

장 정결도 이미지 분류를 위한 이단계 컨볼루션 뉴럴 네트워크 학습 모형

이종범[○], 김대유, 이현규^{*}

인하대학교 컴퓨터공학과, 인하대학교 의예과

bub3690@naver.com, dea0323@gmail.com, hglee@inha.ac.kr^{*}

요 약

대장 수술과 내시경 검사 전, 장 정결(bowel prep)을 진행한다. 하지만, 장 정결 완료 의사결정 과정에서 의사와 간호사에게 업무적 비효율이 발생한다. 따라서 의사 결정과정을 간소화 하기위해 배변사진에 대한 딥러닝 기반 장 정결도 분류 모형을 제안한다. 환자의 배변사진은 각 조명의 색, 변기의 모양이 다르고, 환자의 손가락, 빛 반사, 휴지 등의 폐색 현상(occlusion) 및 노이즈가 존재한다. 데이터 증강 방법과 보조라벨을 사용한 2 단계 학습 모형을 도입하여 이러한 다양성 문제를 해결한다. 그 결과, 제안하는 모델은 80.5% 성능으로 이미지만을 입력으로 사용한 모형에 비해 3.9%의 정확도 향상을 얻을 수 있었다.

1. 서론

장 정결(bowel prep)이란 대장내시경 및 대장 수술을 위해 환자가 전처치제(cool prep, orafang)를 복용하여 장을 정화하는 과정이다. 장 정결이 제대로 수행되지 않을 시, 대장 내시경 검사를 시행하는데 검사 시간이 연장되거나, 검사를 시행하더라도 분변으로 인하여 병변을 놓치게 되고, 검사 자체를 시행할 수 없어 중단할 수 있다. 시술 시에는 과도한 공기 주입으로 심한 복통을 유발하고 시술 후에는 발열, 천공 등의 위험을 초래할 수 있다[1]. 따라서, 효과적이고 안전한 대장 내시경 및 수술을 위해서는 장 정결 시행자와 의사의 확인이 중요하다.

본 병원의 대장 항문 외과의 경우 주당 4-5 경우의 장 정결이 필요한 대장 수술을 진행한다. 익일 장 절제 수술환자들은 오후 3시에 입원하여 장 정결을 시작한다. 환자가 전처치제를 복용 후 배변 시 매번 간호사가 사진을 찍어서 의사에게 전송하면 장 정결이 더 필요할 지 여부를 당직전공의가 결정하는 방식으로 장 정결을 진행한다. 하지만, 간호사들이 장 정결 상태를 확인하기 위해 환자의 배변 사진을 촬영하고 의사에게 다음 단계 진행 확인을 받기까지 매 단계의 시간 지연이 발생한다. 환자 당 최소 2-3 장의 배변 사진 확인 후 최종 확인이 이루어져야 하므로 간단한 업무임에도 업무 과정에서 비효율이 발생하며, 정해진 시간 상관없이 환자의 배변 시간에 따라 확인이 이뤄져야 하므로 간호사 및 의사의 불필요한 업무 과중이 생긴다.

선행 연구로는 대장 내시경 영상에서 딥러닝을 활용해 장 정결도를 분류한 연구들이 있다. 컨볼루션 뉴럴 네트워크를 통해 Boston Bowel Preparation Scale (BBPS) 점수를 예측하여 93%의 정확도로 내시

경 의학자보다 더 높은 결과를 보였고, 거품이 있는 영상에서는 80% 정확도임을 보고했다[2]. 대장 내시경 영상 품질에 따라 분류 결과가 정확하지 않을 수도 있어, 영상 품질 분류와 BBPS 점수 분류 2 단계로 진행한 연구도 있다[3]. 하지만, 이 연구들은 대장 내시경을 통해 촬영된 영상을 활용했기에, 대장 내시경 및 수술 전에 발생하는 의사 및 간호사의 업무 과중을 줄여줄 수는 없다.

따라서, 본 연구는 딥러닝을 활용하여 모바일 환경에서 환자의 배변사진의 장 정결도를 분류하여 전국 병원의 의사 및 간호사들이 장 정결을 판단하는 업무를 줄이고자 하는 것이 목표이다. 이를 위해 보조 레이블을 활용한 2 단계 컨볼루션 뉴럴 네트워크 학습 모형을 제안한다.

