

알약 자동 인식을 위한 딥러닝 모델간 비교 및 검증

이경윤[†], 김영재^{**}, 김승태^{***}, 김효은^{****}, 김광기^{*****}

Comparison and Verification of Deep Learning Models for Automatic Recognition of Pills

GyeongYun Yi[†], YoungJae Kim^{**}, SeongTae Kim^{***}, HyoEun Kim^{****}, KwangGi Kim^{*****}

ABSTRACT

When a prescription change occurs in the hospital depending on a patient's improvement status, pharmacists directly classify manually returned pills which are not taken by a patient. There are hundreds of kinds of pills to classify. Because it is manual, mistakes can occur and which can lead to medical accidents. In this study, we have compared YOLO, Faster R-CNN and RetinaNet to classify and detect pills. The data consisted of 10 classes and used 100 images per class. To evaluate the performance of each model, we used cross-validation. As a result, the YOLO Model had sensitivity of 91.05%, FPs/image of 0.0507. The Faster R-CNN's sensitivity was 99.6% and FPs/image was 0.0089. The RetinaNet showed sensitivity of 98.31% and FPs/image of 0.0119. Faster RCNN showed the best performance among these three models tested. Thus, the most appropriate model for classifying pills among the three models is the Faster R-CNN with the most accurate detection and classification results and a low FP/image.

Key words: Pill Classification, Object Detection, Deep Learning, Artificial Intelligent, Hospital

1. 서 론

대형 병원에서 환자의 호전 상태에 따라 처방이 바뀌거나 하는 일이 발생하면, 조제되었으나 투약되지 않은 약은 회수하게 된다. 약사는 이렇게 회수된 수백 종류의 알약을 재분류하는 작업을 하고 있다. 재분류 작업은 약사가 육안으로 직접 확인하여 수작업으로 이루어진다. 이는 약사들의 노동 손실뿐만 아니라, 분류 과정에서 실수가 발생할 가능성도 높다. 최근에는 이와 같은 문제점을 해결하기 위해 영상

처리 기술을 통해 알약 분류를 자동화하기 위한 연구들이 다수 보고되고 있다.

이법기 등은 알약의 자동 분류를 위해 알약의 색상 및 크기를 영상처리 시스템을 이용하여 분석하였다. 총 9종의 서로 다른 알약 샘플을 대상으로 하여 알약의 형태를 Circle, Rectangle, Quadrangle, Capsule, 및 Ellipse로 설정하고, 색상은 HSV컬러 모델의 Hue값으로 지정하였다. 제안된 방법에서 측정한 알약의 크기와 KPIC에 등록된 알약의 크기와 오차를 비교했을 때, 최대 0.12mm 정도의 오차를 나타내

* Corresponding Author: Kwang Gi Kim, Address: (21565) 21 Namdong-daero 774 beon-gil, Namdong-gu, Incheon, Korea, TEL: +82-32-458-2770, FAX: +82-32-460-2361, E-mail: kimkg@gachon.ac.kr
Receipt date: Sep. 4, 2018, Revision date: Jan. 15, 2019
Approval date: Jan. 30, 2019

[†] Dept. of Biomedical Eng., Gachon University College of Medicine (E-mail: gyyi0718@gmail.com)

^{**} Dept. of Biomedical Eng., Gachon University College of Medicine (E-mail: kimyj10528@gachon.ac.kr)

^{***} Dept. of Pharmacy., Gil Hospital (E-mail: highrx@gmail.com)

^{****} Dept. of Biomedical Eng., Gachon University College of Medicine (E-mail: skeblue27@daum.net)

^{*****} Dept. of Biomedical Eng., Gachon University College of Medicine

* This research was supported by the Bio & Medical Technology Development Program of the National Research Foundation (NRF) funded by the Ministry of Science & ICT(2017M3A9E2072929).

