Classification of degenerative arthritis using Xception model in radiographic images

**Sang-min Lee and Namgi Kim**

Kyonggi university Department of computer science

Suwon, South Korea

[e-mail: d9249@kyonggi.ac.kr, ngkim@kyonggi.ac.kr]

\*Corresponding author: Sangmin Lee

***Abstract***

This paper presents an interactive method to manipulate reconstructed volumetric medical images from medical imagery using Augmented Reality(AR). The proposed system provides a possibility to reconstruct 3D dataset of patients in real-time and overlay the 3D models for manipulating the virtual objects in AR system. For this work, a sequence of CT images from human organ are acquired and segmented to detect region of interest(ROI). Subsequently using the consecutive slices of the segmented image, surface rendering is performed to construct polyhedral 3D surface. The generated 3D surface models are registered specific marker on AR system for the purpose of providing visual information of the organ model in a feasible fashion. By controlling the positions of markers in AR system with ARTookit, the view of the augmented 3D medical model can be dynamically varied.

**Keywords:** Knee osteoarthritis, Deep Learning

1. Introduction

몸은 자유로운 움직임을 위해 수 많은 관절로 이루어져 있다. 이러한 관절에는 뼈와 뼈가 직접 부딫치는 것을 방지하기 위해서, 사이에 연골(물렁뼈)이 있는데 이러한 연골은 정상적인 관절의 기능을 유지하는데 가장 중요한 조직이다. [1]

해당 연골의 중요도는 생활 속에서 계속 사용되기에 나이가 들어가며 마모의 정도가 심해지며, 누구나 연령이 높아감에 따라 퇴행성 관절염이 진행되고 퇴행성 관절염은 관절 연골의 국소적인 퇴행성 변화, 관절의 변형 등의 특징과 함께 지속적인 통증, 관절의 운동범위 제한 및 강직감을 나타내는 질환이다. 퇴행성 관절염은 연골의 사용이 많을수록 더 빠르게 진행되며, 우리나라의 경우 55세 이상에서는 약 80% 이상, 75세 이상에서는 거의 모든 노인환자에게서 방사선 검사상 퇴행성 무릎관절염 소견을 확인할 수 있다. 연골에 심각한 마모가 진행될 경우 정상적인 연골의 기능을 하지 못하게 되어 수술적인 방법과 비수술적인 방법을 통해 치료를 진행하게 된다. 비수술적 치료는 물리치료, 운동치료, 약물치료 등의 방법을 이용한다. 만약 통증이 극심하여 일상생활이 어렵거나, 6개월 정도의 비수술적 치료에도 통증의 호전이 보이지 않을 때, 관절의 구조적 변형이나 불안정성이 심하거나 운동범위의 제한이 심한 경우 등에는 무릎절골술, 인공관절 치환술과 같은 수술적인 치료를 시행하게 된다. 하지만 수술적 치료는 기본적으로 감염 및 이로 인한 재수술의 위험성이 있고, 특히 전치환술의 경우 시행 후 구조적 불안정성으로 인한 통증 및 기능 저하 등의 부작용이 수반되는 경우가 있으며, 치환물의 수명이 약 15년 정도로 제한되어 있어 시행 전 신중한 결정이 필요하다. 이와 같이 수술적 치료가 갖는 한계로 상대적으로 부작용이 덜하고 환자 개개인의 신체적 특성을 바탕으로 진단하고 치료하는 한의학에 대한 수요가 증가하고 있으나, 퇴행성 무릎관절염에 대한 한의학적 진단과 치료에 대한 과학적이고 논리적인 근거 확보 또한 절실하며, 이에 대한 근거가 확보되기 위해서는 임상현장에서 활용할 수 있는 표준화된 진단과 치료 방법의 개발이 필요하다. [2]

표준화된 진단 방식의 개발을 위해 기존 관절염 등급 진단 방식의 객관적 지표를 제공하여 의사가 보다 더욱 빠르고 정확한 진단을 내려 퇴행성 관절염으로 인한 연골의 마모를 사전에 예방하고자 한다.