2. 제안하는 방법

2.1 데이터 셋

데이터 셋은 인하대학교 대장 항문외과에서 2019년 8월부터 2022년 10월까지 대장 절제 수술환자의 851건의 이미지가 수집되었다. 데이터의 라벨은 이미지의 상태에 따라 ‘약물 복용 필요(0)’, ‘물 복용 필요(1)’, ‘장 정결 완료(2)’ 3 가지로 구성되며, 전공의가 라벨링을 수행했다. 그림 1은 각 라벨에 따른 장 정결 상태이다. 훈련, 검증, 평가 데이터는 7:1:2로 분할하여, 랜덤 시드 5번의 반복 실험을 수행한다. 데이터의 분포는 표 1과 같이 구성된다.

뉴럴 네트워크가 주요한 지점에 집중할 수 있도록 보조 라벨을 추가 도입한다. 보조 라벨은 총 3

가지로 ‘색’, ‘탁도’, ‘잔여물’이다. 각 범주의 수준은 색 3 가지, 탁도 2 가지, 잔여물 3 가지이며, 표 2 는 보조 라벨의 분포를 나타낸다. 색의 수준은 변의 색이 검은색과 갈색인 경우 0, 어두운 황색은 1, 밝은 황색 및 맑은 색일 때는 2 이다. 탁도의 수준은 불투명 한 경우 0, 투명한 경우 1 이다. 잔여물의 수준은 잔여물이 많으면 0, 중간 량 1, 소량 2 이다. 보조 라벨의 수준이 높을수록 장 정결이 완성되었음을 의미한다.

표 1. 훈련 및 검증 데이터 셋 분포

라벨	0	1	2
	약물 복용 필요	물 복용 필요	장 정결 완료
훈련	226	186	268
검증	16	15	27
테스트	41	35	37

표 2. 데이터 셋의 보조 레이블 분포

수준	0	1	2
색	129	157	565
탁도	374	477	-
잔여물	157	242	452



그림 1. 장 정결 상태에 따른 이미지.
(a) 약물 복용이 필요한 상태 (라벨 0),
(b) 물 복용이 필요한 상태 (라벨 1),
(c) 장 정결 완료 상태 (라벨 2)

2.2 데이터 증강

또한, 오측 요인을 보완하고 모형을 일반화하기 위해 데이터 증강방법을 시도했다. 그림 2 은 사용된 데이터 증강방법의 예시들이다. 빛 반사에 의한 오류를 줄이기 위해 차량용 컴퓨터 비전에서 활용되는 그림 2. a 의 ‘임의 빛 번짐’ 데이터 증강 방법을 활용했다. 임의 빛 번짐 데이터 증강 방법은 카메라 렌즈에 빛이 번지는 현상을 구현한 방법이다. 손가락, 휴지 등의 폐색 현상으로 인한 오류를 줄이기 위해 그림 2. b 의 ‘오려내기(Cutout)’ 데이터 증강방

법을 활용했다. 오려내기 데이터 증강방법은 학습 영상의 일부분을 무작위로 가리는 데이터 증강 기법으로, 이미지 폐색 현상을 구현한 방법이다[4]. 그림자에 의한 오류를 줄이기 위해 그림 2.c 의 ‘임의 그림자 생성’ 데이터 증강 방법을 활용했다. 그 외에는 촬영 환경의 빛 세기와 대조로 인한 오류를 줄여주는 그림 2.d 의 ‘임의 빛 세기, 대조’ 증강방법과 일반적인 의료 이미지 학습에 사용되는 상하좌우 뒤집기 데이터 증강방법을 활용했다.

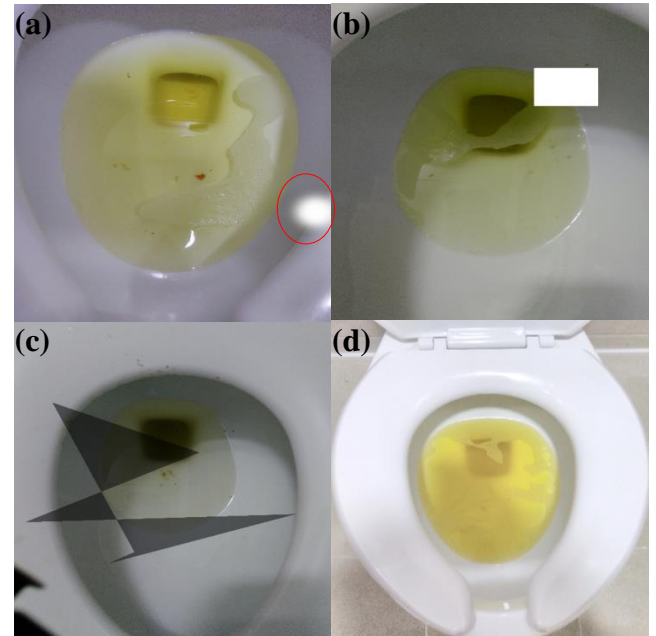


그림 2. 학습에 사용된 데이터 증강 방법.

