

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 ARToolkit, the view of the augmented 3D medical model can be dynamically varied.

Keywords: Knee osteoarthritis, Deep Learning, Model Lightweight, Convolutional Neural Network

1. Introduction

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

해당 연골은 신체의 움직임에 있어 필요불가결하며, 생활 속에서 계속 사용되기에 연령이 높아짐에 따라 연골 마모의 정도가 심해져 퇴행성 관절염이 발생해 움직임에 불편함을 느끼게 될 수 있다. 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]

위처럼 현재 의료계에서는 의사의 진단 과정에 과학적이고 논리적인 근거를 필요로 하고 있으며, 이러한 근거를 기반으로 한 표준화된 진단 방식을 활용할 수 있게된다면, 의사는 보다 객관적인 지표를 활용할 수 있게 되며, 이를 통해 더욱 빠르고 정확한 진단을 내릴 수 있게 될 것이라고 생각한다. 최종적으로는 개선된 방식을 통해 퇴행성 관절염으로 인한 질병을 예방할 수 있을 것이라고 생각한다.

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 그림 2와 같이 시각적으로 보이는 변화 현상을 통해 퇴행성 관절염의 진단이 가능하기에 그동안 축적된 무릎 방사선학 이미지를 딥러닝 모델에 학습시켜 퇴행성 관절염 등급을 판단하고자하는 연구들이 폭넓게 진행되고 있으며, 이러한 시각적인 변화를 바탕으로한 딥러닝 모델을 통해 객관적인 지표로의 정보 전달이 가능해지나, 현재 딥러닝 모델을 활용한 SOTA(State-of-the-art) 예측 성능이 약 70%를 웃돌기에 의료계에 딥러닝 학습 모델을 실제 접목하기 위해서는 퇴행성 관절염 등급 분류의 정확도가 더욱 개선되어야 실제 사용이 가능해지기에 정확도를 주된 목표로한 관련 연구가 폭넓게 진행되었다.

본 논문에서는 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 딥러닝 모델은 성능을 위해 VGGNet와 같은 깊이가 깊은 네트워크를 가진 딥러닝 모델을 활용하여 많은 양의 Parameter를 사용하였으나, 연구 결과 예측 정확도 증가가 보이지 않았기 때문에 이러한 방식 개선을 위해 네트워크 깊이를 깊게 하는것이 아니라 매개변수를 효과적으로 사용하기 위한 컨볼루션 네트워크 모델인 DenseNet과 같은 모델들이 연구되었으나,

눈에 띄는 성능의 향상은 볼 수 없었다.

본 연구에서는 매개변수의 수를 줄여 학습 시간, 비용을 최소화하기 위해서 경량화 모델들을 학습에 사용하였으며, 정확도를 높이기 위해 학습 종료 과정에서 줄어드는 매개변수의 수를 조절하기 위해 Xception 모델의 Exit flow를 수정하여 본래의 모델 구조의 끝단의 filter 수를 조밀하게 구성하여 기존 모델 보다는 매개변수를 많이 사용하지만 보다 높은 정확도를 목표하였다.

네트워크를 겹겹히 깊이 쌓아 구성하는 VGGNet에 비해 모델의 깊이가 얇기 때문에, 비교적 Parameter를 줄일 수 있어 모델을 보다 빠르게 학습이 가능할 것이며, 저명한 경량화 모델들 간의 정확도 비교를 통해 해당 도메인에 적합한 딥러닝 모델을 찾아 학습을 진행하였다.

본 연구에서는 퇴행성 관절염 등급 분류 문제에 경량화 모델을 적용하였으며, 다른 경량화 모델들과의 성능 비교를 통해 결과 분석을 진행하였고, 퇴행성 관절염 등급 분류 문제에 개선된 정확도를 보이는 Xception 변형 모델을 제안한다.

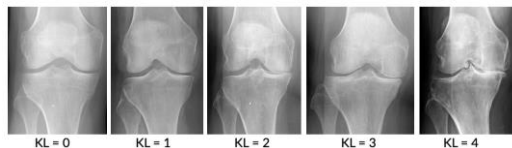


Fig. 1. Stage of knee osteoarthritis []

〈퇴행성관절염의 무릎 관절 방사선 사진〉



Fig. 2. 퇴행성 관절염의 시각적 증상 [3]

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 Parameter를 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 사용하는 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. [4]

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. [5]

Date argmentaion과 경량화 모델을 사용한 해당 연구에서는 전체 이미지의 작은 부분만 학습하는 구조를 가진 DenseNet model이 K-L grade 분류 문제에 대해 효과적인 성능을 보일 것이라고 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. [6].

