

Automatic Anatomical Classification Model of Esophagogastroduodenoscopy Images Using Deep Convolutional Neural Networks for Guiding Endoscopic Photodocumentation

Jung-Whan Park*, Yoon Kim*, Woo-Jin Kim**, Seung-Joo Nam***

*Student, Dept. of Computer Science and Engineering, Kangwon National University, Chuncheon, Korea

*Professor, Dept. of Computer Science and Engineering, Kangwon National University, Chuncheon, Korea

**Professor, Dept. of Internal Medicine and Biomedical Informatics, Kangwon National University, Chuncheon, Korea

***Professor, Dept. of Internal Medicine, Kangwon National University School of Medicine, Chuncheon, Korea

[Abstract]

Esophagogastroduodenoscopy is a method commonly used for early diagnosis of upper gastrointestinal lesions. However, 10-20 percent of the gastric lesions are reported to be missed, due to human error. And countries including the US, the UK, and Japan, the World Endoscopy Organization (WEO) suggested guidelines about essential gastrointestinal parts to take pictures of so that all gastric lesions are observed. In this paper, we propose deep learning techniques for classification of anatomical sites, aiming for the system that informs practitioners whether they successfully did the gastroscopy without blind spots. The proposed model uses pre-processing modules and data augmentation techniques suitable for gastroscopy images. Not only does the experiment result with a maximum F1 score of 99.6%, but it also shows a error rate of less than 4% based on the actual data. Given the performance results, we found the model to be explainable with the potential to be utilized in the clinical area.

▶ **Key words:** Deep Learning, Medical Image Analysis, EsophagoGastroDuodenoscopy, Stomach Anatomy Site Classification, Image Processing

[요 약]

위내시경 촬영은 조기에 위 병변을 진단하기 위해서 주로 사용한다. 하지만 위내시경을 했음에도 불구하고 위 내부를 자세히 관찰하지 못해서 10~20% 위 병변을 놓치는 경우가 생기는 것으로 보고 되고 있다. 미국, 영국, 일본 등의 일부 국가와 세계내시경협회(World Endoscopy Organization)에서는 위내시경 시에 맹점 없는 관찰을 위해서 반드시 촬영해야 할 부위에 대한 촬영지침을 제안한 바 있다. 이에 본 논문에서는 수련의가 내시경을 하는 데 있어 위 내부를 자동으로 맹점 없이 관찰하는데 필요한 딥러닝 기술인 해부학적 분류모델을 제안한다. 제안한 모델은 위내시경 이미지에 적합한 전처리 모듈과 데이터 증강 기술들을 사용한다. 실험결과를 통해 최대 F1 점수 99.6% 분류 성능을 확인하였다. 또한, 실제 데이터를 통한 실험결과에서도 에러율이 4% 미만을 보였다. 이러한 성능을 바탕으로 설명 가능한 모델임을 보여 임상에서의 사용 가능성을 확인하였다.

▶ **주제어:** 딥러닝, 의료영상분석, 상부위장관내시경, 해부학적 위치 분류, 영상처리

- First Author: Jung-Whan Park, Co-author: Yoon Kim, Woo-Jin Kim, Corresponding Author: Seung-Joo Nam
- *Jung-Whan Park (k2p2047@kangwon.ac.kr), Dept. of Computer Science and Engineering, Kangwon National University
- *Yoon Kim (yooni@kangwon.ac.kr), Dept. of Computer Science and Engineering, Kangwon National University
- **Woo-Jin Kim (pulmo2@kangwon.ac.kr), Dept. of Internal Medicine and Biomedical Informatics, Kangwon National University
- ***Seung-Joo Nam (pinetrees@daum.net), Dept. of Internal Medicine, Kangwon National University School of Medicine
- Received: 2021. 01. 18, Revised: 2021. 02. 24, Accepted: 2021. 02. 24.

I. Introduction

위내시경(EGD:EsophagoGastroDuodenoscopy)은 역류성 식도염, 위 십이지장의 궤양과 같은 상부 위장관 질환의 검사와 조기 위암의 발견을 위한 검사방법이다. 검사 방법은 의사들이 위내시경을 촬영하면서 내시경 시스템의 모니터를 통해 육안으로 병변을 진단한다. 의사가 육안으로 병변을 진단할 수 있는 것은 고화질 백색광 등의 기술이 모든 내시경에 적용되어 있기 때문이다. 그로 인해 내시경 영상 연구의 경우 병변 진단에 충분한 화질을 제공하므로 화질 개선에서 영상 분석(해석)으로 바뀌는 추세다. 특히 AI 기술이 발전하면서 딥러닝을 이용한 위내시경 영상 분석 연구가 활발하게 진행되고 있다[1-3]. 최근까지 딥러닝을 이용한 병변 진단 연구들을 메타 분석한 논문[3]에 따르면 AI 알고리즘이 점차 향상되어 경험이 적은 수련의들보다 진단 정확도가 높게 보고되고 있다. 예를 들어 위 논문에 소개된 GRAIDS[4]는 AI 진단 알고리즘, 10년 이상의 전문의, 5년 이상의 준 전문의, 2년 이내의 수련생 각각의 네 개의 그룹을 만들어 평가한 결과 민감도는 94.2%, 94.5% 85.8%, 72.2%가 나왔다. 이처럼 현재 딥러닝을 이용한 내시경 영상 분석 연구는 병변의 정확도를 높이는 것에 초점이 맞춰져 있다.

하지만 의사들이 위내시경을 했음에도 중요한 병변을 놓치는 사례가 10~20% 정도 보고 되고 있다[5-7]. 병변을 놓치는 사례는 의사가 실수로 놓치거나, 위 내부를 꼼꼼하게 촬영하지 못할 때 생긴다. 특히 위 내부를 꼼꼼하게 촬영하지 못한 경우는 숙련도가 낮은 수련의들에게서 주로 발생하고 있다. 따라서 숙련도가 낮은 수련의들이 위내시경을 맹점 없이 관찰할 수 있도록 돕는 위내시경 촬영 가이드 시스템이 필요하다 하지만 현재 실수로 위 병변을 놓치는 경우를 줄이고 모든 부위를 촬영하였는지를 검증해주는 시스템은 없다. 본 논문에서는 위내시경 검사자가 위내시경을 촬영할 때 Fig 1과 같이 필수적으로 찍어야 하는 위치 5곳을 자동으로 찾아주는 위 해부학적 분류모델을 제안한다. 필수적으로 찍어야 하는 위치에 관해서는 위내시경 이미지 획득의 지침을 참고했다[8-10]. 2013년에 일본에서는 Yao K이 제안한 Systematic Screening Protocol for the Stomach (SSS)는 위내시경 의사들이 위의 사각지대까지 촬영하고 나올 수 있도록 22개의 위치가 포함된 위내시경 가이드를 제시했다[8]. 위와 비슷하게 2016년 유럽 소화기내시경학회가 제안한 위내시경 표준 가이드에는 10개의 위치를 찍을 것을 권장했다[9]. 일본과 유럽에서 제시한 표준 지침대로 위 해부학적 분류모델 연구가 기존에 진행되어왔지만, 고전적인 딥러닝을 사용했거나 설명 불가

