AsT: An Asymmetric-Sensitive Transformer for Osteonecrosis of the Femoral Head Detection (Student Abstract)

Haoyang Chen[1], Shuai Liu[1], Feng Lu[1], Wei Li[2], Bin Sheng[3], Mi Li[4], Hai Jin[1], Albert Y. Zomaya[2]

[1] National Engineering Research Center for Big Data Technology and System, Services Computing Technology and System Lab, Cluster and Grid Computing Lab, School of Computer Science and Technology, Huazhong University of Science and Technology, China

[2] Centre for Distributed and High Performance Computing, School of Computer Science, The University of Sydney, Australia

[3] Department of Computer Science and Engineering, Shanghai Jiao Tong University

[4] Tongji Hospital, Tongji Medical College, Huazhong University of Science and Technology

{hychen,shuailiu,lufeng,ortholimit,hjin}@hust.edu.cn,{weiwilson.li,albert.zomaya}@sydney.edu.au,shengbin@sjtu.edu.cn

Abstract

Early diagnosis of *osteonecrosis of the femoral head* (ONFH) can inhibit the progression and improve femoral head preservation. The radiograph difference between early ONFH and healthy ones is not apparent to the naked eye. It is also hard to produce a large dataset to train the classification model. In this paper, we propose *Asymmetric-Sensitive Transformer* (AsT) to capture the uneven development of the bilateral femoral head to enable robust ONFH detection. Our ONFH detection is realized using the self-attention mechanism to femoral head regions while conferring sensitivity to the uneven development by the attention-shared transformer. The real-world experiment studies show that AsT achieves the best performance of AUC 0.9313 in the early diagnosis of ONFH and can find out misdiagnosis cases firmly.

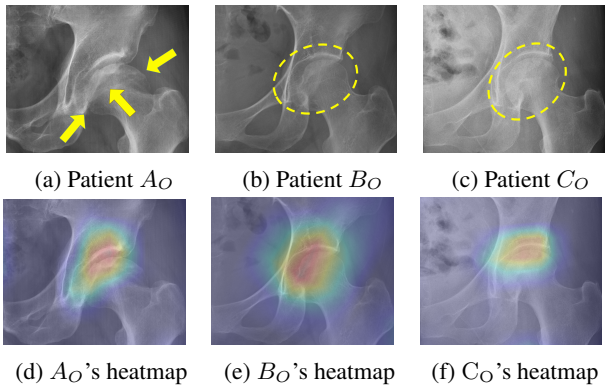
Introduction

Osteonecrosis of the femoral head (ONFH) is an intractable disease often caused by a compromised blood supply to the bone structure with significant clinical morbidity. A recent study shows 6 million ONFH patients worldwide and at least 110 thousand confirmed new cases yearly. Late treatment in ONFH patients could lead to collapse and hip replacement. Early detection with aggressive intervention is thus widely believed to help restore normalcy or halt the progression of ONFH.

The late stage of ONFH is easy to diagnose from plain radiographs, as shown in Figure 1(a). Early ONFH patients' cases are strikingly different since their pathological changes are still subtle. The gold standard for early ONFH diagnosis is MRI (*Magnetic Resonance Imaging*) scanning. However, MRI scanning is a scarce resource in hospitals and is only used to diagnose ONFH when the doctor has concerns about radiographs, as shown in Fgiure 1(b). Figure 1(c) shows a clean shot on plain radiographs, but the patient confirmed ONFH on MRI two months later.

Detecting early ONFH in radiographs is a *fine-grained visual classification* (FGVC) task. The ONFH detection needs to specify the target region to guide the learning for better results. Recently transformer-based methods, e.g., ViT (Dosovitskiy et al. 2020) and Swin Transformer (Liu et al. 2021), can yield impressive results without designated regions. They examine the correlation between different parts of the global image to find the most discriminative regions. Once confined, they can extract high-order features and make a reasonable classification. With sizeable datasets, such weakly supervised learning can attain excellent results through the autonomous discovery of discriminative regions. It is a bitter fact that the samples of ONFH are hard to accumulate. We collected only close to 800 plain radiographs of ONFH patients from 2012 to 2020 in Wuhan Tongji hospital, China.

We propose an *Asymmetric-Sensitive Transformer* (AsT) to guide the self-attention computation on the ONFH-Sensitive regions, thereby reducing the needed learning samples. Whether ONFH is bilateral or not, the progression of femoral head necrosis will develop asymmetrically. That is, the left and right ONFH do not progress the same. The self-attention computation can effectively capture ONFH if well utilizing this unique feature. We thus designed a new fine-grained image classification model to diagnose ONFH, particularly the early ONFH.



(a) Patient A_O (b) Patient B_O (c) Patient C_O

(d) A_O's heatmap (e) B_O's heatmap (f) C_O's heatmap

Figure 1: Comparison of patients diagnosed manually and by AsT. Patient A_O in (a) was diagnosed as ONFH by the orthopedist with a plain radiograph. The orthopedist misdiagnosed (b) and (c). Our AsT model picked these patients, and their lesions were correctly identified by the heatmaps as shown in (d)∼(f).