- (a) 임의 빛 번짐(빨간 영역) 데이터 증강,
(b) 오려내기 데이터 증강, (c) 임의 그림자 생성,
(d) 임의 빛 세기 대조 증강

2.3. 백본 네트워크

컨볼루션 뉴럴 네트워크 백본의 역할은 이미지에서 유의미한 특징을 추출하는 것이다. 장 정결도 이미지 데이터는 다양한 변기 색, 빛 반사와 같은 노이즈를 포함하고 있어 과적합이 될 수 있기에, 일반화 성능이 우수한 모형을 사용한다. 모형에 사용되는 컨볼루션 뉴럴 네트워크 백본 및 베이스라인은 RESNET18 모형을 활용하였다. RESNET 모형은 잔차 블록에서 입력층 x 와 출력층의 $F(x)$ 를 합산하여 기울기 손실과 과적합 문제를 해소하기에, 현재의 문제에 적합한 모형이라 할 수 있다[5]. RESNET18 모형은 RESNET 모형 중 가장 파라미터 수가 적기에, 모바일 환경에 가장 적합하여 선택했다.

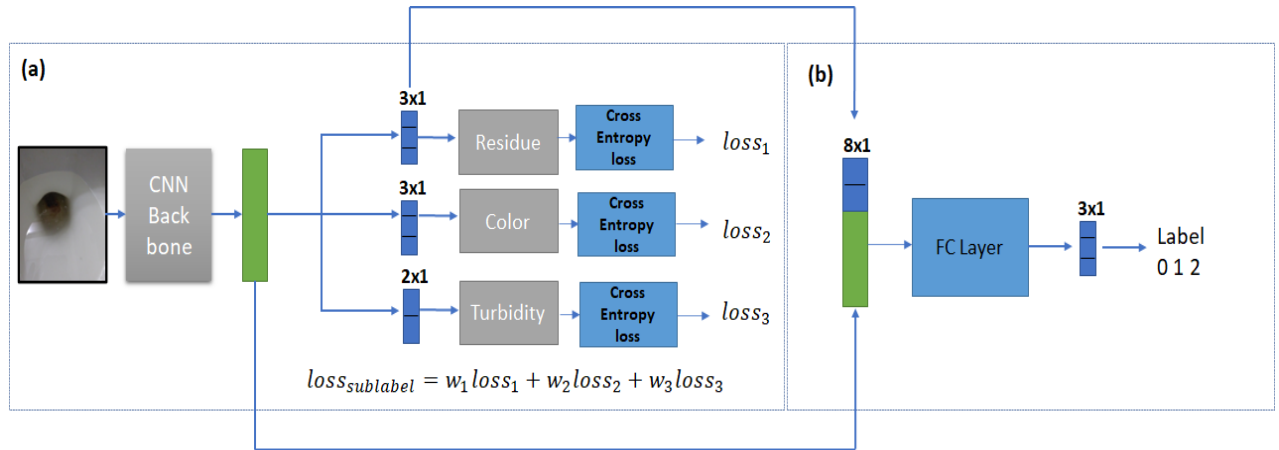


그림 3. 보조 라벨 기반 2 단계 컨볼루션 뉴럴 네트워크 모형

- (a) 보조 라벨 예측 수행 모형,
 (b) 보조 라벨 사전 학습된 네트워크를 이용한 장 정결 라벨 예측 모형, FC Layer: 다중 퍼셉트론 완전 연결 층

dropout 모듈을 사용하여 과적합 문제를 해소한다.

2.4. 보조 라벨 기반 2 단계 학습 모형

본 논문에서 제안하는 모형은 그림 3 로 크게 (a) 보조 라벨 예측 모형 학습 단계와 (b) 장 정결 라벨 예측 모형 학습으로 2 단계로 구성된다.

