

위내시경 디지털 영상에서 정상과 위궤양 딥러닝 분류 모델

Classification model of normal and gastric ulcer Deep Learning on endoscopy Digital Image

이경윤¹, 김영재², 김광기^{3*}

GyeongYun Yi¹, Young Jae Kim², Kwang Gi Kim^{3*}

요 약

내시경 장비의 발전으로 위장 질환의 조기 발견 및 치료에 많은 향상이 있다. 따라서, 많은 내시경 이미지와 함께 내시경 이미지 분류에 대한 연구가 활발히 증가하고 있다. 본 논문에서는 위내시경에서 많이 발견되는 질환인 위궤양을 분류하고자 한다. 위궤양은 초기에 적절한 치료를 하지 않으면 합병증을 일으킬 수 있으므로, 초기에 병변을 진단하는 것이 가장 중요하다. 본 연구에서는 ResNet-50 딥러닝 모델을 이용하여 정상과 위궤양을 분류하고자 하였다. 총 1,525개의 이미지를 이용하여 모델을 생성하였고, 효율적인 학습을 위해 데이터 증강을 적용하였다. 제안된 모델의 분류 성능은 정확도 0.9016, ROC 곡선은 0.83으로 확인하였다.

핵심어: 위궤양, 내시경 영상, 딥러닝, 인공지능

Abstract

Due to the development of endoscopic equipment, there has been much progress in the early detection and treatment of gastrointestinal diseases. Therefore, many endoscopic images have been provided, and studies on endoscopic image classification have been increased actively. In this paper, we aim to classify gastric ulcer, a disease frequently found in endoscopy. Gastric ulcers can lead to complications if not properly treated early on, so it is most important to diagnose the lesions early. Our study used the ResNet-50 in-depth learning model. A model was created using a total of 1,525 images, and data enhancement was applied for efficient learning. The classification performance of the proposed model is 0.9016 with accuracy and 0.83 with ROC curve.

Keyword: Gastric Ulcer, Endoscopic Image, Deep Learning, Artificial Intelligent

1 Department of Health Sciences and Technology, GAIHST, Gachon University, Incheon, Korea [Researcher]
e-mail: gyyi0718@gc.gachon.ac.kr

2 Department of Biomedical Engineering, College of Health Science, Gachon University, Incheon, Korea [Professor]
e-mail: kimyj10528@gmail.com

3 Department of Biomedical Engineering, College of Health Science, Gachon University, Incheon, Korea [Professor]
e-mail: kimkg@gachon.ac.kr (Corresponding Author)

* "This research was supported by the MSIT(Ministry of Science and ICT), Korea, under the ITRC(Information Technology Research Center) support program (IITP-2019-2017-0-01630) supervised by the IITP(Institute for Information & communications Technology Promotion), and Institute for National IT Industry Promotion Agency(NIPA) grant funded by the Korea government(MSIT) (No.A0602-19-1032, Intelligent surgical guide system & service from surgery video data analytics)"

Received(August 16, 2019), Revised(November 2, 2019), Accepted(December 9, 2019)

1. 서론

내시경 장비의 발달은 소화기질환의 조기발견과 및 치료에 많은 발전을 가져왔다. 위장관 질환은 내시경 검사를 통한 조기 진단이 무엇보다 중요하다. 2017년 위내시경 검사 결과 판정현황에서는 정상을 제외하고 가장 많이 발견되는 질환은 위염과 위궤양 순이었다 [1]. 위궤양은 위점막이 염증에 의해 부분적으로 손상되어있는 상태로, 임상에서 가장 흔하게 접하는 질환 중의 하나이다. 위궤양은 지속적인 통증을 동반하며, 출혈, 천공 같은 합병증을 일으킬 수 있다. 이와 같이 위궤양은 환자의 일상생활에 영향을 줄 수 있으며, 적절한 치료를 하지 않으면 합병증이 발생하거나 재발이 자주 일어난다 [2]. 따라서 위궤양의 경우 초기에 병변을 진단하는 것이 중요하다. 진단의 정확성을 높이기 위해 위장조영술과 확대내시경, 협대역 내시경 등이 이용되고 있다. 그리고 2년마다 국가암검진으로 내시경검사를 하고 있어 내시경 검사의 빈도도 높아지고 있다. 하지만 이러한 내시경 영상의 분석과 내시경 검사의 빈도가 높아지면서, 의사들의 피로도도 증가하고 있다. 따라서 내시경을 시술하는 의사의 피로도로 인해 판독 정확성이 떨어지는 문제를 경감시킬 수 있도록 하는 컴퓨터 보조 시스템의 연구가 필요하다.