능한 모델을 제시하였다. 본 연구에서는 우수한 분류모델의 성능을 보이며 설명 가능한 분류모델을 보임으로써 임상에서 사용 가능한 AI 분류모델을 보인다. 본 논문의 구성은 다음과 같다. 2장에서는 위 해부학적 분류모델과 관련된 연구를 기술 한다. 3장에서는 위내시경 이미지 위치 가이드 및 특성을 설명하고 학습에 필요한 이미지 전처리와 위내시경 이미지에 적합한 학습모델을 선정하고 손실함수를 소개한다. 4장에서는 실험 및 실험결과를 정리한다. 5장에서는 4장에서 나온 결과를 가지고 임상에서 사용 가능한 분류모델인지 검증하고 6장에서 결론을 도출한다.

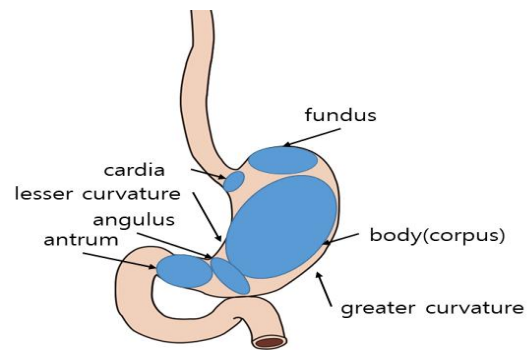


Fig. 1. Stomach Anatomy Site (Cardia, Fundus, Corpus, Angulus, Antrum)

II. Preliminaries

1. Related works

딥러닝을 이용한 위 해부학적 분류모델 연구가 처음으로 제안된 연구는 컴퓨터 비전 분야에서 쓰이는 CNN 모델 중에서 2012년 이미지넷 챌린지에서 우승을 차지한 Alexnet을 이용하여 위의 해부학적 구조를 14가지로 분류한 모델을 2018년에 제안[11]하였으며, 2014년에 제안된 GoogLeNet을 기반으로 후도, 식도, 위, 십이지장과 일본 위암 분류에 따라 위의 위치(위, 중간, 아래)를 분류한 모델을 2018년에 제안하였다[12]. 또한, 일본에서 제안한 SSS 지침에 4개의 위치를 추가하여 26개의 위치를 분류 및 모니터링할 수 있는 강화학습 기반 시스템(WISENSE)이 2019년에 제안되었고[13], 영국에서 제안한 위내시경 검사 지침에 따라 2016년과 2017년에 나온 CNN 모델 중 이미지 챌린지에서 우승한 Resnet50, Densenet121 사용하여 분류한 모델을 2020년에 제안되었다[14]. 기존연구에서 위 해부학적 분류모델을 제안한 방법은 컴퓨터 비전 분야에서 2017년 이전에 나온 CNN 모델을 가지고 진행하였으며, 제안한 분류모델들은 어떻게 정답을 예측했는지 설명할 수 없는 블랙박스의 특징을 가지고 있다. 또한, 기존

연구의 분류문제는 소프트맥스 손실함수를 사용함으로써 학습에 참여하지 않은 위치의 이미지가 분류되는 문제를 발생시키는 한계가 있으며, 가장 최신의 연구[14]에 따르면 분류정확도가 91.1%가 나오므로 임상에서 사용하기에 어려운 점이 있다. 본 연구에서는 학습에 들어가지 않은 위치의 이미지가 오 분류되지 않기 위해 다중 레이블 (Multi-Label)로 문제를 해결하고자 했으며 분류 성과 설명 가능한 모델임을 보이기 위해서 최신 딥러닝 기술을 적용하여 임상에서 사용 가능한 분류모델을 제안한다.

같이 위의 해부학적 위치 형식을 따른다. 해부학적 위치의 주석을 정한 것은 유럽에서 위내시경 촬영 시 최소한 찍어야 할 위치를 정한 ESGE 권고[9]에 따라 11개의 부위를 참고[15]하여 이미지를 수집하였다. 단, 수집한 이미지 중에는 분류에서 제외할 이미지가 포함되어 있다. Fig 3의 a, b와 같이 위내시경 이미지 중 조직검사를 하기 위한 검자가 들어간 사진과, Fig 3의 c, d와 같이 ESGE에서 반드시 찍도록 권고한 부분 외의 부위가 촬영된 이미지가 있다.

III. The Proposed Scheme

1. Data Collection

본 논문에서 사용되는 데이터는 강원대병원의 임상시험 심사위원회(IRB) 허가를 획득하였다. (IRB number 2020-07-016-002). 2015년 1월부터 2020년 11월까지 강원대학교 병원에서 위내시경 검사를 받은 환자의 위내시경 이미지 2457장(Dataset A)을 획득하였다. 딥러닝 모델의 학습과 테스트를 위해서 사용할 이미지는 전문의가 식도, 위, 십이지장 등으로 구분하였다. 또한, 실제 데이터인 60명의 위내시경 이미지(Dataset B) 모델 검증을 위해서 정답이 매겨진 데이터를 획득하였다. 이미지 위치별 주석은 강원대학교 병원으로 받은 위내시경 이미지는 Fig 2와

