

무릎 골관절염 분류에서 다중 영역 특징 융합의 유효성 분석

이재훈, 유재천*

성균관대학교 전자전기컴퓨터공학과

jaehoonlee@skku.edu, *yoojc@skku.edu

An Analysis on the Validity of Multi-Region Feature Fusion in Knee Osteoarthritis Classification

Lee Jae Hoon, Yoo Jae Chern*

Department of Electrical and Computer Engineering, Sungkyunkwan University

요약

본 연구는 무릎 골관절염의 중증도 분류 정확도 향상을 목표로, 임상적 진단 과정을 모방한 다중 영역 특징 융합 네트워크의 실효성을 검증하였다. 단일 전체 이미지를 입력으로 사용하는 표준 ResNet-50 모델과, 5 개 영역의 특징을 병렬적으로 추출하고 단순 융합(Concatenation)하는 다중 경로 모델의 성능을 비교 분석하였다. 시험 결과, 표준 단일 네트워크(정확도 67.7%)와 다중 영역 특징 융합 네트워크(정확도 약 68.2%) 간에 통계적으로 유의미한 성능 차이는 관찰되지 않았다. 이는 다중 영역에 대한 명시적인 특징 추출 및 단순 융합 방식이, 이미 강력한 특징 추출 능력을 갖춘 심층 신경망의 성능을 개선하는 데 제한적임을 시사한다. 본 연구는 해당 접근법의 한계와 향후 어텐션 메커니즘과 같은 고도화된 융합 전략의 필요성을 제언한다.

I. 서론

무릎 골관절염은 높은 유병률을 가진 퇴행성 질환으로, Kellgren-Lawrence (KL) 등급[1]을 이용한 중증도 평가는 치료 방향을 결정하는 데 중요한 역할을 한다. 하지만 방사선 영상 판독은 관찰자의 숙련도와 주관성에 따라 편차가 발생할 수 있어, 일관성 있고 객관적인 평가를 위한 자동화 시스템의 필요성이 대두되고 있다. 실제 임상 환경에서 영상의학과 전문의는 관절 간격 협착이 주로 발생하는 내측(medial) 및 외측(lateral) 구획 등 특정 국소 영역을 집중적으로 관찰하여 진단한다는 점에 착안하여, 본 연구에서는 인공지능 모델 역시 다중 영역의 특징을 종합적으로 판단하도록 설계하면 단일 전체 이미지를 사용하는 것보다 우수한 성능을 보일 것이라는 가설을 설정하였다. 본 연구의 목적은, 5 개 영역에 대한 다중 경로 특징 융합 네트워크를 구축하고, 단일 전체 이미지 네트워크와 성능을 정량적으로 비교 분석하여 제안된 접근법의 유효성을 평가하는 것이다.

II. 본론

본 연구에서는 OAI 데이터셋의 무릎 X-ray 이미지[2]를 사용하여 KL 등급 분류 모델을 학습 및 평가하였다. 그림 1과 같이 각 이미지로부터 5 개의 관심 영역(ROI)을 추출하였다: (a)전체 이미지, (b)넓은 중앙부, (c)중앙부, (d)외측부, (e)내측부. 모든 이미지는 224x224 픽셀 크기로 정규화 되었다. 모델 학습에는 ImageNet 으로 사전 학습된 ResNet-50[3]을 기반 모델로 사용하였으며, Adam 옵티마이저, 초기 학습률 $1e-4$, 최대 50 에포크, 미니배치 크기 8 의 조건으로 미세 조정(fine-tuning)을 수행하였다.

