the disease in the knee. Over 250 million people worldwide are estimated suffering from osteoarthritis (OA), and the prevalence of knee OA in adults aged 60 or over is about 10% in males and 13% in females [1]. Because of the complex structure and its importance, it is necessary to diagnose internal disease in a timely manner. Magnetic resonance imaging (MRI) is the most commonly used method. Manually analysis of MRIs can be time-consuming and prone to errors. Computer vision models have demonstrated their effectiveness in the medical field. In this thesis, vision transformers (ViTs) [2] will be used to find the new state-of-the-art method for diagnosing diseases in the knee, proving that it has a better performance than the traditional CNN models. And with the help of second loss on the attention layer, model can achieve a better performance.

Datasets

MRNet dataset: It consists of 1250 cases of knees, with 3 different diseases (abnormal, ACL and meniscus) as label. The most famous public dataset created by Standford University [3]. (Fig.1 is one example case)

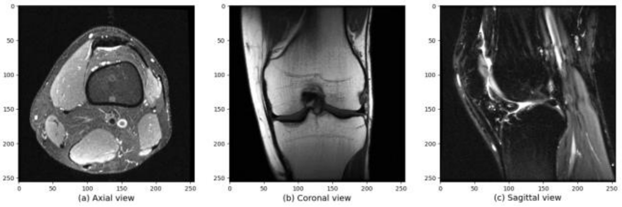

Figure 1: An example case from MRNet [3].

Praxis dataset: It has a similar structure with 3794 cases, the labels provide detailed information in the form of coordinates for numerous diseases. Here only 4 diseases are grouped for the further training. (Fig.2 is an ACL case inside)

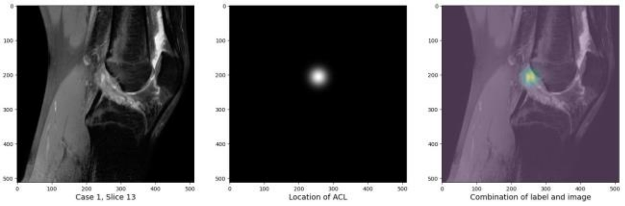

Figure 2: An example case of ACL from Praxis, the coordinates are highlighted as a Gaussian ball.

Methods

Multi-axis vision transformer (MaxViT) is a transformers model which is published last year. It achieves a better performance than any other ConvNets and transformers because of the linear-complexity of the grid attention, MaxViT is able to see globally throughout the entire network, even in earlier, high-resolution stages [4]. Here a new model combined by baseline model [3] and MaxViT is created for further training. Every single slice in knee MRIs will be inputted inside MaxViT separately and then pass the spatial attention structure to get the global features. After filtering them by max pooling layer, the features remain can be used to make the final classification. (Fig.3 is the framework of this model, spatial attention is optional here)

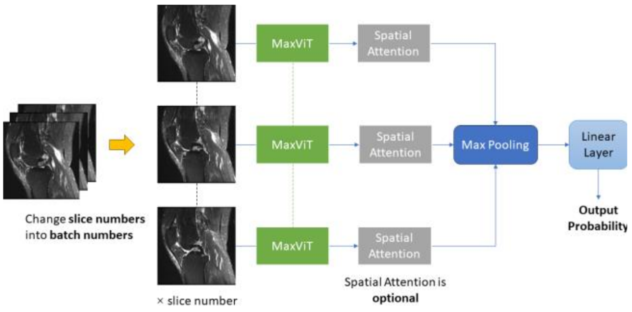

Figure 3: The framework of the combination of MaxViT [4] and baseline structure [3].

In order to better use the information from z direction in MRIs and the attention layer in ViTs, a new 3D ViTs is created. The second loss is used to compare with label in Praxis, help computer to focus on this area more. (Fig.4 is the structure, final loss is the combination of 2 losses)

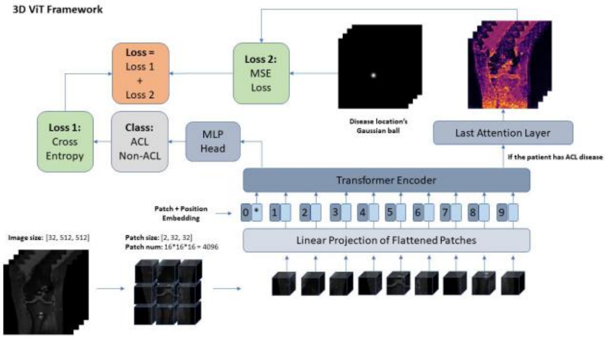

Figure 4: The framework of 3D ViTs.

Results

Method (MRNet)	*Abnormal*	*ACL*	*Meniscus*
Baseline	0.929	0.860	0.766
MaxViT	**0.945**	**0.903**	**0.782**

3D ViT (Praxis)	*ACL*
Only 1 loss	0.522
2 losses, with first attention	0.534
2 losses, with last attention	**0.575**

Discussion

The new structure with MaxViT can achieve a better result compared with baseline model. And with the help of the second loss, ViTs can improve the final performance.

References

[1] D. Primorac et al., “Knee Osteoarthritis: A Review of Pathogenesis and State-Of-The-Art Non-Operative Therapeutic Considerations,” Genes, vol. 11, no. 8, 2020, doi: 10.3390/genes11080854.

[2] A. Dosovitskiy et al., “An image is worth 16x16 words: Transformers for image recognition at scale,” arXiv preprint arXiv:2010.11929, 2020.

[3] N. Bien et al., “Deep-learning-assisted diagnosis for knee magnetic resonance imaging: development and retrospective validation of MRNet,” PLoS medicine, vol. 15, no. 11, e1002699, 2018.

[4] Z. Tu et al., “Maxvit: Multi-axis vision transformer,” in Computer Vision - ECCV 2022: 17th European Conference, Tel Aviv, Israel, October 23 - 27, 2022, Proceedings, Part XXIV, 2022, pp. 459 - 479.

UNIKLINIK RWTH AACHEN Klinik für Orthopädie
medi TEC Chair of Medical Engineering at Helmholtz-Institute of Biomedical Engineering
RWTH AACHEN UNIVERSITY