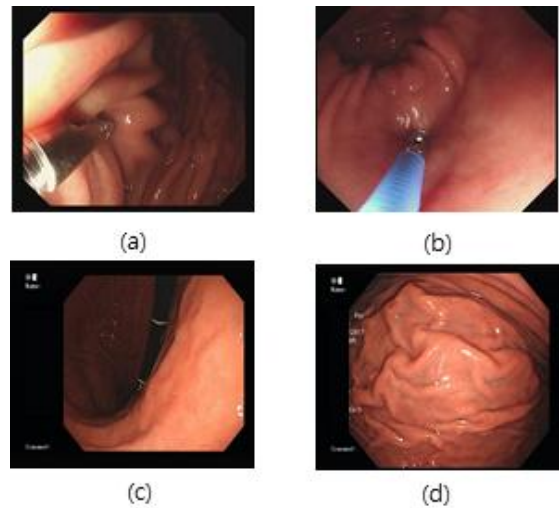


Fig. 3. Example Images to be removed from Classification

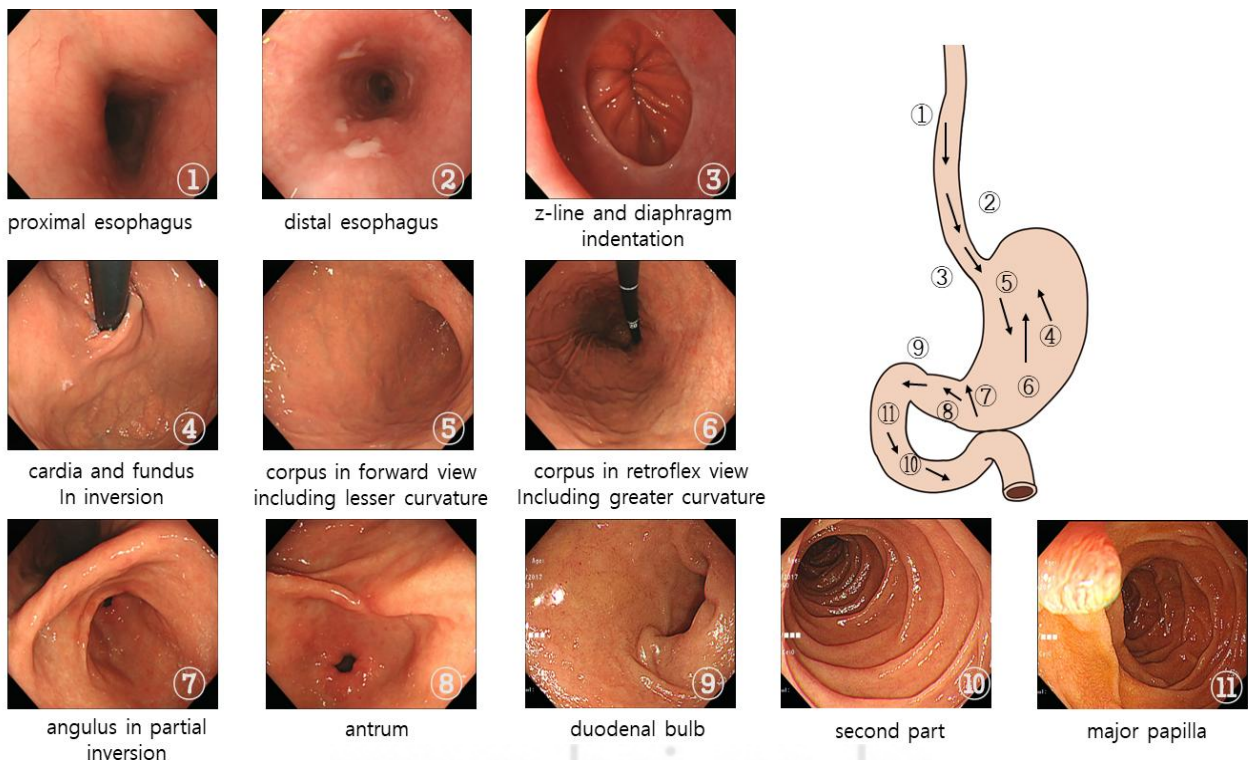


Fig. 2. Anatomical Annotation and Guideline, Preprocessed Image

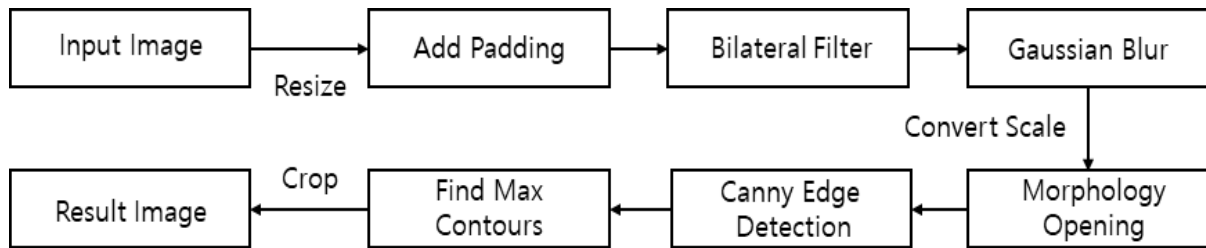


Fig. 4. Process ROI Extraction

2. ROI Extraction

Fig 3과 같이 강원대학교 병원에서 쓰이는 올림푸스 위 내시경의 경우, 내시경 카메라 영상이 그대로 저장되지 않고 병원 내 의료영상 저장 및 전송 시스템인 PACS에 연결되어 의사가 특정 영상의 위치를 캡처하는 방식으로 이미지가 저장된다. 캡처하는 과정에서 올림푸스 위내시경 장비 영상 설정에 따라 Fig 3과 같이 이미지마다 해상도가 다르며, 이미지 옆에 텍스트가 함께 보이는 경우가 생긴다. 따라서 딥러닝의 성능을 높이기 위해 위 이미지만을 추출할 수 있는 전처리 알고리즘이 필요하다. 내시경 이미지 영역만을 추출하는 과정은 Fig 4와 같다. 이미지 전처리 과정은 서로 각기 다른 해상도를 가진 원본 이미지를 640×480 크기로 조정한다. 조정한 이미지는 경계선을 찾아 내시경 이미지를 추출하기 위해서 안쪽 여백을 추가한다. 여백을 추가한 이미지는 경계선을 뚜렷이 하는 양방향 필터(Bilateral Filter)와 잡음(Noise)을 제거할 수 있는 가우시안 블러(Gaussian Blur)을 진행한다. 필터링을 진행한 이미지는 이미지 속 텍스트를 제거를 위한 모폴로지(Morphology)의 오픈링(Opening)을 적용한다. 경계선 검출을 위해 캐니 엣지 검출(Canny Edge Detection)을 사용한 후 가장 큰 경계선의 관심 영역(Region Of Interests)을 찾아 이미지를 자른다.