내시경 영상 분야에서 컴퓨터 보조진단 시스템으로 다양한 연구가 이루어지고 있으며 최근 딥러닝 기술을 활용하는 연구가 활발히 이루어지고 있다. 딥러닝은 영상 인식, 음성 인식 자연어 처리 등의 연구분야로도 활용되고 있으며, 특히 영상 인식 분야에서 활발한 연구가 이루어지고 있다. 일반적인 영상 인식뿐만 아니라, X-Ray, CT, MRI, 내시경 영상과 같은 의료 영상 데이터 분석 또한 딥러닝 기술을 적용하여 자동 분할 모델, 분류 모델에 대한 연구 결과가 보고되고 있다. 그리고 위의 연구들과 같이 내시경 영상의 연구 분야에서도 기존의 영상처리 알고리즘에 비해 성능이 높은 딥러닝 모델을 활용하고 있다.

2. 관련 연구

2.1 내시경 영상에서 컴퓨터 보조진단 연구

내시경 영상을 이용한 보조진단 관련 연구에서 Liu, Xiaoqi 등은 ME-NBI 내시경 영상에서 HSV 색 공간에서 가우시안 필터를 적용하여 H와 S채널 값과 텍스처 기반으로 병변을 탐지하였다 [3]. 제안된 방법에서는 F-Measure 값과 FPR(False Positive Rate)값이 각각 0.61과 0.16의 결과를 보였고, 병변을 분류하는 모델이 아니라, 비정상적인 패턴을 자동으로 분할하는데 그쳤다. 그리고 후속연구로 협대역영상(M-NBI) 시스템을 이용하여 확대 내시경 영상에서 위병변 유형을 평가하는 방법을 제안하였다 [4]. 제안하는 연구에서는 만성 위염(chronic gastritis), 저등급신생물(low grade neoplasia),

조기위암(early gastric cancer) 세 종류의 병변은 ResNet-50 딥러닝 모델을 적용하여 분류하였다. 정확도는 0.96으로 높은 정확도를 보였으나, 병변이 있는 확대 내시경 영상으로 대상으로만 진행하는 한계점이 있었다.

2.2 딥러닝 기술을 이용한 내시경 영상 연구

김동현 등은 위내시경 영상에서 병변 진단을 위해 딥러닝 기술 중 CNN(Convolutional Neural Network) [5][6] 이용하여 분석하였다 [7]. 총 500장의 영상을 대상으로 실험하였으며, 정상과 비정상적으로 분류하는 모델, 비정상은 용기형, 표면형, 함몰형 등의 위암 영상과 출혈, 위염 위궤양 등의 병변들을 모두 비정상적으로 분류하여 학습을 하였다. 학습결과 정상이미지 68개 중 55개를 정상으로 분류하고 비정상 이미지 56개 중에서 38개를 비정상으로 분류하여 민감도(Sensitivity)와 특이도(Specificity) 값은 각각 0.81과 0.68의 결과를 보였다. 제안된 방법은 민감도와 특이도가 다소 낮은 결과를 보였으며, 비정상 영상의 병변을 구분 없이 학습을 하여 한계점이 있었다.

따라서 본 논문에서는 별도의 영상처리 기술을 거치지 않고, 위내시경 영상에서 정상과 위궤양을 자동으로 분류하기 위해 딥러닝 모델을 사용하여 모델의 성능을 검증해보고자 한다.