었다[1]. 제안된 방법은 알약의 정보 추출의 정확도는 높지만, 비슷한 크기와 색상을 지닌 알약을 분류하는데 한계점이 있었다. 이영범은 영상처리 기술을 기반으로 정상 알약과 불법 알약을 구별하기 위한 매칭시스템인 Pill-ID를 개발하였다. 제안된 방법은 알약의 각인을 SIFT알고리즘과 MLBP descriptor를 사용하여 가장 유사도가 높은 알약 top-20개의 데이터를 선정하였다. 1029개의 불법 알약 이미지와 12,002개의 합법 알약으로 이루어진 데이터베이스로 실험하여 top-20개의 데이터에서 검색하고자 하는 알약이 rank-one으로 분류되는 경우는 73.04%, top-20개의 데이터로 선정되는 경우는 84.47%의 정확도를 보였다[2]. 이 방법은 유사도가 높은 20개의 알약을 제시하기 때문에 실제 대형 병원의 재분류 작업에는 적합한 모델이 아니고, descriptor를 이용한 방법은 빛에 의한 잡음에 취약하여 인식률을 더 개선시키기 어려운 단점이 있다.

앞의 두 연구 모두 알약의 크기, 모양과 같은 특징을 추출하여 기존의 알약의 데이터베이스와 매칭하는 방법을 적용하였다. 특징을 매칭하는 방식의 경우, 알약의 특징이 비슷한 경우 매칭의 실패율이 높아지는 단점이 있다. 이는 영상으로부터 영상처리 기술을 통해 특징을 직접 추출해서 사용하는 방법으로, 분류의 최적화된 특징을 찾아내기가 쉽지 않고, 과적합의 위험성도 높다. 이에 최근에는 영상으로부터 직접 데이터에 최적화된 특징들을 추출하고 분류함으로써 이미지 인식 및 검출에 강점을 보이는 인공지능 기술이 주목을 받고 있으며, 전 세계적으로 다양한 인공지능 관련 연구들이 이루어지고 있다. 예를 들어 사진을 찍어서 이미지에 있는 글자를 추출하여 자동으로 검색해주는 사전 시스템이나 필기체 인식 등과 같이 다양한 분야에서 비전 분야의 문제 해결을 위해 널리 활용되고 있다. 2012년도 ILVRC 대회 우승 모델인 AlexNet[3]은 기존의 이미지 인식 오류율을 크게 낮춤으로써 획기적인 기술 개선이 있었다. 그리고 2015년 ResNet은 오류율을 3.5%까지 낮춤으로써 인간의 인지 오차율과 근접한 수준에 이르게 되었다[4].

이와 같은 인공지능 기술을 기반으로 알약 분류를 시도한 연구 결과도 보고된 바 있다. 김대욱은(2017) 스마트폰으로 알약의 사진을 찍어 인공지능 기술을 통해 알약을 인식할 수 있는 연구를 진행하였다. 제안하는 방법은 알약에서 에지(Edge)를 추출하여 알

약의 모양과 글자 정보를 추출한다. 그리고 추출된 글자는 Deep Learning 학습으로 글자를 인식하여 글자 정보를 획득하고, 추출된 모양 정보와 취합하여 알약을 분류한다[5]. 하지만 이 방법은 알약 영상에서 에지(Edge)를 추출하지 못하는 경우, 글자 정보가 손실되어 인식률이 저하되는 문제점이 있었다.

이에 본 논문에서는 별도의 영상처리 기술을 거치지 않고, 검출과 분류를 동시에 수행하는데 특화된 딥러닝 모델을 사용하여 알약을 자동으로 검출 및 분류하고자 하였다. 또한 YOLO, Faster R-CNN, RetinaNet과 같이 다양한 딥러닝 모델들을 통해 알약을 검출 및 분류하고, 알약 분류에 적합한 딥러닝 모델을 찾기 위해 각 모델 간의 성능을 비교 검증해 보고자 하였다.

2. 제안한 방법

2.1 개발환경

본 논문에서 딥러닝 학습을 위한 시스템은 4개의 NVIDIA TITAN Xp (NVIDIA, Santa Clara, Calif) 그래픽 처리 장치, Xeon E5-1650 v4 (Intel, Santa Clara, Calif) CPU 및 128 GB RAM으로 구성하였으며, Ubuntu 14.04 운영체제에서 Python 2.7 및 Keras 2.1.3 (TensorFlow 백엔드 포함) 프레임 워크를 통해 수행되었다. 딥러닝 학습에 사용한 이미지는 로지텍 BRIO 4K PRO로 촬영하였다.

