

합성곱 신경망을 활용한 위내시경 이미지 분류에서 전이학습의 효용성 평가

박성진¹ · 김영재¹ · 박동균² · 정준원² · 김광기¹

¹가천대학교 의과대학 의공학교실, ²가천대학교 길병원 소화기내과

Evaluation of Transfer Learning in Gastroscopy Image Classification using Convolutional Neural Network

Sung Jin Park¹, Young Jae Kim¹, Dong Kyun Park², Jun Won Chung² and Kwang Gi Kim¹

¹Department of Biomedical Engineering, College of Medicine, Gachon University

²Department of Gastroenterology, Gil Medical Center, Gachon University College of Medicine

(Manuscript received 18 July 2018 ; revised 1 October 2018 ; accepted 10 October 2018)

Abstract: Stomach cancer is the most diagnosed cancer in Korea. When gastric cancer is detected early, the 5-year survival rate is as high as 90%. Gastroscopy is a very useful method for early diagnosis. But the false negative rate of gastric cancer in the gastroscopy was 4.6~25.8% due to the subjective judgment of the physician. Recently, the image classification performance of the image recognition field has been advanced by the convolutional neural network. Convolutional neural networks perform well when diverse and sufficient amounts of data are supported. However, medical data is not easy to access and it is difficult to gather enough high-quality data that includes expert annotations. So This paper evaluates the efficacy of transfer learning in gastroscopy classification and diagnosis. We obtained 787 endoscopic images of gastric endoscopy at Gil Medical Center, Gachon University. The number of normal images was 200, and the number of abnormal images was 587. The image size was reconstructed and normalized. In the case of the ResNet50 structure, the classification accuracy before and after applying the transfer learning was improved from 0.9 to 0.947, and the AUC was also improved from 0.94 to 0.98. In the case of the InceptionV3 structure, the classification accuracy before and after applying the transfer learning was improved from 0.862 to 0.924, and the AUC was also improved from 0.89 to 0.97. In the case of the VGG16 structure, the classification accuracy before and after applying the transfer learning was improved from 0.87 to 0.938, and the AUC was also improved from 0.89 to 0.98. The difference in the performance of the CNN model before and after transfer learning was statistically significant when confirmed by T-test ($p < 0.05$). As a result, transfer learning is judged to be an effective method of medical data that is difficult to collect good quality data.

Key words: Gastroscope, Convolutional Neural Network, Transfer learning, Resnet, Inception, VGGnet

Corresponding Author : Kwang Gi Kim
Department of Biomedical Engineering, Gachon University
College of Medicine, 21, Namdong-daero 774 beon-gil,
Namdong-gu, Incheon, 21565, Republic of Korea
TEL: +82-32-458-2770 / E-mail: kimkg@gachon.ac.kr
E-mail: kimkg@gachon.ac.kr, drgreen@gilhospital.com
Corresponding Author : Jun Won Chung
Department of Gastroenterology, Gil Medical Center,
Gachon University College of Medicine, 21, Namdong-daero
774 beon-gil, Namdong-gu, Incheon, 21565, Republic of Korea
TEL: +82-32-460-2178 / E-mail: drgreen@gilhospital.com
이 연구는 2018년도 과학기술정보통신부의 재원으로 정보통신기술
진흥센터의 지원(2018-0-00861, 의료데이터분석 지능형 SW 기술
개발)과 2017년도 가천의생명융합연구원의 MD-PhD공동과제
(2017-5294) 지원을 받아 수행하였음.

1. 서 론

위암은 2015년도 기준, 국내에서 13.6%로 가장 흔한 암 중 하나이다[1]. 조기 발견하여 수술하였을 때, 5년 생존 기간이 90%로 높은 생존율을 보인다[2]. 국가적으로도 2002년도부터 40세 이상의 사람들에게 위장조영촬영 또는 위내시경을 지원하고 있다[3]. 그 중 위내시경은 이상이 의심되는 부위에 대해 조직검사를 할 수 있는 이점이 있어 조기진단에 매우 유용한 방법이다. 하지만 위내시경은 실시간으로

진단이 이루어지는 방법이기 때문에 의사의 숙련도가 많은 영향을 끼친다. 의사의 주관적 판단의 영향으로 위내시경의 위양 위음성율(False negative rate)은 4.6~25.8%에 이른다[4-7]. 만약에 의사를 보조하여 위내시경 영상을 분류하는 진단 프로그램이 도입 된다면, 의사의 주관적 판단에 의한 일정하지 않은 진단 결과에 도움이 될 것이라고 기대한다.