경량화를 고려한 모델로는 SqueezeNet, Xception, MobileNet, ShuffleNet등이 모델 경량화에 많은 영향을 주었으며, SqueezeNet[7]은 0.5MB도 안되는 용량을 가지고 있기에 연산량이 적어 학습이 빠릅니다. 이러한 구조를 갖기 위해서 Squeeze layer와 expand layer로 이루어진 Fire Module이라는 구조를 제안하였고, 이를 활용하여 AlexNet보다 50배 적은 매개변수로 동일한 정확도를 가진 모델이지만, 해당 모델은 실시간으로 업데이트가 이루어지는 모델에 더욱 적합합니다.

Xception[8] Depthwise Separable Convolution이라는 효율적인 Convolutional Layer를 제안하였으며, 이를 통해 일반적인 컨볼루션 네트워크 모델 계산량의 1/9정도로 줄여 약 9배 빠른 학습 속도를 가졌으며, Depthwise Separable Convolution로 인해 채널마다 filter를 넣어 spatial feature를 추출할 수 있게 되었다.

MobileNet[9]은 Depthwise Separable Convolutions 구조를 적절히 활용해 모바일 기기나 어플에서 동작가능한 수준의 경량한 구조를 제안했고, 구조를 더욱 개량해 버전 3까지 발표되었습니다.

ShuffleNet[10]은 Depthwise Separable Convolutions된 결과를 섞어서(Shuffle) 경량화되었으며, 1x1 pointwise convolution의 연산량을 감소시키기 위해 제안한 Pointwise Group Convolution과 3x3 group

convolution에서 발생하는 문제점을 해결하기 위한 방법인 Channel Shuffle을 제안하였고 이를 통해 연산량에서도 이점을 가지며, 제한된 연산량 내에서 더 많은 피쳐맵을 활용하는 모델입니다.

정확도 개선을 위한 연구는 폭 넓게 진행되었으며, 데이터 부족을 해결하기 위해 증강을 적용하는 방식과 Network 최적화를 통해 파라미터의 수를 줄여 학습 비용과 시간을 줄이려는 연구들이 있었다.

하지만, 눈에 띄는 정확도 개선이 이루어지지 않았기에 많은 양의 Parameter를 요하는 VGG와 같은 모델에 해당 도메인을 적용하는 것이 아니라 Depthwise Separable Convolution를 통해 효과적으로 spatial feature를 추출해낼 수 있는 경량화 모델 중 Xception 모델에 적용하고자 하였다.

3. 실험

제안된 Xception 모델 성능 평가는 DPhi의 “Data Sprint #35: Osteoarthritis Knee X-ray” 대회[5]에 참가하여 해당 대회에서 제공된 test dataset에 대한 예측 정확도를 모델의 성능 평가의 주된 기준으로 삼았으며, Kaggle의 “Knee Osteoarthritis Dataset with KL Grading – 2018”에서 제공한 무릎 관절염 데이터셋[6]을 검증 데이터로 사용하고, 모델의 실험에 대한 Confusion Matrix를 분석하여 모델의 성능을 판단하였다.

퇴행성 관절염 진행 증상 중 점진적으로 나타나는 관절 간격의 감소와 연골 아래 뼈의 음영이 짙어지는 경화 현상을 통해 등급 구분을 하기 위한 feature를 추출하기에 Xception 모델이 가장 적합한 모델이라고

생각한다. 왜냐하면 Xception은 Depth-wise convolution를 활용하여서 채널마다의 spatial feature를 추출할 수 있기 때문에 해당 도메인에 가장 적합한 모델이기에 해당 연구에 적용하였으며, 추출된 feature가 급격하게 줄어드는 과정에서 feature 손실이 일어나 올바른 예측을 하지 못하였다고 생각하여 올바른 feature를 추출하기 위해 filter를 조밀하게 구성하여 Xception model의 Exit flow를 적용하였지만, 조밀한 filter의 추가로 인해 기존 Xception보다 약 2.5배정도로 매개변수가 늘어났다.

3.1 Data

해당 실험에 사용된 데이터셋의 세부 구성은 표 2에 나타있으며, DPhi에서 제공받은 Train, Test data를 학습과 테스트에 사용하였으며, Validation 데이터는 제공되지않아 Kaggle data을 validation data로 사용하였다.

표 2, 그림 3, 4를 보면 알수있듯이 Train, Test Data imbalance가 아주 심하기 때문에 이를 해결하기 위한 방법으로 가중치 부여를 통하여 데이터 수가 적은 class는 더 높은 가중치를 갖게 하여 학습을 진행하였다.