2.2 데이터

데이터 셋에는 Fig. 1과 같이 10종류의 알약을 사용하였다. 알약의 특징 추출을 하기 위해, 카메라와 알약의 거리, 조도와 배경을 같은 환경으로 설정하여 이미지를 촬영하였다. 각인된 알약의 이름은 알약을 구별하기 위한 특징 중 하나로, 약의 이름이 각인된 표면을 촬영하였다. 알약 이미지를 한 클래스당 100개의 이미지를 사용하였고, 총 10개의 클래스로 구성하였다. 학습에 사용된 데이터의 수는 전체 데이터의 90%이며, 검증에 사용된 데이터는 10%로 나누어 데이터를 설정하였다. 알약은 실제로 가천대 길병원에서 처방되고 있는 약으로 선정하였으며, 색과 크기가 비슷한 고풍제로 뭉쳐있는 타입인 테블릿 형태의 알약 9종류와 1종류의 캡슐 형태의 알약을 사용하였다. 본 논문에서 사용한 알약의 종류는 Fig. 1에 나타내었다.

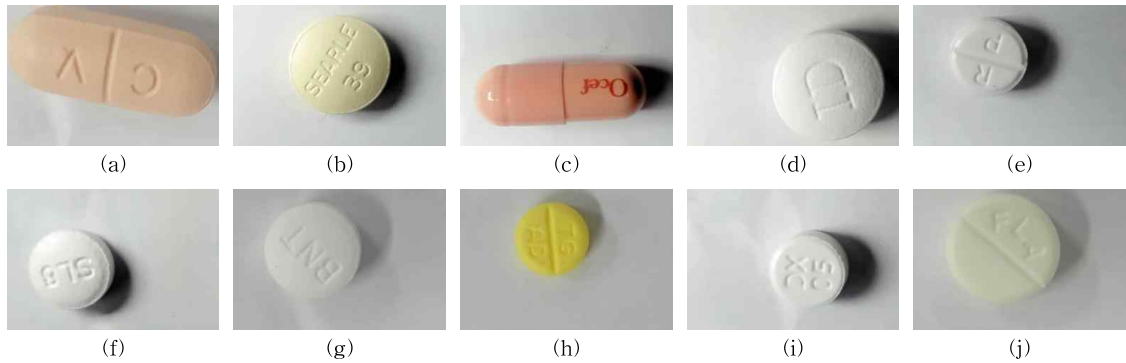


Fig. 1. Dataset of pill types, (a)Cravit, (b)Aladacton, (c)Omnicef, (d)Curan, (e)Repison, (f)Soleton, (g)Banana, (h)Adipam, (i)Dexamethasone, (j)Flasinyl.

2.4 모델 학습

YOLO(You Only Look Once) 모델은 Fig. 2(a)와 같은 구조로 구성되어 있으며, 초당 45프레임의 속도로 이미지를 처리하며 실시간 검출이 가능하다는 장점이 있는 객체검출 알고리즘이다. 입력 이미지를 $n \times n$ 개의 그리드(Grid)로 분할하여 각 그리드의 확률을 계산하여 가장 높은 정확도를 가진 객체의 경계 상자(Bounding Box)를 찾고, 객체의 클래스 확률을 예측하여 객체를 검출한다[6, 7, 8].

본 논문에서 YOLO v3를 사용하였고, 학습 환경은 배치 사이즈(Batch Size) 64, 에폭(epochs) 100을 고정하고, 학습률(Learning Rate)은 0.0001로 설정하여 학습하였다.

Faster R-CNN은 CNN[9, 10]을 기반으로 한 객체 검출 알고리즘중 하나로, RPN(Region Proposal

Network)과 컨볼루션 맵을 공유하여 객체를 검출한다[11, 12]. Faster R-CNN에 대한 구조는 Fig. 2(b)와 같다. 본 논문에서의 적용한 Faster R-CNN 학습 환경은 배치 사이즈 16, 에폭은 100으로 고정하고, 학습률은 0.00001로 설정하여 학습하였다.