퇴행성 관절염 진단의 기존 과정은 그림 1과 같은 방사선학 이미지를 촬영 후 결과 이미지를 통해 점진적으로 나타나는 관절 간격의 감소와 연골 아래 뼈의 음영이 짙어지는 경화 현상과 환자에게 설문조사를 통한 종합적인 소견을 의사를 통해 등급을 진단하였으며 이러한 검사진단에 있어서 과학적이고 논리적인 근거를 의사에게 제공하기 위해서 그림 1과 같이 시각적으로 보이는 변화 현상을 통해 퇴행성 관절염의 진단이 가능하기에 라벨링된 무릎 방사선학 이미지를 Deep Learning 모델에 학습시켜 퇴행성 관절염 등급의 객관적인 지표로의 정보 전달을 목표로 하였으며, 이를 위해서는 기존의 정확도보다 더욱 개선되어야 하기에 정확도를 주된 목표로 하였다.

본 연구에서는 무릎 관절의 연골 마모에 따른 등급 진단의 지표 중 가장 보편적으로 사용되는 등급 분류 지표인 K-L grade를 기준으로 삼았으며, 그림 1과 표 1과 같이 5단계로 구분하고자 하였다. 해당 연구 분야에 기존의 보편적으로 사용되었던 Deep Learning 모델의 SOTA(State-of-the-art) 성능이 약 70%를 웃돌며, 기존 연구에서는 성능을 중점적인 관점으로 설계되었던 모델들을 사용하였기에 많은 Parameter를 요했으며, 본 연구에서는 이를 개선하기 위해서 보다 경량화된 모델인 Xception을 연구에 활용하였으며, 기존 모델들에 비해 높은 정확도와 이와 같이 경량화된 모델을 사용하여 모델의 크기 축소를 통한 학습 및 추론 과정의 시간, 비용을 감소를 목표로 하였으며, 이를 통해 퇴행성 무릎 관절염 분류의 다른 연구보다 높은 정확도와 시간, 비용 감소를 통한 효과적인 등급 분류를 목표로하여 해당 연구를 진행하였다.

텍스트, 응시하는, 흐림이(가) 표시된 사진

자동 생성된 설명

**Fig. 1.** Stage of knee osteoarthritis

|  |  |
| --- | --- |
| 등급 | 증상 |
| K-L grade 0 | 정상 |
| K-L grade 1 | 정상에 비해 관절 간격이 좁아진 것으로 의심될 경우, 골극이 있을 수 있음 |
| K-L grade 2 | 관절 공간이 명확하게 좁아짐. 골극이 확실함 |
| K-L grade 3 | 관절 공간이 좁고 골경화와 골형태에 심한 변형이 관찰. 중증도의 다발성 골극 관찰 |
| K-L grade 4 | 관절 간격이 현저하게 좁아져 골형태에 심한 변형이 관찰 |

**Table. 1.** Kellgren lawrence grade 평가 지표

2. 관련 연구

그동안 퇴행성 관절염 진단에 Deep Learning을 활용한 연구는 많은 수의 Paramter를 요하는 Deep Learning 모델을 활용하여 방사선학 image Classification을 진행하였으며, 이러한 연구들의 정확도는 약 70%를 웃돌기에 이를 개선하기 위한 주된 연구 방향성은 학습 데이터에 변형을 통해 증강에 활용하거나, 전처리를 통해 방사선학 이미지의 불필요한 정보를 줄임으로써 주요한 정보만 parameter로 사용하여 학습에 사용용하는 방안이 활발히 진행되었으며, 이러한 관련 연구로 방사선학 이미지 관절 탐지에 Yolo-V2 모델을 활용하여 Joint Detector의 역할로 활용하였으며, 이러한 전처리를 통해 방사선학 이미지에서 연골을 탐지하여 전체 무릎 방사선학 이미지에서 관절 부분만 Crop하여 사용하였고, Classification을 위해서 144만개의 파라미터를 사용하는 VGG-19 모델을 사용하였으며, 전체 테스트 이미지에 69.58%의 정확도를 보였다. [2] Parameter를 줄이는 연구로 FCN(Fully Connected Network)을 Detection에 활용하여 앞선 연구와 같은 역할로 사용하였으며, Parameter를 줄이기 위해 Convolution Layer의 개수와 다른 매개변수의 분석을 통해 약 5.4만개의 파라미터만을 사용하는 퇴행성 무릎관절염 등급 분류에 가장 적합한 Classification 모델을 제작하여 모델 학습을 진행한 결과, 약 63.5%의 성능을 보였다. [3] Date argmentaion과 경량화 모델을 사용한 해당 연구에서는 전체 이미지의 작은 부분만 학습하는 구조를 가진 DenseNet model이 K-L grade 분류 문제에 대해 효과적인 성능을 보일것이라고 생각하여 DenseNet-169 모델과 Inception-V3 모델을 학습에 사용하였으나, Inception-V3는 낮은 성능을 보여 제외하였으며, DenseNet-169 모델의 사전학습으로 ImageNet을 학습하였고, 학습 데이터 부족 문제를 해결하기 위해서 학습 데이터에 80%에 대해서는 Crop, upscale, add noise, flip, randomize contrast를 단계적으로 적용하여 학습을 진행하였고, 20%에 대해서는 Crop만을 적용하였으며, 증강된 데이터를 바탕으로 한 학습의 결과 약 71%에 웃도는 성능을 보였지만, 같은 테스트 데이터 셋에 대한 방사선학 전문가의 정확도가 약 61%에 웃도는 것을 확인하여 딥러닝 모델의 정확도가 퇴행성 관절염 등급 진단의 방사선학 전문가의 정확도를 넘어섰다는 것을 검증하였다. [4].