2.4.1 보조 라벨 예측 모형 학습

보조 라벨 예측 모형에서는 3 가지 보조 라벨들을 사전 학습하여 주요 영역에 집중하도록 하는 단계이다. 이미지가 Pretrained RESNET18 모형을 통과하면 1000×1 크기의 노드가 생성된다. 노드들은 완전 연결층을 통해 각 보조 라벨의 수준으로 축소된다. 그리고 크로스 엔트로피 손실 함수를 통과하여 각 보조 라벨의 손실이 계산된다. 각 라벨의 손실들은 수식 1 의 가중합을 통해 하나의 손실로 계산되고 역전파에 사용된다.

$$loss_{sublabel} = w_1 loss_1 + w_2 loss_2 + w_3 loss_3 \quad (1)$$

2.4.2 장 정결 라벨 예측 모형 학습

장 정결 라벨 예측 모형에서는 3 가지 보조 라벨들이 사전 학습된 모형을 활용하여, 목표 장 정결 라벨을 학습시키는 모형이다. 컨볼루션 뉴럴 네트워크 백본에서 생성된 노드와 보조 라벨 예측 노드들을 특징 연결(Feature concat)하여 완전 연결 층을 통해 분류하는 구조를 가진다. 이때, 보조 라벨 예측 노드들의 가중치 동결 또는 컨볼루션 뉴럴 네트워크 백본의 가중치 동결 등의 학습방법이 있다. 완전 연결 층은 다중 퍼셉트론 구조로 ReLU 함수와

3. 실험 결과 및 분석

베이스라인, 보조 라벨 학습 수행 모형, 장 정결 라벨 분류 예측 모형 모두 공통 랜덤 시드 5 회 환경에서 실험이 수행되었다. 이때 베이스라인은 백본 네트워크에 보조 라벨 사용없이 이미지만을 사용하여 장 정결 라벨을 학습한 모형이다. 학습 세대(Epoch)는 최대 50 으로 수행되었지만, 조기 중단(Earlystop)은 10 으로 설정하여 과적합을 방지하고 학습 시간을 줄였다. 보조 라벨 예측 모형 학습에 사용되는 수식 1 의 라벨의 손실 가중치 w_1, w_2, w_3 는 실험적으로, $w_1 = 0.3, w_2 = 0.35, w_3 = 0.35$ 로 각각 설정했다.

3.1 데이터 증강

그림 4 은 학습된 백본 컨볼루션 뉴럴 네트워크가 예측을 수행하여 틀린 샘플에 대하여 GRAD-CAM(Class Activation Map)[6]을 시각화 하고 데이터 증강 후 GRAD-CAM 을 확인한 자료이다. GRAD-CAM 을 통해 환자의 배변사진에는 그림 4. a 에서 빛 반사에 의한 과적합 현상, 그림 4. b 에서 신발과 같은 복잡한 배경 영역에 의한 오류, 그림 4. c 에서 그림자에 의한 색감의 오류, 그림 4. d 에서 휴지와 거품 등의 폐색(occlusion)현상으로 인한 오류가 존재하는 것을 파악했다.

그림 4. a 의 3 열에서 임의 빛 번짐 데이터 증강을 통해 번기의 빛에 적합 하는 것이 완화되어 빛 반사된 영역 외에 GRAD-CAM 이 생성되었다. 그림 4.

b 의 복잡한 배경 요인은 오래내기 데이터 증강에 의해 신발이 아닌 변기에 주의 집중했다. 그림 4. c 는 그림자가 없는 변기의 일부에만 집중한 것으로, 임의 그림자 생성 증강을 통해 변기의 더 넓은 영역에서 집중했다. 그림 4. d 에서 기존에는 거품의 일부에만 집중했지만, 오래내기 데이터 증강 후 거품 외의 더 넓은 영역에 집중했다.