RetinaNet은 ResNet[13]을 기반으로 한 Backbone network와 2개의 전용 Sub network로 구성된 하나의 통합 네트워크 모델이고[14], Fig. 2(c)와 같은 구조이다. 본 논문에서 적용한 RetinaNet의 학습 환경은 배치 사이즈는 16, 에폭은 100으로 고정하고, 학습률은 0.00001로 설정하여 학습하였다.

3. 실험 결과 및 고찰

본 논문에서 각 딥러닝 모델에 대해서 데이터를 Table 1과 같이 10-fold cross validation을 검증데이

Table 1. The Result of the cross-validation for trained each model

	YOLO		Faster R-CNN		RetinaNet	
	Sensitivity	FPs/image	Sensitivity	FPs/image	Sensitivity	FPs/image
1	0.9429	0.0286	1.0	0	1.0	0
2	0.9423	0.0577	1.0	0	1.0	0
3	0.951	0.0294	1.0	0	1.0	0
4	0.8713	0.0792	1.0	0	1.0	0
5	0.97	0.02	1.0	0	0.98	0.01
6	0.9	0.09	1.0	0	1.0	0
7	0.899	0.0505	1.0	0	1.0	0
8	0.9192	0.0505	1.0	0	1.0	0
9	0.8265	0.0612	0.9592	0.051	0.8673	0.0918
10	0.8776	0.0408	1.0	0.0408	0.9796	0.0204
ALL	0.9105	0.0507	0.996	0.0089	0.9831	0.0119