관련 연구들을 통해 퇴행성 관절염 등급 분류의 정확도 개선을 위한 연구 방향성은 학습 데이터의 변형을 적용해 관절 부분만 있는 데이터 셋을 제작하여 데이터 증강을 적용하는 방법과 파라미터의 수를 줄이기 위한 모델 경량화와 noise 제거를 위해 전체 방사선학 이미지에서 관절 부분만 탐지하여 분류에 필요한 내용만을 학습에 활용하도록 하는 경량 모델을 설계하여 사용하는 방법이 연구되었으나, 눈에 띄는 정확도 개선이 이루어지지 않았기에 본 논문에서는 정확도와 경량화를 가장 중요한 기준으로 연구를 진행하기 위해 많은 양의 Parameter를 요하는 VGG와 같은 모델을 사용하거나 데이터 증강을 통해 학습에 사용하지 않고, 학습 비용을 줄이기위해 경량화 모델 중 하나인 Xception 모델을 활용하고자 하였다.

3. 실험

제안된 Xception 모델 성능 평가는 2가지의 방법을 통해 이루어졌으며, DPhi의 “Data Sprint #35: Osteoarthritis Knee X-ray” 대회[5]에 참가하여 해당 대회에서 제공된 테스트 데이터셋에 대한 예측 정확도와, 추가 테스트 데이터 셋을 제작하여 Confusion Matrix를 기준으로 하여 모델의 성능을 판단하였다.

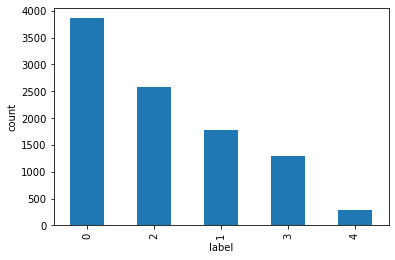
하지만 기존 연구보다 높은 정확도를 위해 기존 Xception 모델에 구조 변형을 가했으며, 해당 결과로 본래의 Xception 모델의 설계된 Parameter보다 더 많은 양의 파라미터를 사용하였다.

3.1 사용 데이터

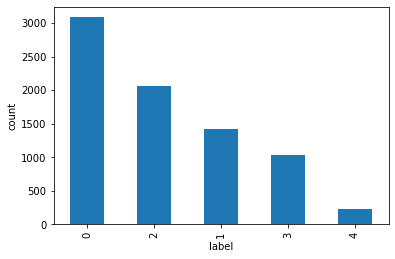
해당 실험에는 표 1과 같이 구성되어있는 3개의 데이터 셋이 사용되었으며, 학습 데이터로는 앞선 관련 연구의 Yolo-V2를 활용하여서 제작되었던 데이터를 학습 데이터로만 사용하였으며, DPhi에서 제공되는 학습, 검증 데이터셋을 합쳐 검증 데이터로 활용하였다. 테스트 데이터로 DPhi의 Test dataset과 직접 제작한 dataset을 모델 테스트 데이터 셋으로 사용하였다.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
|  | Grade 0 | Grade 1 | Grade 2 | Grade 3 | Grade 4 |
| Yolo-V2  (Fig. 2.) | 3857 | 1770 | 2578 | 1286 | 295 |
| DPhi train  (Fig. 3.) | 3085 | 1416 | 2062 | 1029 | 236 |
| DPhi test | 1958 | | | | |
| Created data  (Fig. 4.) | 63 | 199 | 139 | 84 | 46 |