3. Deep Learning Model

위내시경 이미지에 적합한 분류모델을 제안하기 위해 기존연구에서 사용한 Resnet50, Densenet121과 Attention Module을 포함한 모델인 SE_Resnet50, CBM_Resnet50 그리고 최신의 CNN 모델인 Efficientnet-B1을 사용한다. 기존연구에서 사용한 딥러닝 모델들은 2017년 이전에 이미

지넷 챌린지에서 우승을 거둔 모델들이다. 하지만, 2017년 이후로도 꾸준히 이미지 인식의 성능을 높이려는 방법들이 제안되어왔다. 대표적으로 이미지의 어떠한 특성을 집중해서 학습시키고자 하는 Attention Module을 CNN 모델에 추가하거나 강화학습을 사용하여 최적의 모델을 찾는 방법이 연구됐다. Attention Module을 CNN에 적용하여 이미지넷 챌린지의 분류 성능을 높인 모델은 Squeeze-and-Excitation(SE)[16], Convolutional Block Attention Module(CBM)이 있다[17]. SE와 CBM 모두 범용적인 CNN 네트워크인 VGG, Goog- LeNet, ResNet 등에 쉽게 붙여서 사용할 수 있다. SE는 Feature Map을 Fig 5와 같이 Global Average Pooling을 통해 각 채널의 중요한 정보만 추출해서 학습을 시킬 수 있다. CBM은 SE가 Channel Attention만 고려했다면 Spatial Attention까지 순차대로 계산하는 특징을 가진 모듈이다. 본 논문에서는 Resnet50의 Residual block마다 SE, CBM을 넣은 SE_Resnet50, CBM_Resnet50을 사용한다. 강화학습을 이용하여 찾은 모델 중 대표적인 모델은 Efficientnet이다[18]. Efficientnet은 같은 저자의 이전 논문에서 제안한 MnasNet 모델을 가지고 Width, Depth, Resolution을 조절하며 모델 파라미터 개수에 따른 최적의 분류모델 8개(B0~B7)를 제안하였다. 분류모델 8개 중에서 이미지넷 챌린지의 분류 성능에서 Resnet50 모델보다 2.8% 높은 Efficientnet-B1을 사용한다.

4. Loss Function

본 연구는 다중 레이블로 분류문제를 해결하고자 Loss Function을 Binary Cross-Entropy(BCE) Loss를 사용한다. CNN 모델의 분류기에 시그모이드 활성화 함수를 사용하여 모델의 예측값을 정답 레이블에 대해서 Loss를 계

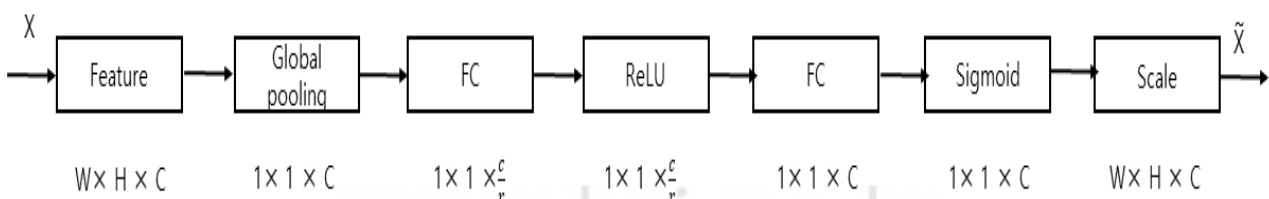


Fig. 5. Design Attention Module - SE

산한다.

$$Loss = -\frac{1}{N} \sum_{i=1}^{N_i} t_i \log(y_i) + (1 - t_i) \log(1 - y_i) - (1)$$

IV. Experiment

1. Dataset

본 실험에서는 Fig 2를 참고하여 Table 1과 같이 클래스를 정의한다. 위의(S1~S5) 위치의 분류 성능을 높이기 위해서 식도와 십이지장의 세부 클래스를 각각 한 개의 클래스로 지정한다. 정지영상만으로 식도와 십이지장의 세부 클래스를 분류하기에는 어려운 점이 있어서 한 개의 클래스로 구분한다. 또한, 내시경 검자가 들어간 이미지를 분류에서 제거하기 위해 Fig 3을 클래스 C로 정의한다.

Table 1. Class Definition for Stomach Anatomy Site

Organ	Segment	Class
Esophagus	Proximal Esophagus	E1
	Distal Esophagus	E2
	Z-line & Diaphragm Indentation	E3
Stomach	Cardia and Fundus in Inversion	S1
	Corpus in Forward View including Lesser Curvature	S2
	Corpus in Retroflex View including Greater Curvature	S3
	Angulus in Partial Inversion	S4
	Antrum	S5
Duodenal	Duodenal Bulb	D1
	Second Part	D2
	Major Papilla	D3

2. Experiment

본 논문에서 제안한 위 해부학적 분류모델의 성능을 평가하기 위해 실험을 진행하였다. 실험은 2203장의 훈련데이터와 254장의 테스트 데이터를 나누어 실험하였다. 실험 방법은 데이터 크기를 224×224와 448×448 나누어 진행하였고, 스크래치 모델의 경우 배치 크기 8, Epoch 100회, 이미지넷 사전훈련 모델을 사용한 경우는 분류기를 수정하여 배치 크기 8, Epoch 30회를 진행하였다. 본 연구에서는 분류 성능 향상을 위해 Bag of Tricks[19]에서 제안된 방법으로 Adam 최적화 기법을 사용하였고, Cosine Annealing Decay을 이용하여 Epoch마다 학습률을 조정하였다. 데이터 증강(Data Augmentation)기법의 경우 Google Brain 팀이 CVPR 2019에서 제안한 방법인

AutoAugment[20] 방법에서 제시한 CIFAR-10에 대해서 제안한 방식을 이용하였다. AutoAugment는 강화학습 기반으로 Imagenet, CIFAR-10, CIFAR-100, Stanford Cars, SVHN에 대한 데이터 세트에 대해 최적의 Data Augmentation을 보여준 방법이다. Augment 방법은, Shear X/Y, Translate X/Y, Rotate, Solarize, Cutout, Invert, Equalize, AutoContrast, Contrast, Brightness, Sharpness Sample Pairing 16가지 방법을 조합해서 5개의 데이터 세트에 대해 각각 5개의 25가지 정책을 제시하였다. 그중 본 학습에서는 CIFAR-10의 데이터 세트에 대해 최적의 데이터 증강을 찾은 정책을 사용하였다. Table 2와 같이 클래스 불균형 문제는 클래스 비율을 고려해서 적은 개수의 클래스는 더 많은 비율로 가중치가 학습되도록 하였다. 실험 환경은 Table 3과 같다.