	Grade 0	Grade 1	Grade 2	Grade 3	Grade 4
Kaggle 2018 (Fig. 2.)	3857	1770	2578	1286	295
DPhi train (Fig. 3.)	3085	1416	2062	1029	236
DPhi test	1958				

Table. 2. Used dataset Configuration

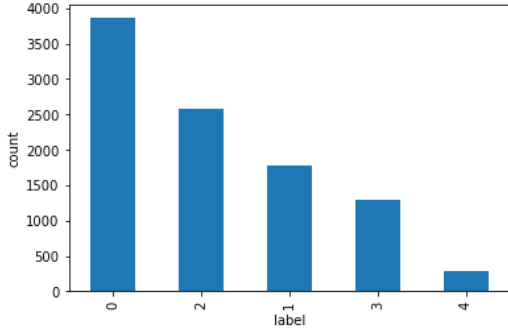


Fig. 3. Kaggle 2018 Dataset Configuration

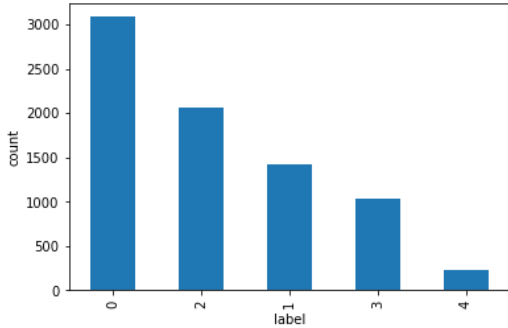


Fig. 4. DPhi Train Dataset Configuration

3.2 학습

그림 5는 변형된 Xception 모델을 나타내며, 학습은 1, 2차로 나누어져 학습되었으며, 1차 학습에는 train data로 DPhi train, validation data로는 kaggle 2018 data를 사용하였으며, 2차 학습에는 kaggle 2018 data를 나누어 2차 학습을 진행하였다.

학습 이미지는 224x224x3을 사용하였으며, Pretraining = imagenet, Batch size = 8, Optimizer = Adam (learning rate = 0.00001)로 Early stoping을 설정하였으며, ubuntu 18.04, V100, Tensorflow 2.6.0, Keras 2.6.0 환경에서 학습을 진행하였다.

딤러닝 모델간의 성능 비교의 전체 성능은 표 3에서 나타내었으며, 경량화 모델의 발전에 큰 영향을 준 MobileNet, ShuffleNet,

SqueezeNet, Xception과 추가로 깊은 Network Layer를 사용하는 딤러닝 모델도 동일한 환경에서 학습을 진행하였으며, 관련 연구에서 설명하였던 딤러닝 모델들간의 성능도 함께 비교하였다.