3. 위 내시경 영상에서 정상과 위궤양 분류 방법

3.1 개발환경

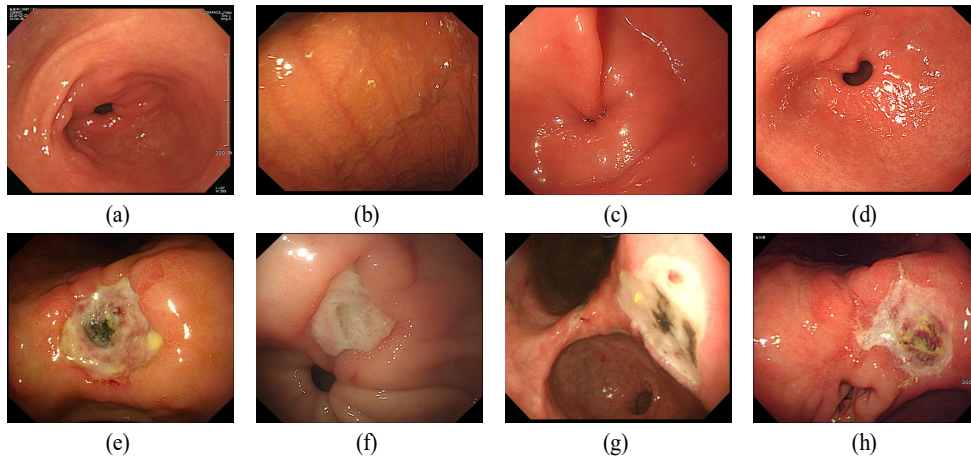
본 논문에서는 딥러닝 학습을 위해 4개의 NVIDIA TITAN Xp (NVIDIA, Santa Clara, Calif) 그래픽 처리 장치와 Xeon E5-1650 v4 (Intel, Santa Clara, Calif) CPU 및 128GB RAM으로 구성된 시스템을 사용하였으며, Ubuntu 14.04 운영체제에서 Python 2.7 및 Keras 2.1.3 (TensorFlow 백엔드 포함) 프레임워크를 통해 모델을 학습하였다.

3.2 데이터

본 논문에서 실험을 위해 상부위장관 내시경 검사를 받은 환자들의 내시경 영상 총 1,525장을 수집하였다. 수집된 영상 데이터는 874장의 정상 내시경 영상과 651장의 위궤양 병변이 있는 내시경 영상으로 구성되었다. 수집된 영상은 24비트의 RGB 칼라의 화소 값을 가지는 jpg 포맷으로 구성되었다. [그림 1]에서는 수집된 정상 내시경 영상과 위궤양 내시경 영상의 예를 보여준다.

수집된 1,525장의 영상 데이터 중, 1,220장은 학습 데이터(Training Data)로 사용되었고, 305장은 테스트 데이터(Test Data)로 사용되었다. 학습 데이터는 699장의 정상 데이터와 521장의 위궤양 데이터로 구성되었고, 테스트 데이터는 175장의 정상 데이터와 130장의 위궤양 데이터로 구성되었다.

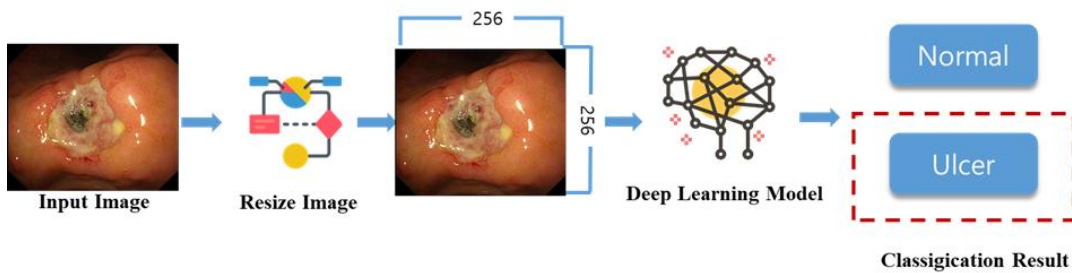
본 논문에서는 학습 데이터의 부족을 보완하기 위해 데이터 증강(Data Augmentation) 기법을 통해 새로운 데이터를 생성하였다. 데이터 증강은 영상 회전(Rotation), 스케일(Scale), 이동(Translation)의 각 지표 간의 무작위 조합으로 이루어졌다. 데이터 증강은 학습 데이터에 대해서만 이루어졌으며, 정상 데이터는 학습 데이터 699장에서 5배 증강하여 3,495장, 위궤양 데이터는 학습 데이터 521장에서 5배 증강하여 2,605장으로 만들어 사용하였다. 이때 증강된 학습 데이터에서 20%는 학습 과정에서 튜닝을 목적으로 한 검증 데이터(Validation Data)로 사용되었으며, 무작위로 선택되었다.