Table 2. Dataset A Configuration for Model Evaluation

Class	Train(N)	Valid(N)
C	31	4
D	237	30
E	382	49
S1	270	29
S2	328	36
S3	333	36
S4	291	36
S5	331	34

Table 3. Experiment environment

OS	Windows 10 Pro
CPU	Xeon W-2123
GPU	RTX 2080 Ti 11GB
RAM	64GB
Framework	pytorch

3. Performance Metrics

본 모델의 평가는 다중 레이블분류와 클래스 불균형 데이터를 위한 성능 평가지표인 Recall, Precision, F1 점수를 사용하였다. 평가는 모델의 예측값과 정답 데이터 사이의 TR(True Positive), FP(False Positive), FN(False Negative)을 구해서 평가를 진행하였다. Recall-(Sensitivity), Precision(PPV), F1 점수의 수식은 다음과 같다.

$$(True)Recall = \frac{TP}{TP + FN} - (2)$$

$$(True)Precision = \frac{TP}{TP+FP} \quad (3)$$

$$F1\ Score = \frac{2 \times Precision \times Recall}{Precision + Recall} \quad (4)$$

4. Experiment Results & Analysis

Table 4, 5는 실험을 진행한 모든 모델의 민감도 (Sensitivity), 정밀도(PPV), F1 점수를 정리한 표이다. 이미지넷 사전훈련을 사용한 모델의 결과를 보인 Table 5는 Scratch 모델의 결과를 보인 Table 4보다 성능이 0.5~2% 올라갔으며, Resnet50 모델에 Attention Module을 추가한 SE_Resnet50, CBM_Resenet50 두 모델이 1~2% 높게 나왔다. 입력 크기에 따른 결과는 CNN의 경우 입력 이미지의 크기가 클수록 모델이 이미지의 특징을 잘 학습하게 되어 분류 성능을 높일 수 있다고 알려졌지만, 위내시경 이미지의 도메인에서는 클래스 간 비슷한 특성이 있어서 크기에 따른 차이는 두드러지게 나타나지 않았다. 특히 본 논문에서 사용한 데이터 증강 방식에는 기울임, 색상변화, 밝기변화 등을 주로 포함하고 있어 test 데이터에 대한 성능 향상을 가져올 수 있었다. 이러한 분류 모델 비교 실험의 결과를 바탕으로 본 논문에서는 임상에서 사용 가능한 모델의 타당성을 검증한다.

Table 4. Prediction Result Using Scratch Model

Model	Sensitivity	PPV	F1 score
(a) Resnet50_224	0.9764	0.9502	0.9631
(b) Resnet50_448	0.9665	0.9723	0.9704
(c) SE_Resnet50_224	0.9567	0.9681	0.9624
(d) SE_Resnet50_448	0.9685	0.9647	0.9666
(e) CBM_Resnet50_224	0.9724	0.9802	0.9763
(f) CBM_Resnet50_448	0.9764	0.9802	0.9783
(g) Densenet121_224	0.9921	0.9805	0.9863
(h) Densenet121_448	0.9724	0.9686	0.9705
(i) Efficientnet-B1_224	0.9764	0.9612	0.9688
(j) Efficientnet-B1_448	0.9882	0.9580	0.9729

Table 5. Prediction Result Using Pretrained Model

Model	Sensitivity	PPV	F1 score
(a) Resnet50_224	0.9685	0.9685	0.9685
(b) Resnet50_448	0.9606	0.9760	0.9683
(c) SE_Resnet50_224	0.9724	1.0	0.9880
(d) SE_Resnet50_448	0.9724	0.9880	0.9802
(e) Densenet121_224	0.9803	0.9803	0.9803
(f) Densenet121_448	0.9764	0.9920	0.9841
(g) Efficientnet-B1_224	0.9921	0.9921	0.9921
(h) Efficientnet-B1_448	0.9961	0.9961	0.9961

V. Model Validation for Clinical Area

위내시경 검사자가 맹점 없이 위 내부를 촬영했는지 확인하는 시스템에 필요한 분류모델은 Table 1에서 정의한 클래스 S1~S5를 정확하게 예측했는지 여부(Sensitivity, 민감도)뿐만 아니라 S1~S5가 아닌 위치에 대해서 오 분류를 발생시키지 않는 것이 중요하다. 하지만, 본 논문에서 사용한 다중 레이블 분류문제에서는 모델을 평가할 때 모델의 출력값인 클래스 확률이 50%를 기준으로 분류하고 있으므로 민감도는 높게 나올 수 있지만, 분류모델이 학습 시 보지 못한 이미지에 대해서는 오 분류를 할 가능성을 가지고 있다. 따라서 본 논문에서는 실험적으로 클래스 확률의 기준을 높일 때 오 분류가 줄어드는 것을 보이며, 히트맵을 통해서 설명 가능한 모델을 제시하여 임상에서 사용 가능성을 확인한다.

1. Experiment

검증에 사용할 모델은 F1 점수 기준 상위 4개인 Efficientnet-B1_224, 448 Densenet121_448, SE_Resnet50_224이고, 데이터는 60명의 임상 데이터를 갖고 실험을 한다. 데이터는 Table 1에 정의한 클래스와 그 밖의 이미지들을 X로 정의한 클래스로 구성되어 있다. 또한, Table 6과 같이 60명에 대해서 10명씩 6개의 그룹으로

S1~S5의 위치를 모두 포함하는 한 개의 그룹과 각각 한 개의 위치가 빠져있는 5개 그룹으로 나누어져 있다. 실험 방법은 클래스 확률 50%과 확실한 이미지에 한해서 정답으로 분류할 수 있도록 하는 클래스 확률 99%로 나누어 모델을 진행한다. 실험에 이용한 모델은 Table 7로 정의한다.

Table 6. Dataset B Configuration for Assuming Clinical Test

Index	Case	N(Person, Image)
1	All	10, 357
2	No S1	10, 403
3	No S2	10, 345
4	No S3	10, 331
5	No S4	10, 319
6	No S5	10, 303

Table 7. Model Definition

Model	Name
A	Efficientnet-B1_224
B	Efficientnet-B1_448
C	Densenet121_448
D	SE_Resnet50_224

2. Results

실험은 그룹 All과 그 외 남은 그룹에 각각 적합한 방식으로 진행하였다. 그룹 All은 분류모델이 S1~S5를 모두 분류한 사람에 대해서 실험을 하였고, 그 외 남은 그룹에 대해서는 분류모델이 S1~S5 클래스 중 하나라도 빠진 사람을 찾아내었는지 확인하였다. Table 7, 8, 9는 위와 같은 방식의 실험결과이며 Table 7은 클래스 확률 조건을 50%로 한 결과이고 Table 8은 99% 조건으로 한 결과다. Table 9는 Table 7, 8을 바탕으로 사람의 수를 더한 수치와 그에 따른 정확도를 보여준다. Table 10과 같이 클래스 확률이 99%인 경우 50%로 했을 때보다 정확도 향상을 12~25% 가져왔다. 특히 Model D는 60명 중 45명을 정확하게 검증하였다. 자세한 결과 분석과 모델이 놓친 15명에 대해서 S1~S5가 아님에도 오 분류된 이미지와 모델 검증을 위한 히트 맵을 다음 절 Results Analysis & Visualization에서 보인다.