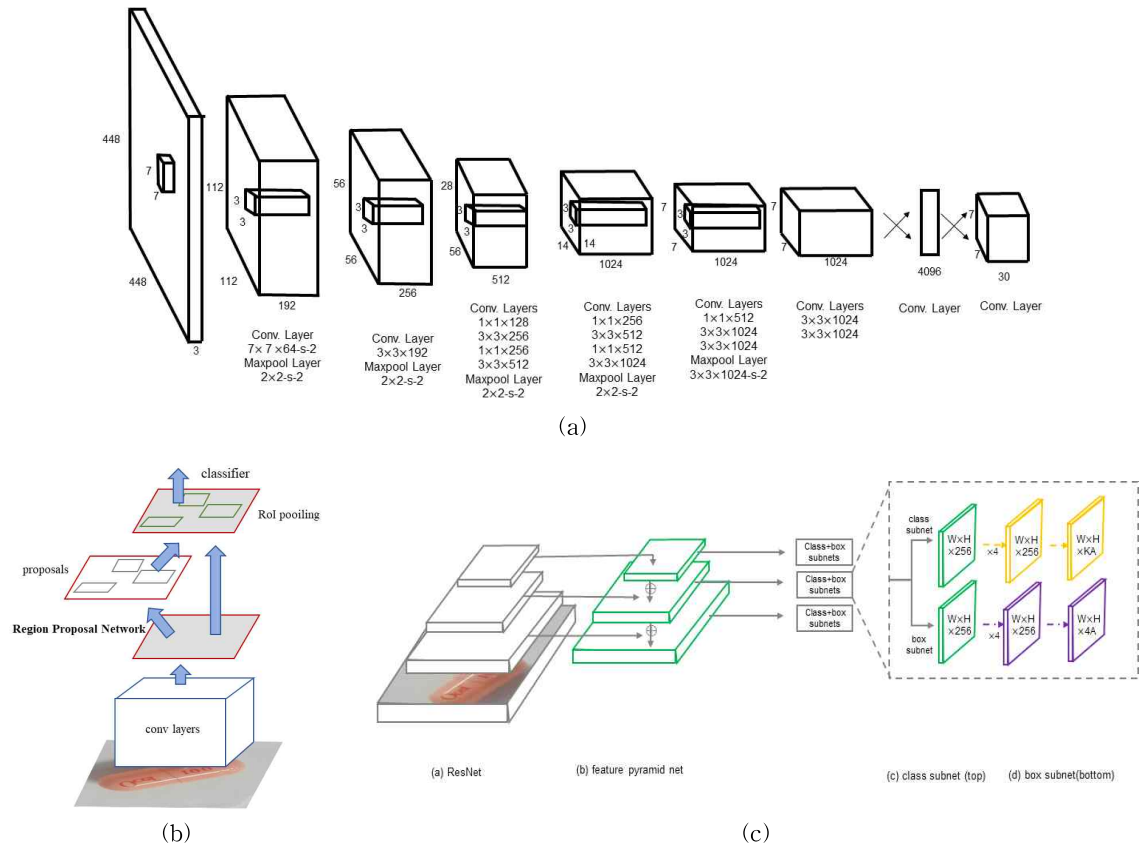


Fig. 2. Architecture of each model, (a)architecture of YOLO model, (b)architecture of Faster R-CNN, (c)architecture of RetinaNet.

터를 적용하고, 그 결과의 평가척도로 객체 검출에서 알고리즘의 성능을 평가하기 위해 사용하는 지표인 IOU(intersection over union)를 사용하였고, IOU>=0.5로 설정하였다[15]. 각 딥러닝 모델의 객관적인

평가를 위해 FROC Curve를 이용하였고[16], FROC Curve 그래프는 Fig. 3과 같다. 실험결과 YOLO 모델의 민감도는 91.05%, Faster R-CNN모델의 민감도는 99.6%, RetinaNet모델의 민감도는 98.31%이다.

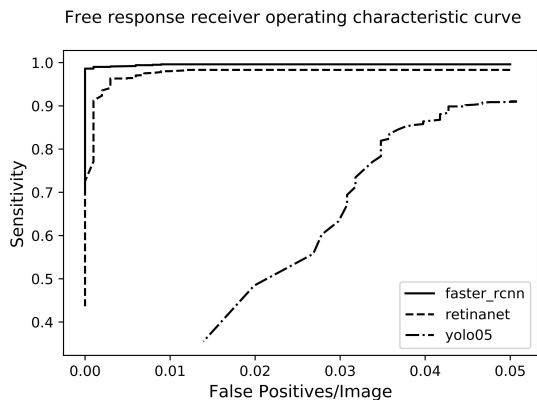


Fig. 3. FROC Curve comparing the each model.

Fig. 4는 YOLO, Faster R-CNN, RetinaNet 각각의 모델에서 알약을 검출 및 분류한 결과이다. 실제 경계 상자는 연두색이고, 예측된 경계 상자는 빨간색으로 표시하였다.

Fig. 5는 YOLO, Faster R-CNN, RetinaNet 모델에서 알약 검출에 실패한 경우이다.

4. 결론 및 토의

본 논문에서는 알약을 자동으로 분류하기 위해 딥러닝 모델인 YOLO, Faster R-CNN, RetinaNet의 3가지 모델을 통해 촬영된 알약 영상들을 학습하여 분류 모델을 각각 만들고, 각 모델간 성능을 비교 검

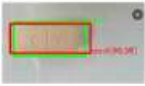



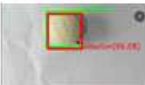
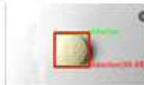

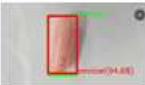
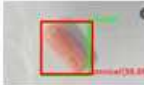















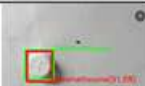





	YOLO	Faster RCNN	RetinaNet
Cravit			
Aladacton			
Omnicef			
Curan			
Repison			
Soleton			
Banan			
Adipam			
Dexamethasone			
Flasinyl			

Fig. 4. The result of detection 10 kinds of pills for each model.

증 해보았다. 실험에 사용한 영상의 개수가 많지 않았기 때문에 검증 결과의 신뢰성을 높이기 위해 각 모델별로 10-fold 교차검증을 수행하였다. 그 결과, 각 알약에 대해서 100장이라는 적은 학습 데이터에도 불구하고, 뛰어난 성능을 보이는 것을 확인하였다.