[그림 1] 위내시경 정상영상 (a), (b), (c), (d), 위내시경 위궤양 영상 (e), (f), (g), (h)

[Fig. 1] Dataset of normal gastric images, (a), (b), (c), (d) and gastric ulcer image (e), (f), (g), (h)

3.3 전처리



[그림 2] 위내시경 영상 분류 도식도

[Fig. 2] Flowchart for classification endoscopic images.

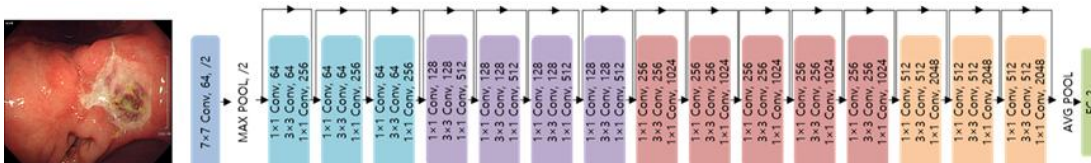
합성곱 신경망 기반의 딥러닝 학습은 구조의 특성상 입력 영상의 가로 및 세로축의 크기가 동

일해야만 한다. 따라서 본 논문에서는 [그림 2]와 같은 전처리 과정을 통해 영상의 가로 및 세로축의 길이를 동일하게 만들었다. 이때 가로축의 길이와 세로축의 길이 간의 비율을 유지한 상태에서 영상을 만들기 위해, 우선 영상의 가로축과 세로축 중 큰 축의 길이를 가진 정사각형의 빈 영상을 만든 후, 중심축을 기준으로 원본 영상을 붙였다. 만들어진 정사각형의 영상들은 모두 동일한 크기인 가로 256 픽셀, 세로 256픽셀로 리샘플링(Resampling) 하여 실험에 사용하였다.

3.4 합성곱 신경망 모델 학습

본 논문에서는 내시경 영상에서 정상과 위궤양을 분류하기 위한 딥러닝 모델로는 ResNet-50을 사용하였다. ResNet-50모델은 ResNet모델에서 컨볼루션(convolution)연산과 연결된 모든 레이어를 계산했을 때, 총 레이어의 개수가 50개가 되는 모델이다 [8][9]. ResNet은 대용량 이미지셋에서 이미지 인식 알고리즘의 성능을 평가하는 ILSVRC 대회에서 2015년에 우승한 모델로 성능이 우수하다. ResNet의 이미지 인식 성능은 오분류율이 3.6%의 수치이며, 이 수치는 사람이 분류하는 수준인 오분류율의 5% 내외를 뛰어넘었다 [10].

ResNet-50의 구조는 [그림 3]과 같고, 본 논문에서 적용한 RetinaNet-50의 학습 환경은 배치 사이즈(Batch Size)는 40, 에폭(Epochs)은 300으로 고정하고, 학습률(Learning Rate)은 0.0001, ImageNet [9]으로 전이학습 하였다.



[그림 3] ResNet-50 구조

[Fig. 3] ResNet-50 Architecture

4. 성능 평가

본 논문에서는 ResNet-50 모델을 이용하여 내시경 영상에서 정상과 위궤양을 자동으로 분류하는 모델을 학습하였다. 학습된 모델의 성능을 검증하기 위해 별도로 수집된 테스트 데이터를 대상으로 학습된 모델에 적용하여 분류 결과를 확인하였다. 학습 모델에 대한 성능은 오차 행렬(Confusion Matrix)을 통해 True Positive (TP), False Positive (FP), True Negative (TN), False Negative (FN)을 각각 산출하고, 아래의 수식에 따라 정밀도(Precision), 재현율(Recall), F1 Score를 계산하였다.

$$\text{Precision} = \frac{TP}{TP+FP} \quad (1)$$

$$\text{Recall} = \frac{TP}{TP+FN} \quad (2)$$

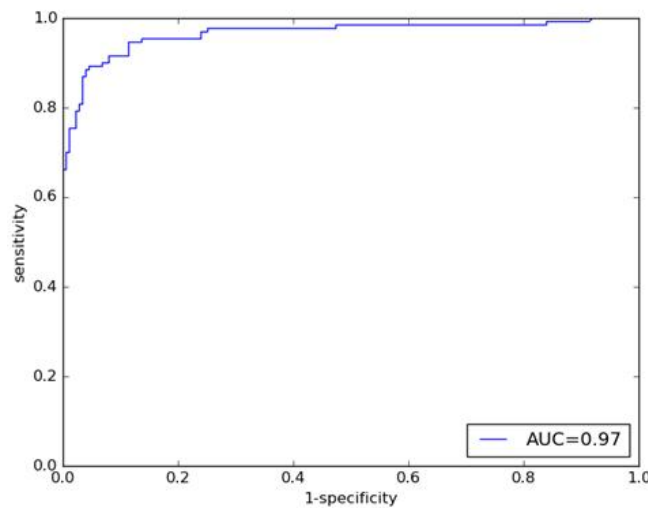
$$\text{F1 Score} = 2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \quad (3)$$