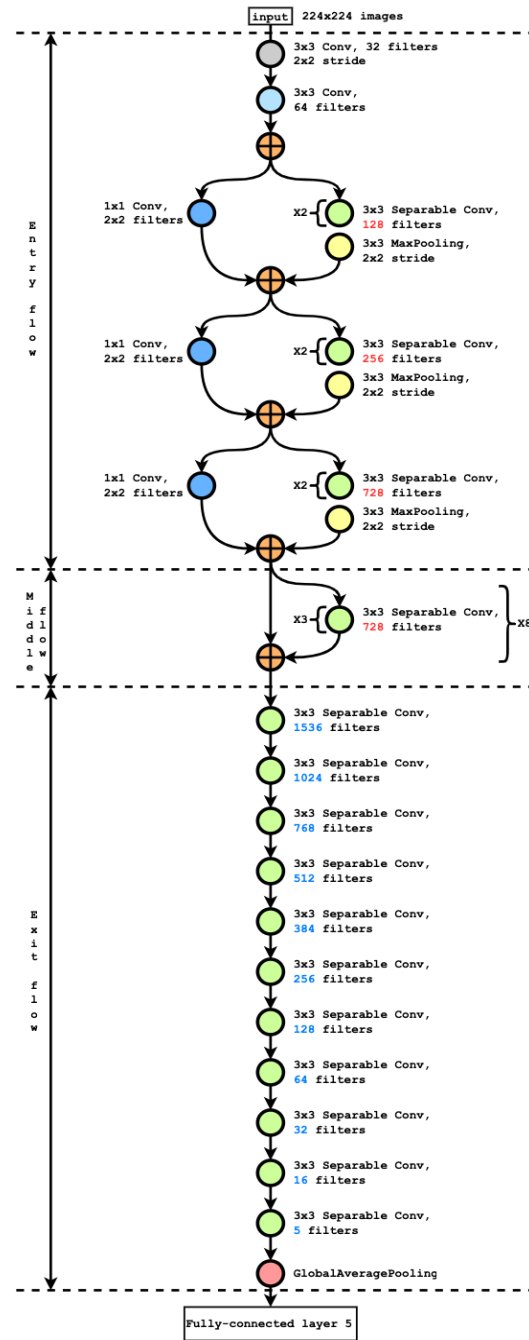


Fig. 5. Stage of knee osteoarthritis

4. Conclusions

학습 결과 모델의 정확도는 DPhi의 Test dataset을 통하여 계산하였으며, 같은 환경에서 실험을 진행한 정확도 결과이다. 우리가 제안한 모델이 가장 많은 파라미터를 사용하였지만, 다른 연구에서 사용되었던 모델이나, 경량화 모델들 중에서 Xception model이 가장 높은 성능을 보였으며, 제안된 모델은 변경전 모델보다 파라미터를 약 2.5배 더 사용하기는 하지만, 정확도에서 약 7% 정도의 성능 차이를 보이는 것으로 보아 유의미한 개선이 이루어졌다고 볼 수 있으며, 추후 연구를 통해 모델의 네트워크 개선이 가능해 보인다.

추후 연구로는 다른 무릎 관절 방사선학 이미지 데이터 셋을 활용한 테스트를 통해 제안된 Xception 모델의 성능을 더욱 명확하게 검증을 진행해야할 필요성이 있다.

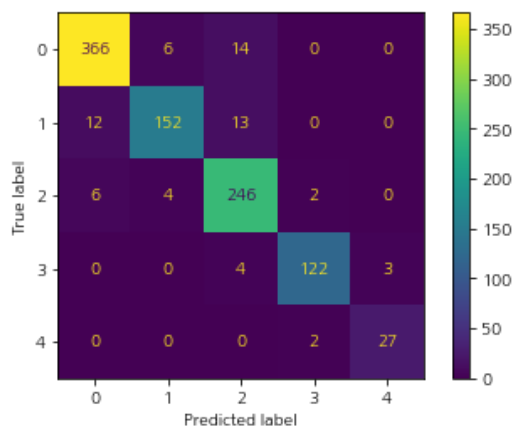


Fig. 6. Improved Xception Validation data result, at Kaggle 2018 dataset

Model	Parameter	Accuracy
DenseNet121	6,958,981	55.46
DenseNet169	12,492,805	55.76
DenseNet201	18,102,533	66.85
EfficientNetB0	4,013,953	39.68
EfficientNetB1	6,519,589	39.17
EfficientNetB2	6,519,589	38.51
EfficientNetB3	10,703,917	38.51
ResNet50	23,544,837	65.17
ResNet101	42,563,077	50.56
ResNet152	58,229,765	40.60
SqueezeNet [7]	737,989	39.43
Xception [8]	20,899,127	72.22
MobileNet [9]	3,212,101	40.50
MobileNetV2	2,230,277	40.60
MobileNetV3Small	1,522,981	40.60
MobileNetV3Large	4,208,437	44.18
ShuffleNet [10]	918,125	43.05
Ours*	56,755,255	79.62

Table 3. 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, WOMAC 과의 상관관계에 대한 탐색적 연구”, Journal of Korean Medicine Rehabilitation Vol. 29 No. 3, July 2019
- [3] <https://www.100ssd.co.kr/news/articleView.html?idxno=70447>
- [4] 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
- [5] 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.
- [6] 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
 - [7] Forrest N. Iandola, Song Han, Matthew W. Moskewicz, Khalid Ashraf, William J. Dally, Kurt Keutzer, “SqueezeNet: AlexNet-level accuracy with 50x fewer parameters and <0.5MB model size”, ICLR 2017
 - [8] François Chollet, “Xception: Deep Learning with Depthwise Separable Convolutions”, 2017 IEEE(CVPR)
 - [9] Andrew G. Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, Hartwig Adam, “MobileNets: Efficient Convolutional Neural Networks for Mobile Vision Applications”, 2017 CVPR
 - [10] Xiangyu Zhang, Xinyu Zhou, Mengxiao Lin, Jian Sun, “ShuffleNet: An Extremely Efficient Convolutional Neural Network for Mobile Devices”, CVPR 2017
 - [11] <https://www.kaggle.com/datasets/tommyngx/kneeoax>
 - [12] <https://dphi.tech/challenges/data-sprint-35-osteoarthritis-knee-x-ray/81/overview/about>