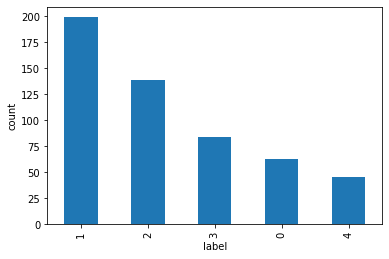
**Table. 2.** Used dataset Configuration



**Fig. 2.** Yolo-V2 Dataset Configuration



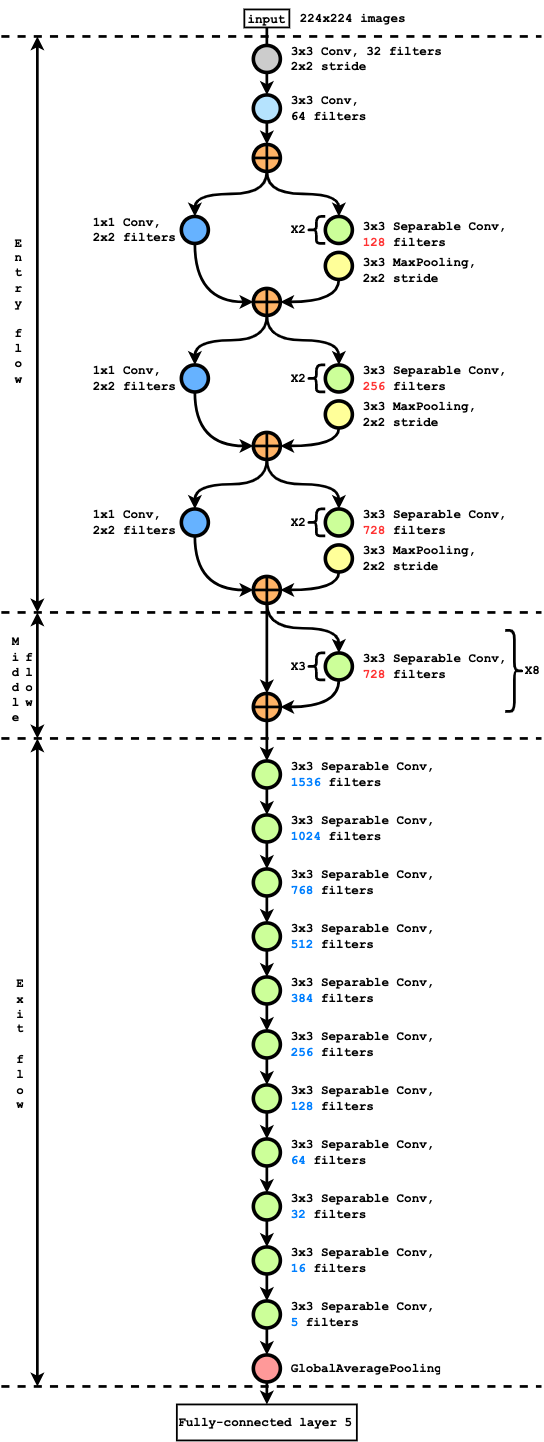
**Fig. 3.** DPhi Train Dataset Configuration



**Fig. 4.** Created Dataset Configuration

3.2 학습

변형된 Xception 모델은 그림 2와 볼 수 있으며, 1차 학습 과정에는 Yolo-V2 데이터를 학습시켰으며, DPhi train data를 통해 2차 학습을 진행하였다. 학습 매개변수는 전체 이미지 사이즈 224 by 224, Batch size = 8, Optimizer = Adam (learning rate = 0.00001)로 설정하였으며, 사전학습으로 ImageNet을 사용하여, Tensorflow 2.6.0, Keras 2.6.0, RTX 3090 환경에서 학습을 진행하였다.