학습된 모델에 대한 오차 행렬을 산출한 결과는 [표 1]과 같고, 정상 데이터는 총 175장 중 170장을 정확히 분류하고 5장을 오분류하여 97.14%의 분류 정확도를 보였다. 또한 위궤양 데이터는 총 130장 중 105장을 위궤양으로 정확히 분류하고, 25장을 오분류하여 80.77%의 분류 정확도를 보였다. 이에 학습된 모델 기반의 위궤양 분류에 대해 95.45%의 정밀도와 80.77%의 재현율, 0.87의 F1 Score를 보였다. 본 논문에서는 분류된 결과에 따라 ROC curve를 분석하여 AUC를 계산하였다. ROC 곡선은 [그림 4]와 같고, 0.97의 높은 AUC를 보였다.

[표 1] 정상 및 위궤양 영상 분류 결과

[Table 1] Classification result for normal images and ulcer images

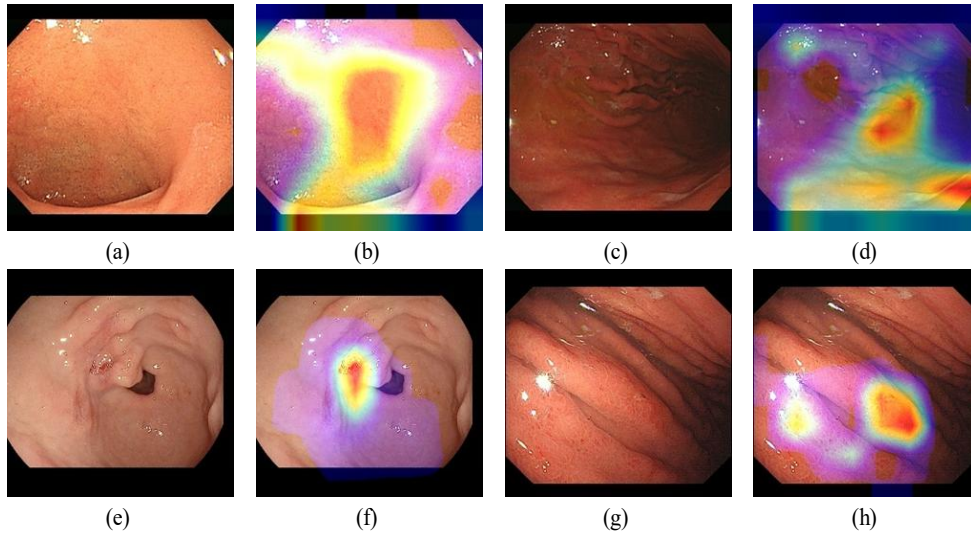
Confusion matrix (50 images)	Predicted Normal	Predicted Ulcer	Accuracy (%)	Total Accuracy (%)
Actual Normal	170	5	97.14	90.16
Actual Ulcer	25	105	80.77	



[그림 4] 분류모델 성능에 대한 ROC 곡선

[Fig. 4] ROC curve on classification model performance

본 논문에서는 학습된 모델이 분류 결과에 미치는 영향을 확인하기 위해 학습된 모델의 Class Activation Map (CAM)을 확인하였다. 그 결과는 [그림 5]이며, 정상 데이터에서 CAM은 영상의 전반에 걸쳐 활성화(Activation)가 이루어졌다. 그리고 위궤양 데이터의 경우, 위궤양이 위치한 부분에서 활성화가 주로 이루어지는 것을 확인하였다.