Table 8. Result Assuming Clinical Test of Class Probability 50%

Case	Model			
	A	B	C	D
All	10	10	10	10
No S1	1	3	3	2
No S2	0	0	0	2
No S3	8	4	8	6
No S4	4	4	5	5
No S5	5	4	6	6

Table 9. Result Assuming Clinical Test of Class Probability 99%

Case	Model			
	A	B	C	D
All	9	9	8	9
No S1	4	3	3	6
No S2	3	2	1	3
No S3	10	8	10	10
No S4	10	7	9	9
No S5	7	6	8	8

Table 10. Result Clinical Test

Model	Probability	Sum	Accuracy
A	50%	28	0.467
	99%	43	0.717
B	50%	25	0.417
	99%	35	0.583
C	50%	32	0.533
	99%	39	0.65
D	50%	31	0.517
	99%	45	0.75

3. Results Analysis & Visualization

3.1 Model

4장 Experiment에서의 실험결과는 최신의 CNN 모델인 Efficientnet-B1-448 모델의 Sensitivity가 99.61%의 성능을 보였으나 임상을 고려한 실험에서는 현저히 줄어든 41.7%과 58.3%를 보였다. 하지만 SE_Resnet50-224는 Sensitivity가 97.24%를 보였음에도 모델이 참으로 예측한 것에 대해서 실제 정답이 참인 평가 매트릭(PVP)에서 100%의 성능을 보여주었기에 임상에서의 결과도 가장 높은 51.7%과 75% 성능이 나올 수 있었다. 이로써 범용적인 분야에서 우수한 성능을 보여주고 있는 CNN 모델들이 의료 도메인인 위내시경 이미지의 분류에서 좋은 결과를 내지 못하고 있음을 확인했으며, 오히려 의료 도메인에서 기존에 사용하고 있는 분류모델인 Resnet50에 Attention Module을 추가한 SE_Resnet50이 위내시경 이미지에 적합한 분류모델임을 확인하였다.

3.2 Model Visualization

본 논문에서는 정량적인 평가지표뿐 아니라 설명 가능한 모델임을 보이기 위해 Fig 2의 4~7에 해당하는 S1~S5 위치의 이미지를 가지고 모델의 히트 맵을 시각화하여 Fig 6에 보였다. 히트 맵은 CNN 모델이 'S1' 이미지를 보고 'S1'이라고 예측을 할 때, 그 'S1' 이미지의 어떤 부분을 보고 'S1'이라고 예측했는지 보여주는 특징을 지니고 있다. 이를 보여줄 방법은 CNN Model의 분류기에 Fully Connected Layer가 아닌 Global Average Pooling

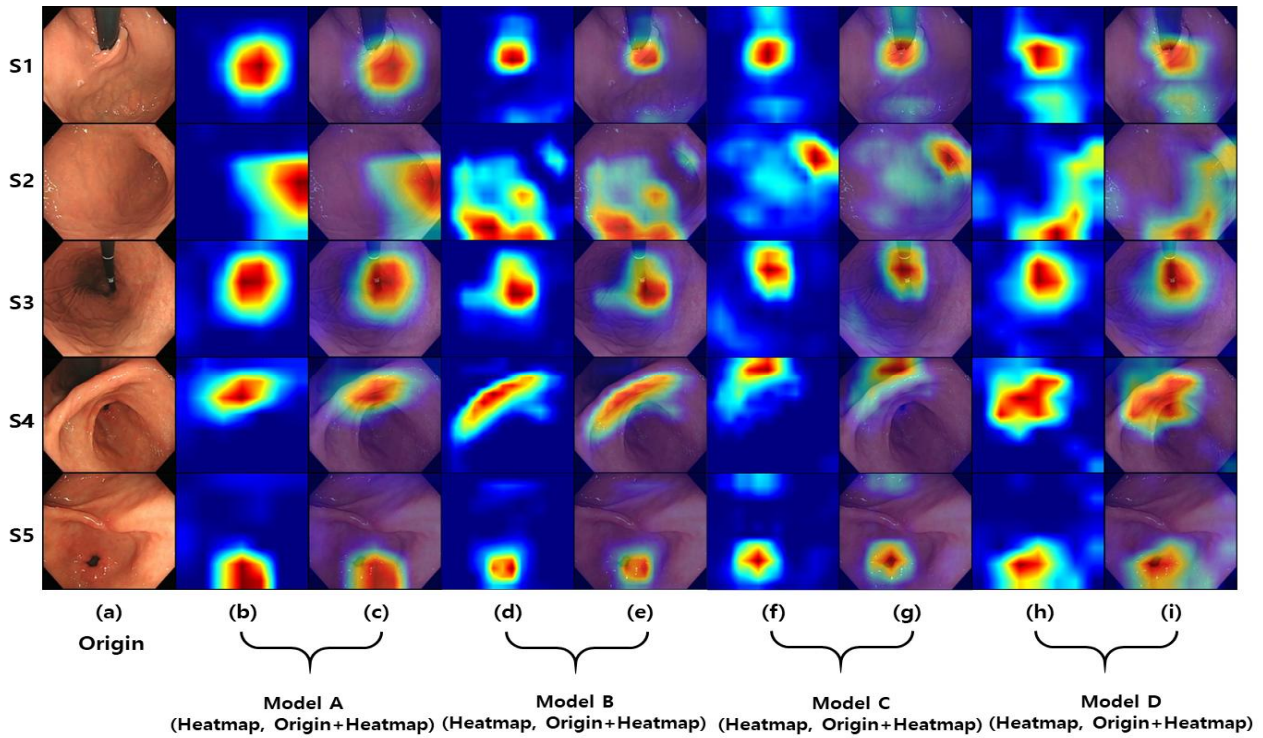


Fig. 6. Grad CAM

Layer를 가지고 있을 때, 학습한 모델을 변경 없이 히트맵을 그릴 수 있게 하는 Grad-CAM[22]을 사용하였다. Model A~D 모두 S1~S5 각각의 특징을 보고 분류하고 있음을 볼 수 있다. 다만 Table 9에서 S2가 다른 위치에 비해 오 분류가 두드러지게 발생하고 있음을 볼 때, Fig 6에서 모델 A와 D가 다른 모델에 비해 S2의 특징인 홀을 보고 예측한 것을 보였다.