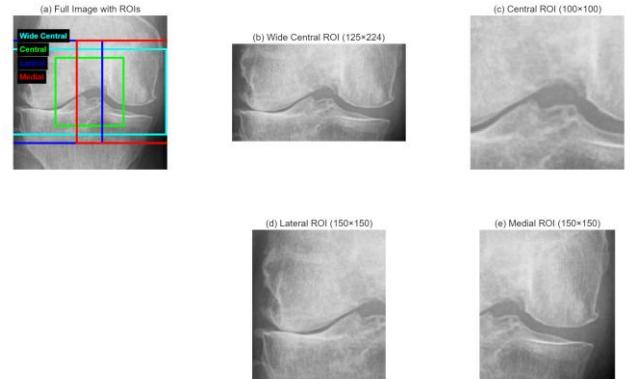


그림 1. 추출된 무릎 골관절 부위의 관심영역

실험은 두 단계로 진행되었다. 먼저, 각 단일 ROI로 학습시킨 모델들의 성능을 평가하였다. 표 1에서 보듯이, 전체 이미지를 사용한 모델(Accuracy: 0.6769)과 넓은 중앙부를 사용한 모델(Accuracy: 0.6757)이 다른 국소 영역(내측, 외측, 중앙) 모델들보다 월등히 높은 성능을 기록했다. 이는 진단에 있어 국소적 정보뿐만 아니라 이미지의 전반적인 맥락 정보가 중요함을 시사한다.

Model	Accuracy	Macro F1-Score	Macro Sensitivity	Macro AUC
Full Image	0.6769	0.6496	0.6448	0.8921
Wide-Central	0.6757	0.6711	0.6631	0.8991
Medial	0.5809	0.5192	0.5046	0.8382
Lateral	0.5707	0.4829	0.4654	0.8282
Central	0.5731	0.5028	0.5028	0.8161

표 1. 각 단일 관심 영역(ROI) 모델의 성능 비교

다음으로, 5 개 ROI 특징을 모두 사용하는 다중 경로 융합 모델(제안 모델)과 전체 이미지 모델(기준 모델)의 성능을 비교하였다. 그림 2 는 두 모델의 혼동 행렬을 보여준다. 기준 모델의 최종 정확도는 67.7%였으며, 제안 모델의 정확도는 68.2%로, 두 모델 간 성능에 통계적으로 유의미한 차이는 나타나지 않았다. 두 모델 모두 KL 등급 1 과 2 를 혼동하는 유사한 오류 패턴을 보였으며, 이는 해당 등급 간의 미세한 차이를 구분하는데 어려움이 있음을 의미한다.



그림 2. 기준 모델(좌)과 제안 모델(우)의 혼동 행렬

비교

III. 결론

본 연구를 통해, 임상적 지식을 반영한 다중 영역 특징 융합 접근법이 표준 단일 이미지 모델 대비 뛰어난 성능 향상을 보이지 않음을 확인하였다. 이러한 결과는 단순 특징 융합 방식의 내재적 한계와 심층 CNN 의 강력한 자체 특징 추출 능력에 기인하는 것으로 분석된다. 각 영역 특징의 중요도를 고려하지 않고 물리적으로 연결하는 단순 융합 방식은 핵심 정보의 희석을 유발할 수 있으며, ResNet-50 과 같은 모델은 이미 전체 이미지 내에서 진단에 중요한 국소적 특징을 식별하는 능력을 갖추고 있기 때문이다. 또한, 이미지 분할 과정에서 다리 축 정렬과 같은 거시적 정보가 손실되는 점도 성능 향상을 저해하는 요인으로 작용할 수 있다.

따라서 향후 연구는 단순 융합의 한계를 극복하는 데 집중해야 한다. 구체적으로, 각 영역 특징의 중요도를 데이터로부터 동적으로 학습하여 가중치를 부여하는 어텐션 메커니즘(Attention Mechanism)[4]의 도입을 제안한다. 어텐션 메커니즘은 진단에 결정적인 영역의 특징에 더 높은 가중치를 부여함으로써 모델이 핵심 정보에 집중하도록 유도할 수 있다. 이는 분류 성능의 향상뿐만 아니라, 각 예측에 대한 모델의 판단 근거를 시각화 하여 해석 가능성은 높이는 데에도 기여할 수 있을 것으로 기대한다.

참 고 문 헌

- [1] Kellgren, J. H., & Lawrence, J. S. (1957). Radiological assessment of osteo-arthrosis. *Annals of the rheumatic diseases*, 16(4), 494– 502.
- [2] Chen, Pingjun (2018), “Knee Osteoarthritis Severity Grading Dataset”, Mendeley Data, V1, doi: 10.17632/56rmx5bjcr.1

[3] He, K., Zhang, X., Ren, S., & Sun, J. (2016). Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition* (pp. 770–778).

[4] Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).