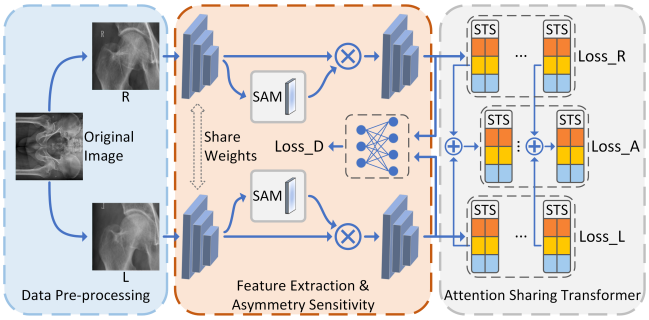


Figure 2: Overview of the AsT.

Methodology

Figure 2 provides the overview of the AsT consisting of three modules. The Data Pre-processing module first selects the bilateral femoral head's *region of interest* (ROI) on radiographs. Then, Feature Extraction & Asymmetry Sensitivity module extracts the ROI features and amplifies the asymmetry difference. The last module, Attention Sharing Transformer, shares features and attention to compare bilateral femoral heads for lesion detection.

Feature Extraction & Asymmetry Sensitivity

Once the femoral head's bilateral ROIs are obtained, AsT performs feature extraction using two convolutional blocks. It incorporates the spatial attention module (SAM in Figure 2) to highlight semantic pathological asymmetries and intensify attention to regions where pathology may be salient. This deep feature optimization can also reduce the interference of non-pathological regions. The process can be expressed as Eq. 1 and Eq. 2.

$$M_s = Sig(Conv(AvgPool(F_{out1}))) \tag{1}$$

$$F_{in2} = F_{out1} \cdot M_s \tag{2}$$

where $F_{out1} \in R^{H\times W\times C}$ is the output feature of the convolution block, $M_s \in R^{H\times W\times 1}$ is the obtained spatial attention map for F_{out1}, and $F_{in2} \in R^{H\times W\times C}$ is the input of the next convolution block reassigned by M_s and F_{out1}.

The pathological fact inspires us that differences in the ROIs can reveal the possibility of ONFH. This module thus performs further analysis on the features extracted from bilateral femoral heads to generate a difference Loss (Loss_D), which makes the model more sensitive to pathological asymmetry. The calculation procedure is shown in Eq. 3.

$$Loss\_D = \begin{cases} \frac{1}{n*n}\sum_{i}^{n}\sum_{j}^{n}\|F_L(x)_{ij} - F_R(x)_{ij}\| & y=0 \\ m - \frac{1}{n*n}\sum_{i}^{n}\sum_{j}^{n}\|F_L(x)_{ij} - F_R(x)_{ij}\| & y=1 \end{cases} \tag{3}$$

where y is the case label, m aims to control the upper limit of difference Loss, and F is the multichannel feature vector.

Experiments

We evaluated AsT on 1592 cases collected from Wuhan Tongji hospital in China from March 2012 to December 2020. We employed the ViT (Dosovitskiy et al. 2020), Swin Transformer (Liu et al. 2021), MCL (Chang et al. 2020), and SPS (Huang, Wang, and Tao 2021) as performance benchmarks. Table 1 shows the experimental results. It is easy to see that AsT outperformed all benchmarks in all aspects. We further investigated the plain radiographs of 1232 cases diagnosed to be ONFH-free between January 2021 to June 2022 from the same hospital. We found that radiologists misdiagnosed nine patients, and they all eventually developed ONFH. It is worth noting that three out of nine patients were confirmed shortly by MRI, but the other six were confirmed several months later. AsT could have identified these nine patients if it had been used.

Model	AUC	Precision	Recall	F1-score
ViT	0.6984	0.7086	0.6984	0.7004
Swin Transfomer	0.7048	0.7019	0.7048	0.7021
MCL	0.7431	0.7053	0.7000	0.6998
SPS	0.7852	0.7333	0.7321	0.7305
AsT(ours)	**0.9313**	**0.8527**	**0.8500**	**0.8501**

Table 1: The result of different methods to retrospectively classify the ONFH patients in Wuhan Tongji hospital, China.

Conclusions

This paper proposes a fine-grained classification model (AsT) to address the FGVC task of diagnosing ONFH using plain radiographs. Unlike existing methods, our approach focuses on the uneven development of the bilateral femoral head to enable ONFH detection effectively. Experimental results on the real-world data show that the AsT has superior performance and can find out misdiagnosed ONFH cases.

Acknowledgments

This work is supported by the Key Project of the National Natural Science Foundation of China (62232012) and the Hubei Big Data Analysis Platform and Intelligent Service Project for Medical and Health.

References

Chang, D.; Ding, Y.; Xie, J.; Bhunia, A. K.; Li, X.; Ma, Z.; Wu, M.; Guo, J.; and Song, Y.-Z. 2020. The devil is in the channels: Mutual-channel loss for fine-grained image classification. *IEEE Transactions on Image Processing*, 29: 4683.

Dosovitskiy, A.; Beyer, L.; Kolesnikov, A.; Weissenborn, D.; Zhai, X.; Unterthiner, T.; Dehghani, M.; Minderer, M.; Heigold, G.; Gelly, S.; et al. 2020. An image is worth 16x16 words: Transformers for image recognition at scale. *arXiv preprint arXiv:2010.11929*.

Huang, S.; Wang, X.; and Tao, D. 2021. Stochastic partial swap: Enhanced model generalization and interpretability for fine-grained recognition. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 620–629.

Liu, Z.; Lin, Y.; Cao, Y.; Hu, H.; Wei, Y.; Zhang, Z.; Lin, S.; and Guo, B. 2021. Swin transformer: Hierarchical vision transformer using shifted windows. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 10012–10022.