3.3 Misclassification Image

1, 2항목 분석을 통해 본 논문에서는 오 분류가 가장 적게 발생한 모델 D인 SE_Resnet50_224를 가지고 오 분류 데이터에 관해서 설명한다. Fig 7은 모델 D에서 오 분류가 발생한 이미지다. Fig 6과 비교했을 때, 모델 D에서 오 분

류가 발생한 이미지는 실제 정답 이미지와 유사한 특성을 보인다. 실제로 같은 위치를 촬영한 이미지이지만 각도에 따른 차이가 있어서 X로 분류된 이미지다. 오 분류가 일어난 이미지 개수는 Table 11과 같다. Table 6에서 보인 그룹별 전체 이미지 수를 고려했을 때 모델이 그룹별 오 분류 한 비율은 최대 4% 미만으로 나왔음을 확인하였다. 또한, 본 모델이 전혀 다른 위치의 이미지를 S1~S5로 오 분류는 발생시키지 않으므로 임상에서 사용할 수 있는 유의미한 결과를 도출하였다.

$$Error\ rate = \frac{Misclassification}{All} \quad (5)$$

Table 11. Result Misclassification Image

Case	Class	N(Image)	Error rate
No_S1	S1	15	0.037
No_S2	S2	9	0.026
No_S3	S3	0	0.000
No_S4	S4	1	0.003
No_S5	S5	2	0.006

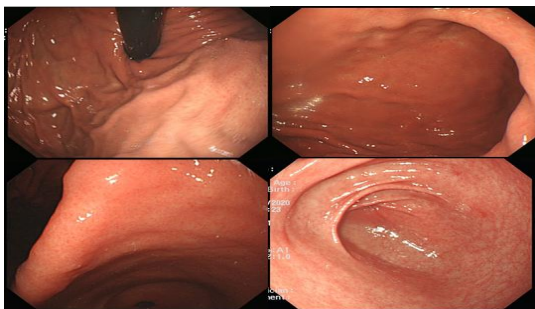


Fig. 7. Example Misclassification Image(Top left : X to S1, Top right : X to S2, Bottom left : X to S4, Bottom right : X to S5)

3.4 Discussion

본 논문에서 보인 위내시경 해부학적 분류모델을 바탕으로 실시간 위내시경 촬영 가이드 시스템을 구축하면 위내시경 검사자가 내시경을 촬영할 때 S1, S2, S3, S4, S5

영역을 모두 검증할 수 있게 되어 맹점 없이 내시경을 하는 데 도움을 줄 수 있을 것으로 보인다.

VI. Conclusions

본 연구에서는 최신 CNN 모델 기반의 위내시경 해부학적 분류모델의 우수한 성능을 검증하고 이를 바탕으로 위내시경 검사자들을 위한 위내시경 촬영 가이드 시스템에 들어갈 분류모델을 제안하였다. 딥러닝 모델은 기존연구에서 사용하고 있는 Resnet50, Densenet121을 비교 모델로 삼았고, 본 논문에서는 위내시경 이미지에 적합한 Resnet50 모델에 Attention Module을 추가한 SE_Resnet50 모델과 최신의 분류모델인 Efficientnet-B1을 선정하여 비교하였다. 모델을 비교할 때 입력 이미지를 224×224와 448×448 두 가지 방법을 적용하였고, 위내시경 이미지 특성에 맞는 이미지 전처리와 데이터 증강 기법을 사용하였다. 실험결과 Dataset A에 대해서 Efficientnet-B1_448이 F1 점수 기준 99.6%로 가장 높게 나왔다. 본 실험결과에서 우수한 분류모델을 4개를 선정하여 실제 데이터인 Dataset B를 가지고 평가를 진행하였고 그 중 SE_Resnet50_224 모델이 어려움이 4% 미만으로 나와서 임상에서 위 해부학적 분류모델로 사용 가능성을 확인하였다. 임상에서 사용 가능한 분류모델을 가지고 위내시경 검사자들에게 위내시경 촬영 가이드 시스템을 제공하면 위 내부를 맹점 없이 촬영하고 나올 수 있는데 도움을 줄 수 있다고 생각한다. 다만, 본 논문은 위 내부를 5개의 클래스가 아닌 더 세분화해서 분류하지 못한 것이 한계점으로 여겨진다. 하지만 후속 연구로 위 전체를 촬영한 위내시경 비디오 데이터를 가지고 딥러닝 모델을 구축하려고 한다. 비디오 데이터를 가지고 분류모델을 구축하면 시간 정보와 데이터를 많이 확보할 수 있으므로 위 내부를 보다 정교하게 분류해낼 수 있어서 위내시경 수련의들에게 더욱 효과적인 촬영 가이드를 제공해줄 수 있을 것을 기대한다.

ACKNOWLEDGEMENT

This work was supported by Institute for Information & communications Technology Planning & Evaluation(IITP) grant funded by the Korea government(MSIT)(No.2019-0-01343, Regional strategic Industry convergence security core talent training business)