YOLO 모델은 10-fold 교차검증 결과, 평균 91.05%의 민감도를 보였고, Faster R-CNN 모델은 평균 99.6%의 민감도, RetinaNet은 평균 98.31%의 민감도를 보였다. FP/image는 YOLO 모델이 0.0507, Faster RCNN 모델이 0.0089, RetinaNet 모델이 0.0119의 수치를 보였다. 모델의 성능을 비교한 결과, 민감도는 Faster R-CNN, RetinaNet, YOLO의 순서로 높았으며, FP/image는 Faster R-CNN, RetinaNet, YOLO

의 순서로 낮았다.

각 모델별 검출 및 분류에 실패한 영상들을 살펴보면 주로 흰색의 원형 알약의 경우가 많았다. 이는 흰색의 원형 알약이 표면에 새겨진 문자만 다르고 크기가 작고 형태가 비슷하기 때문인 것으로 파악된다. 또한 본 실험에 사용된 데이터의 개수가 10 종류의 알약을 분류하기에는 다소 부족하다. 따라서 향후 흰색 알약을 포함하여 각 알약들의 데이터들에 대하여 다양한 환경에서 촬영된 영상들을 추가적으로 수집하여 일반화된 데이터를 구축하고, 이를 통해 학습 모델을 보완할 필요가 있다.

지금까지의 결과들을 종합한 결과, 3개의 모델 중 알약을 분류하는데 가장 적합한 모델은 가장 정확한


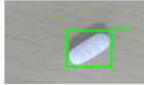


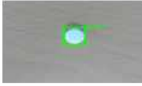




YOLO				
	Adipam	Cravit	Flasinyl	Repison
Faster RCNN				
	Banan	Cravit		
RetinaNet				
	Banan	Cravit	Flasinyl	

Fig. 5. The result of detection failure pills for each model.

검출 및 분류 결과와 낮은 FP/image를 보인 Faster R-CNN이라고 판단된다. 다만 이 연구결과는 정지 영상에서 각 모델의 성능을 비교한 것으로써, 속도에 대한 문제는 배제된 상태에서 진행된 것이다. 따라서 실시간 검출과 같이 속도가 중요한 경우에는 속도가 느리다고 알려진 Faster RCNN을 최적의 모델로 선택하기에는 다소 무리가 있을 수 있다. 따라서 속도까지 반영하여 모델을 비교하기 위해서는 각 모델별 속도를 측정하는 것이 필요하다. 본 논문에서는 속도 측정까지는 이루어지지 않았기 때문에 향후 연구에서는 각 모델별 속도를 측정하여 성능을 비교 검증하는 것이 필요하다고 판단된다. 또한 본 논문에서는 별도의 영상처리를 이용한 전처리 단계 없이 원 영상을 그대로 실험에 사용하였지만 조명 보정과 같은 다양한 전처리 과정을 거쳐 학습을 시킬 경우 향상된 결과를 얻을 수 있을 것으로 생각된다.

결론적으로, 향후 다양한 환경에서의 추가 데이터를 확보하고 다양한 전처리 과정과 속도 측정 등을 통한 추가 연구가 이루어진다면 알약 모델의 성능을 높이고 최적의 모델을 선택하는데 도움이 될 수 있을 것으로 생각된다. 그리고 분류할 알약의 개수를 늘리는 연구가 추가적으로 진행된다면 실제 알약을 자동 분류하는 시스템을 만드는데 많은 도움을 줄 수 있을 것으로 기대한다.

REFERENCE

[1] B.K. Lee and S.G. Kwon, "Automatic Classification System of Tablets with Various

Colors and Shapes," *Journal of Korea Multi-media Society*, pp. 659-666, Vol. 21, No. 6, 2018.