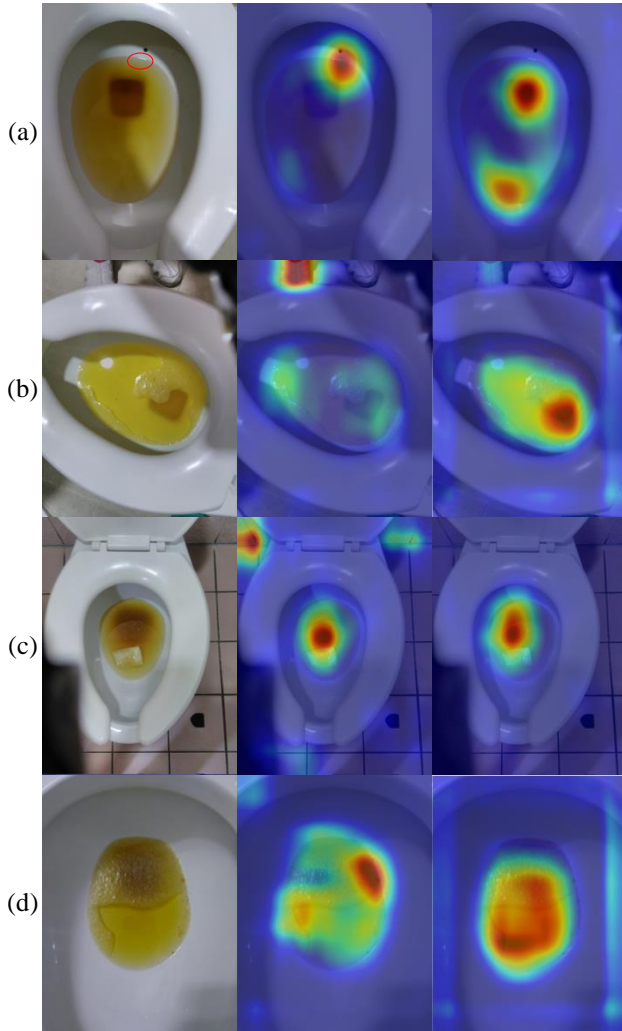


그림 4. 오측 요인 분석을 위한

(1 열) 원본 이미지와 (2 열) GRAD CAM 이미지
(3 열) 데이터 증강 훈련 후 GRAD CAM 이미지.

- (a) 빛 반사에 의한 오류(빨간 영역),
(b) 복잡한 배경 영역으로 인한 오류,
(c) 그림자에 의한 오류, (d) 거품으로 인한 오류

3.2 장 정결 예측 모형 구조 비교 실험

표 3 은 보조 라벨 예측 수행 모형에서 학습을 수행 후, 장 정결 라벨 예측 모형의 구조 변경에 따른 학습 결과를 나타낸다. 표 3 의 장 정결 예측 모

형은 그림 3 의 모형에서 가중치 동결이나 예측 레이어 제거가 없는 보조라벨 기반 2 단계 컨볼루션 뉴럴 네트워크 모형의 기본 형태를 의미한다. 백본 가중치 동결 모형은 보조 라벨 예측 수행 모형 학습 후, 백본의 가중치를 동결하고 장 정결 예측 모형을 학습한 모형이다. 백본 가중치를 동결했기에 기본 장 정결 예측 모형에 비해 서브 라벨의 정확도는 높지만 라벨의 정확도는 0.7% 감소했다. 서브 라벨 예측 가중치 동결 모형은 8×1 크기의 예측 레이어의 가중치를 동결한 모형이다. 기본 장 정결 예측 모형에 비해 정확도 1.1% 상승했고, 평균 서브 라벨 정확도는 1.2% 상승한 결과를 보였다. 서브 라벨 예측 정확도가 떨어지면 라벨의 예측도가 떨어지는 문제를 해소하기 위해 서브 라벨 예측 레이어를 제거한 모형은 예측 가중치 동결한 경우보다 낮은 정확도를 보였다. 서브 라벨 정확도와 예측 정확도 모두 유지하기 위해 백본 네트워크를 복제하여 한 개는 가중치를 동결하여 서브 라벨을 예측하고 한 개는 라벨을 학습한 경우, 평균 서브 라벨 예측 정확도는 기본 장 정결 예측 모형에 비해 5.4% 향상했지만 라벨 예측 정확도는 1.5% 감소했다.

표 3. 장 정결 예측 모형 구조 비교 실험. ‘+’는 장 정결 예측 모형에서 구조를 변경한 방법을 의미한다.