REFERENCES

- [1] Misawa, Masashi, Shin-ei Kudo, Yuichi Mori, Hiroki Nakamura, Shinichi Kataoka, Yasuharu Maeda, Toyoki Kudo et al. "Characterization of colorectal lesions using a computer-aided diagnostic system for narrow-band imaging endocytoscopy." *Gastroenterology* VOLUME 150, ISSUE 7, pp 1531~1532, JUNE 01, 2016, DOI: <https://doi.org/10.1053/j.gastro.2016.04.004>
- [2] Urban, Gregor, et al. "Deep learning localizes and identifies polyps in real time with 96% accuracy in screening colonoscopy." *Gastroenterology* Volume 155, Issue 4, pp 1069~1078, October 2018, DOI: <https://doi.org/10.1053/j.gastro.2018.06.037>
- [3] Lui, Thomas KL, Vivien WM Tsui, and Wai K. Leung. "Accuracy of artificial intelligence-assisted detection of upper GI lesions: a systematic review and meta-analysis." *Gastrointestinal endoscopy* Volume 92, Issue 4, pp 821-830. October 2020, DOI: <https://doi.org/10.1016/j.gie.2020.06.034>
- [4] Luo, Huiyan, et al. "Real-time artificial intelligence for detection of upper gastrointestinal cancer by endoscopy: a multicentre, case-control, diagnostic study." *The Lancet Oncology* 20.12 pp 1645-1654, Dec 2019 DOI: [https://doi.org/10.1016/S1470-2045\(19\)30637-0](https://doi.org/10.1016/S1470-2045(19)30637-0)
- [5] Chadwick G, Groene O, Riley S, et al. "Gastric Cancers Missed During Endoscopy in England". *Clin Gastroenterol Hepato* 113, no. 7 pp: 1264-1270, Jul 2015 DOI: <https://doi.org/10.1016/j.cgh.2015.01.025>
- [6] Wang YR, Loftus EV, Jr., Judge TA, Peikin SR. "Rate and Predictors of Interval Esophageal and Gastric Cancers after Esophagogastroduodenoscopy in the United States. *Digestion*", *Digestion* 94, no. 3, pp 176-180, 2016 DOI: <https://doi.org/10.1159/000452794>
- [7] Yalamarthi S, Witherspoon P, McCole D, Auld CD. "Missed diagnoses in patients with upper gastrointestinal cancers. *Endoscopy*", *Endoscopy* 36, no. 10, pp 874-879, Sep 2004 DOI:<https://doi.org/10.1055/s-2004-825853>
- [8] Yao, Kenshi. "The endoscopic diagnosis of early gastric cancer." *Annals of Gastroenterology: Quarterly Publication of the Hellenic Society of Gastroenterology* 26.1 p 11, 2013:
- [9] Bisschops, Raf, et al. "Performance measures for upper gastrointestinal endoscopy: a European Society of Gastrointestinal Endoscopy (ESGE) quality improvement initiative." *endoscopy* 48, no. 09 pp 843-864. Sep 2016 DOI <http://dx.doi.org/10.1055/s-0042-113128>
- [10] Emura, Fabian, et al. "Principles and practice to facilitate complete photodocumentation of the upper gastrointestinal tract: World Endoscopy Organization position statement." *Digestive Endoscopy* 32, no. 2 pp 168-179, Jan 2020 DOI: <https://doi.org/10.1111/den.13530>

- [11] He, Qi, et al. "Deep learning-based anatomical site classification for upper gastrointestinal endoscopy." *International Journal of Computer Assisted Radiology and Surgery* 15, pp 1085-1094 May 2020. DOI: <https://doi.org/10.1007/s11548-020-02148-5>
- [12] Takiyama, Hirotohi, et al. "Automatic anatomical classification of esophagogastroduodenoscopy images using deep convolutional neural networks." *Scientific reports* 8, no. 1 pp 1-8, May 2018 DOI: <https://doi.org/10.1038/s41598-018-25842-6>
- [13] Wu, Lianlian, et al. "Randomised controlled trial of WISENSE, a real-time quality improving system for monitoring blind spots during esophagogastroduodenoscopy." *Gut* 68, no. 12 pp 2161 ~ 2169 DOI: <http://dx.doi.org/10.1136/gutjnl-2018-317366>
- [14] Igarashi, Shohei, et al. "Anatomical classification of upper gastrointestinal organs under various image capture conditions using AlexNet." *Computers in Biology and Medicine* 124 p.103950, Sep 2020 DOI: <https://doi.org/10.1016/j.combiomed.2020.103950>
- [15] Januszewicz, Wladyslaw, and Michal F. Kaminski. "Quality indicators in diagnostic upper gastrointestinal endoscopy." *Therapeutic Advances in Gastroenterology* 13 p.1756284820916693, May 2020 DOI: <https://doi.org/10.1177/1756284820916693>
- [16] Hu, Jie, Li Shen, and Gang Sun. "Squeeze-and-excitation networks." *Proceedings of the IEEE conference on computer vision and pattern recognition*. pp. 7132-7141, 2018.
- [17] Woo, Sanghyun, et al. "Cbam: Convolutional block attention module." *Proceedings of the European conference on computer vision (ECCV)*. pp. 3-19. 2018
- [18] Tan, Mingxing, and Quoc V. Le. "Efficientnet: Rethinking model scaling for convolutional neural networks." *arXiv preprint arXiv:1905.11946*, 2019.
- [19] Cubuk, Ekin D., Barret Zoph, Dandelion Mane, Vijay Vasudevan, and Quoc V. Le. "Autoaugment: Learning augmentation policies from data." *arXiv preprint arXiv:1805.09501*, 2018.
- [20] He, Tong, Zhi Zhang, Hang Zhang, Zhongyue Zhang, Junyuan Xie, and Mu Li. "Bag of tricks for image classification with convolutional neural networks." In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pp. 558-567. 2019.
- [21] Sin-ae Lee, Kim, Dong-hyun, and Hyun-chong Cho. "Deep Learning based Gastric Lesion Classification System using Data Augmentation" *The Transactions of The Korean Institute of Electrical Engineers* 69, no. 7 pp. 1033-1039, 2020, DOI : 10.5370/KIEE.2020.69.7.1033
- [22] Selvaraju, Ramprasaath R., Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. "Grad-cam: Visual explanations from deep networks via gradient-based localization." In *Proceedings of the IEEE international conference on computer vision*, pp. 618-626. 2017.

Authors



Jung-Whan Park is now a B.S. and M.S. student in Department of Computer and Communications Engineering at Kangwon National University. His research interests are in the areas of medical image analysis, deep learning and computer vision.



Yoon Kim received a B.S. degree in 1993, an M.S. degree in 1995, and a Ph.D. degree in 2003, in electronic engineering with the Department of Electronic Engineering from Korea University In 2004,

he joined the Department of Computer Science and Engineering, Kangwon National University, where he is currently a professor From 1995 to 1999, he was with the LG-Philips LCD Co., where he was involved in research and development on digital image equipment. His research interests are in the areas of machine learning, multimedia communications, and computer vision



Woo-Jin Kim received a B.S degree in 1994, am M.S. degree in 2004, in Medical College from Seoul National University, and a PhD. degree in 2006 in Medicine from Hallym university.

In 2004, he joined the Department of Internal Medicine, Kangwon National University, where he is currently a professor. His research interests are in the areas of biomedical informatics.



Seung-Joo Nam received a B.S. degree in 2004 and M.S. degree in 2008 in Medical College from Korea University, and a PhD. degree in 2013 in Graduate School of Medical Science and Engineering from

KAIST. In 2015, he joined the Department of Internal Medicine, Kangwon National University, where he is currently a professor. His research interests are in the areas of immunology and gastrointestinal endoscopy.