- [2] Y.B. Lee, *Matching and Retrieval of Drug Pill Images*, Master's Thesis of Korea University, 2011.
- [3] A. Krizhevsky, I. Sutskever, and G.E. Hinton, "Imagenet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems*, pp. 1097-1105, Vol. 60, No. 6, 2012.
- [4] B.Y. Yoo, W.H. Hwang, S.J. Han, S.M. Rhee, J.B. Kim, J.J. Han, et al., "The Trend of Image Recognition Close to The Human Level in Deep Learning," *Communications of the Korean Institute of Information Scientists and Engineers*, pp. 32-41, Vol. 33, No. 9, 2015.
- [5] D.U. Kim, *Shape and Text Imprint Recognition of Pill Image Taken with a Smartphone*, Master's Thesis of Seoul National University, 2017.
- [6] R. Joseph, D. Santosh, G. Ross, and F. Ali, "You Only Look Once: Unified, Real-Time Object Detection," *Proceeding of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 779-788, 2016.
- [7] YOLO: Real-Time Object Detection, <https://pjreddie.com/darknet/yolo/> (accessed May, 21, 2018).

- [8] How to Train YOLOv2 to Detect Custom Objects, <https://timebutt.github.io/static/how-to-train-yolov2-to-detect-custom-objects/> (accessed May, 25, 2018).
- [9] Y. LeCun, L. Bottou, Y. Bengio, and P. Haffner, "Gradient-Based Learning Applied to Document Recognition," *Proceedings of the IEEE*, Vol. 86, No. 11, pp. 2278-2324, 1998.
- [10] D.A. Forsyth, J.L. Mundy, V.D. Gesú, R. Cipolla, *Shape, Contour and Grouping in Computer Vision*, Springer Verlag, Berlin, Heidelberg., 1999.
- LeCun, Yann, et al. "Object recognition with gradient-based learning." Shape, contour and grouping in computer vision. Springer, Berlin, Heidelberg., 1999. 319-345.*
- [11] R. Girshick, "Fast R-CNN," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1440-1448, 2015.
- [12] S. Ren, K. He, R. Girshick, and J. Sun "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *Advances in Neural Information Processing Systems*, pp. 91-99, Vol. 39, No. 6, 2015.
- [13] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 770-778, 2016.
- [14] T.Y. Lin, P. Goyal, R. Girshick, K. He, and P. Dollar, "Focal Loss for Dense Object Detection," *In 2017 IEEE International Conference on Computer Vision (ICCV)*, pp. 2999-3007, 2017.
- [15] F. Ahmed, D. Tarlow, and D. Batra, "Optimizing Expected Intersection-over-union with Candidate-constrained CRFs," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 1850-1858, 2015.
- [16] K.S. Song, H.S. Kang, J.G. Im, M.C. Chung, C.W. Kim, J.H. Kim, et al., "Effects of Digital Image Processing on the Detection of Simulated Lesions in Chest Radiographs: An Experimental Study," *Journal of Korean Radiological Society*, Vol. 28, No. 6, pp. 858-864, 1992.



이 경 운

2016년 2월 신라대학교 컴퓨터공학과(학사)
2018년 8월 부산대학교 의과학과(의학석사)
2018년 10월~현재 가천대 길병원 의료기기 R&D센터 연구원

관심분야 : 의료영상, 의료정보, 인공지능



김 효 은

2013년 2월~현재 호서대학교 전자디스플레이공학부(학사)
관심분야 : 의료영상, 머신러닝



김 영 재

2007년 2월 남서울대학교 컴퓨터학과(학사)
2013년 2월 광운대학교 플라즈마 바이오디스플레이학과(공학석사)
2018년 2월 광운대학교 플라즈마 바이오디스플레이학과(공학박사)

2009년 2월~2017년 4월 국립암센터 의공학연구과
2018년 5월~현재 가천대학교 의과대학 의예과 연구교수
관심분야 : 의료영상, 인공지능, 3D 디스플레이



김 광 기

1996년 2월 안동대학교 물리학과(학사)
1998년 2월 포항공과대학교 물리학과(이학석사)
2005년 2월 서울대학교 의용공학과(공학박사)

2007년 7월~2017년 1월 국립암센터 의공학연구과
2017년 2월~현재 가천대학교 의용생체공학과 교수
관심분야 : 의료영상, 의료정보, 인공지능, 로봇



김 승 태

1984년 2월 성균관대학교 약학과(학사)
1984년 3월 약사면허증 취득
2003년 6월 미국 워싱턴주 약사면허증 취득
2012년 2월~현재 가천대 길병원 근무