Test Acc (%)	라벨	색	잔여물	탁도
베이스라인	76.6	-	-	-
장 정결 예측 모형	77.7	76.6	63.7	82.3
+ 백본 가중치 동결	77.0	78.8	68.0	81.8
+ 서브 라벨 예측 가중치 동결	78.2	77.0	68.3	80.9
+ 서브 라벨 예측 레이어 제거	77.2	-	-	-
+ 두 모형 사용	76.1	82.3	71.5	85.0

3.3 장 정결 예측 모형 데이터 증강 비교 실험

표 4 은 장 정결 예측 모형 구조 중 가장 라벨 정확도가 우수한 서브 라벨 예측 가중치 동결 방법에 데이터 증강을 수행한 결과이다. 상하좌우 뒤집기 방법은 기존 모형에 비해서 라벨 예측 정확도는 2.3% 향상했고, 평균 서브 라벨 예측 정확도는 3.2% 향상하여 가장 우수한 성능을 보였다. 주요 영역을 가릴 수 있는 문제로 인해 오래내기, 임의 및 변검 방법은 3% 이하의 성능 하락을 보였다. 임의 밝기 대조 방법과 그림자 생성은 0.3% 이내의 차이를 보였다. 데이터 증강 조합에서는 라벨 예측 정확도의 큰 하락이 없는 상하좌우 뒤집기, 임의 밝기 대조, 그림자 생성 방법을 모두 적용한 방법이다. 기존 방법에 비해 라벨 예측 정확도는 1.6% 향상했고, 평균 서브 라벨 예측 정확도는 1.3% 향상했다.

표 4. 장 정결 예측 모형 데이터 증강 비교 실험.
‘+’는 장 정결 예측 모형에서 각 데이터 증강 방법을 이용해서 학습을 수행한 결과이다.

Test Acc (%)	라벨	색	잔여물	탁도
서브 라벨 예측 가중치 동결	78.2	77.0	68.3	80.9
+ 상하좌우 뒤집기	80.5	80.0	70.6	85.3
+ 오려내기	73.1	74.7	68.1	84.2
+ 임의 빛 번짐	75.0	74.9	61.9	81.0
+ 임의 밝기 대조	77.9	78.2	64.8	81.9
+ 임의 그림자	77.9	78.4	68.2	82.3
+ 데이터 증강 조합	79.8	76.8	67.6	85.7

4. 결론

본 논문에서는 환자 배변사진의 장 정결도 분류를 위해 보조 레이블을 활용한 2 단계 컨볼루션 뉴럴 네트워크 학습 모형을 제안한다. 이미지만을 사용한 RESNET18 백본 네트워크의 성능 76.6%에 비해 보조 레이블을 활용한 2 단계 컨볼루션 뉴럴 네트워크 학습모형은 장 정결도 예측 정확도 1.6% 향상한 78.2% 성능을 달성했다. 또한, 2 단계 컨볼루션 뉴럴 네트워크 학습모형에 상하좌우 뒤집기 데이터 증강 방법을 적용하여 정확도 2.3% 향상한 80.5%의 성능을 달성했다. 향후에는 장 정결 이미지에 적합한 데이터 증강 방법 개발 및 2 단계 모형의 구조를 수정하여 성능을 높일 예정이다.

감사의 글

감사의 글은 2-3 줄로 짧게 기술하기 바랍니다.

참고문헌

- [1] 조유영, 김현욱, “동영상 교육이 대장내시경 대상자의 장 정결 이행에 미치는 영향,” 대한간호학회지, 제 45 권, 제 5 호, pp. 704-712, 2015.
- [2] Zhou J, Wu L, Wan X, et al, “A novel artificial intelligence system for the assessment of bowel preparation (with video),” Gastrointest Endosc, 91(2), pp. 428-435.e2, 2020.
- [3] Zhou, Wei, et al, "Multi-step validation of a deep learning-based system for the quantification of bowel preparation: a prospective, observational study," The Lancet Digital Health, e697-e706, 2021, 3.11.
- [4] 김민규, 배현진, “딥러닝 기반 의료영상 분석을 위한 데이터 증강 기법,” 대한영상의학학회지, 제 81 권, 6 호, pp. 1290-1304. 2020.
- [5] 박성욱, 김종찬, 김도연, “양상블 학습 알고리즘을 이용한 컨볼루션 신경망의 분류 성능 분석에 관한 연구,” 멀티미디어학회논문지, 제 22 권, 6 호, pp. 665-675. 2019.
- [6] Ramprasaath R. Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, “Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization” Dhruv Batra; Proceedings of the IEEE International Conference on Computer Vision (ICCV), 2017, pp. 618-626