[그림 5] 클래스 활성화 이미지. 정상 이미지 (a), (b), (c), (d), 위궤양 이미지 (e), (f), (g), (h)

[Fig. 5] The result of class activation map images. (a),(b),(c),(d) normal images, (e), (f), (g), (h) gastric ulcer images

5. 결론

본 논문에서는 위내시경 영상에서 정상과 위궤양 병변을 분류하기 위해 딥러닝 모델인 ResNet-50을 적용하여 분류 모델을 학습하고, 비교 분석을 통해 모델의 성능을 검증하고자 하였다. 총 1,220장의 학습 데이터를 5배 증강하여 구성된 6,100장으로 모델을 학습하고, 별도로 구축된 305장의 테스트 데이터로 검증한 결과, 위궤양 분류에 대해 95.45%의 정밀도와 80.77%의 재현율, 0.87의 F1 Score로 높은 성능을 보였으며, ROC 분석을 통한 AUC도 0.97로 우수한 성능을 보이는 것을 확인하였다. 다만 정밀도에 비해 재현율은 다소 떨어졌으며, 분류에 실패한 영상들을 분석한 결과, 정상에서 위궤양으로 오검출한 경우에는 거품으로 인한 잡음, 위장의 주름 등이 원인으로 파악되었으며, 위궤양에서 정상으로 오검출된 경우에는 일부 영상에서 미세한 위궤양으로 육안으로도 구분이 어려운 경우가 다소 있었음을 확인할 수 있었다. 이는 학습 데이터의 부족이 가장 큰 원인으로 분석되며, 향후 추가 데이터 수집을 통해 딥러닝 모델에 대한 추가 학습이 이루어지고,

미세하여 잘 보이지 않는 위궤양 영역에 대한 영상 개선이 추가적으로 이루어진다면 모델의 성능은 더욱 높아질 수 있을 것으로 판단된다.

References

- [1] Korean Statistical Information Service, “Status of Gastrointestinal Imaging and Endoscopy Results”, kosis.kr, http://kosis.kr/statHtml/statHtml.do?orgId=350&tblId=DT_35007_N140&conn_path=I2, (accessed August 1, 2019).
- [2] J. G. Yoon, J. K. Kim, D. H. Lee, J. I. Kim and S. W. Kim, “Clinical Study for Peptic Ulcer”, The Korean Journal of Gastroenterology, vol. 36, no. 2, August 2000, pp. 336-347.
- [3] X. Liu, C. Wang, J. Bai, G. Liao, and Y. Zhao, “Hue-texture-embedded region-based model for magnifying endoscopy with narrow-band imaging image segmentation based on visual features”, Computer methods and programs in biomedicine, vol. 145, July 2017, pp. 53-66, doi: 10.1016/j.cmpb.2017.04.01.
- [4] X. Liu, C. Wang, J. Bai, and G. Liao, “Fine-tuning Pre-trained Convolutional Neural Networks for Gastric Precancerous Disease Classification on Magnification Narrow-band Imaging Images” Neurocomputing, April 2019, doi: 10.1016/j.neucom.2018.10.100.
- [5] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, “Gradient-Based Learning Applied to Document Recognition” Proceedings of the IEEE, vol. 86, issue. 11, November 1998, pp. 2278-2324, doi: 10.1109/5.726791.
- [6] Y. LeCun, P. Haffner, L. Bottou, and Y. Bengio, “Object Recognition with Gradient-Based Learning”, Shape, Contour and Grouping in Computer Vision, Book. 1681, December 1999, pp. 319-345.
- [7] D. H. Kim and H. C. Cho, “Deep Learning based Computer-aided Diagnosis System for Gastric Lesion using Endoscope”, The Korean Institute of Electrical Engineers, vol. 67, no. 7, July 2018, pp. 928-933, doi: 10.5370/KIEE.2018.67.7.928.
- [8] K. He, X. Zhang, S. Ren, and J. Sun, “Deep Residual Learning for Image Recognition”, 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR), June 26-July 1, 2016, Las Vegas, Nevada, USA, pp. 770-778, doi: 10.1109/CVPR.2016.90.
- [9] K. He, X. Zhang, S. Ren, and J. Sun, “Identity Mappings in Deep Residual Networks”, Computer Vision-ECCV 2016, October 11-14, 2016, Amsterdam, Netherlands, pp. 630-645, doi: 10.1007/978-3-319-46493-0_38.
- [10] A. Krizhevsky, I. Sutskever, and G. E. Hinton, “ImageNet classification with deep convolutional neural networks”, Communications of the ACM, vol. 60, no. 6, June 2017, pp. 84-90, doi: 10.1145/3065386.