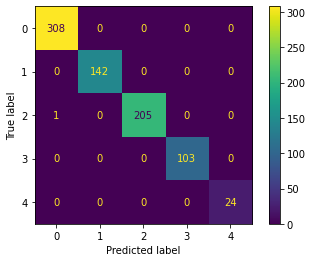


**Fig. 5.** Stage of knee osteoarthritis

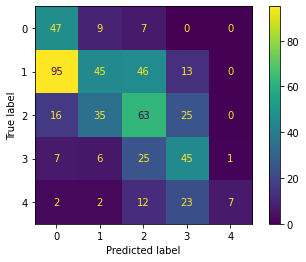
4. Conclusions

학습을 마친 후 test dataset을 통해 Confusion matrix를 그린 결과 그림 3, 4와 같이 나타났으며, 학습된 모델은 DPhi test dataset에 대해서는 99.84%로 높은 성능을 보였지만, 제작된 dataset에 대해서는 38.98%로 기대에 미치지 못한 성능을 보였다. 이러한 결과가 나온 이유는 표 1에서 나타났듯 학습, 검증 테스트 데이터 분포가 거의 유사하게 구성되어있으며, 이러한 이유로 학습 과정에 Data leakage 현상이 발생하여 정확도가 높게 나왔다고 생각하며, 그 근거로는 제작된 테스트 데이터 셋의 분포는 1 > 2 > 3 > 0 > 4 순으로 학습, 검증에 사용된 데이터 셋과 다른 분포를 가지고 있으며, 그림 4의 에측된 데이터 분포 개수는 0 > 2 > 3 > 1 > 4로 학습 데이터의 분포와 유사하게 예측하여 Data leakage 현상이 발생하였다고 볼 수 있다.

추후 연구로는 다른 무릎 관절 방사선학 이미지 데이터 셋을 활용한 테스트를 통해 Xception 모델의 성능을 더욱 명확하게 검증을 진행할 예정이며, 관련 연구에서 언급되었던 연구와 같은 환경에서의 학습을 진행한 후 제작된 테스트 데이터 셋에 대한 검증도 진행할 예정이다.



**Fig. 6.** Accuracy result, DPhi validation dataset



**Fig. 7.** Accuracy result, created dataset

텍스트, 모니터, 여러개이(가) 표시된 사진

자동 생성된 설명

**Fig. 8.** Accuracy result, DPhi validation dataset

텍스트, 모니터이(가) 표시된 사진

자동 생성된 설명

**Fig. 9.** Accuracy result, DPhi validation dataset

|  |  |  |
| --- | --- | --- |
|  | 정답(%) | 오답(%) |
| Fig. 6 | 99.8468(%) | 0.1532(%) |
| Fig. 7 | 38.98(%) | 61.02(%) |

**Table. 3.** Accuracy result

|  |  |  |
| --- | --- | --- |
| Label | Created data | Predict |
| Grade 0 | 63 | 167 |
| Grade 1 | 199 | 97 |
| Grade 2 | 139 | 153 |
| Grade 3 | 84 | 106 |
| Grade 4 | 46 | 8 |

**Table. 4.** Accuracy result

5. Acknowledgement

이 논문은 2020년도 정부(교육부)의 재원으로 한국연구재단의 지원을 받아 수행된 기초연구사업임

(No. 2020R1A6A1A03040583)

References

1. <https://www.snubh.org/dh/main/index.do?DP_CD=JRC&MENU_ID=002008024>
2. 오승준, 장은수, 오영선, 강위창, 이은정, 정인철, “퇴행성 무릎관절염 변증도구의 신뢰도 평가 및 무릎 ROM, VAS, WOMAC과의 상관관계에 대한 탐색적 연구”, Journal of Korean Medicine Rehabilitation Vol. 29 No. 3, July 2019
3. Pingjun Chenm , Linlin Gao, Xiaoshuang Shi, Kyle Allen, Lin Yanga, “Fully automatic knee osteoarthritis severity grading using deep neural networks with a novel ordinal loss” Computerized Medical Imaging and Graphics Volume 75, Pages 84-92, July 2019
4. Joseph Antony, Kevin McGuinness, Kieran Moran, and Noel E O’Connor, “Automatic Detection of Knee Joints and Quantification of Knee Osteoarthritis Severity using Convolutional Neural Networks” MLDM 2017, pp 376-390, 2017.
5. Kevin A. Thomas, Łukasz Kidziński, Eni Halilaj, Scott L. Fleming, Guhan R. Venkataraman, Edwin H. G. Oei, Garry E. Gold, Scott L. Delp, “Automated Classification of Radiographic Knee Osteoarthritis Severity Using Deep Neural Networks” Radiology: Artificial Intelligence Volume: 2 Issue 2, 2020
6. https://dphi.tech/challenges/data-sprint-35-osteoarthritis-knee-x-ray/81/overview/about