과거에는 이미지를 인식하고 분류하기 위해 이미지의 파형 및 질감 등을 특징으로 추출하여 서포트 벡터 머신(Support Vector Machine, SVM), 베이지안(Bayesian) 등으로 구분하는 방법을 시도하였다[8-9]. 이러한 방법은 다양한 환경에서 제공되는 이미지를 적용하기에 어려운 단점이 있다. 최근 이미지 인식 분야의 이미지 분류 성능은 딥러닝(Deep learning)의 한 종류인 합성곱 신경망(Convolutional Neural Network, CNN)을 통해 기술발전의 고도화가 이루어지고 있다. 합성곱(Convolution)을 통해 이미지의 특징을 추출하고 서브샘플링(Subsampling)으로 그 크기를 효과적으로 제어하는 CNN 구조는 이미지 인식 분야에서 최고의 성능을 발휘하고 있으며[10-12], 사람 이상의 이미지 분류 수준을 보이고 있다[13].

의료계에서도 합성곱 신경망을 활용한 연구가 활발하게 이루어지고 있다. 대표적인 예로 구글에서 발표한 당뇨병 망막증의 진단[14]과 스탠퍼드 대학에서 발표한 피부암 진단[15]이 크게 주목 받았으며, 일본 암연구병원은 위내시경 이미지에서 위암을 검출하는 연구를 진행하였다[4]. 구글에서 발표한 당뇨병 망막증 진단에 사용된 이미지는 총 139,886 장이며, 그 중 128,175장으로 딥러닝 학습에 사용하여 ROC(Receiver Operating Characteristic)의 AUC(Area Under Curve)값이 0.95 이었다[14]. 스탠퍼드 대학에서 발표한 피부암 진단에는 총 129,450장의 이미지를 사용하여 0.96의 AUC값을 기록하였다[15]. 또한 일본 암연구병원은 13,584장의 위내시경 이미지를 합성곱 신경망에 적용하여 위암 검출 시도하였고 92%의 민감도(sensitivity)를 보였다[4].

합성곱 신경망은 다양하고 충분한 양의 데이터가 뒷받침 되었을 때 좋은 성능을 발휘한다[10-11]. 하지만 많은 양의 데이터를 획득하지 못하였을 때, 이미 충분한 양의 데이터로 학습된 모델의 가중치를 비교적 작은 데이터에 초기값으로 전이하여 재학습 시키는 전이학습(Transfer learning) 방법이 주목 받았다[16-18]. 자동차 이미지로 제조사 및 모델을 분류하는 연구에서 전이학습이 적용되기 전, 후 분류 정확도가 12% 향상 되었고[18], 폐 CT 데이터에서 병변을 검출하는 연구에서는 전이학습이 적용되기 전, 후 AUC가 0.12 향상되었다[16].

본 논문은 서로 다른 구조의 합성곱 신경망(VGG16[19], Inceptionv3[20], Resnet50[21])이 실생활 이미지 125만 장, 1000개의 클래스로 학습된 이미지넷(ImageNet) 모델

의 가중치[22]를 초기값으로 전이 되었을 때와 임의의 초기값으로 학습이 시작되었을 때를 비교하여, 위내시경 영상 분류 및 진단 에서 전이학습의 효용성을 평가한다.

II. 연구 재료 및 방법

본 논문에서 사용되는 데이터는 가천대 길병원의 임상시험심사위원회(IRB)의 허가를 득하였다(2018-052). 가천대 학교 길병원에서 위내시경 검사를 받은 환자의 위내시경 이미지 787장을 획득하였다. 전문의가 각각의 이미지를 진단하여 그림 1과 같이 정상(Normal) 이미지와 비정상(Abnormal) 이미지로 구분 지었으며, 정상 이미지의 수는 200장, 비정상 이미지의 수는 587장으로 구성되었다. 정상 이미지 수가 딥러닝 학습을 진행하기에 비교적 적은 200장 이었기 때문에, 학습 데이터와 테스트 데이터 비율에 대해 적절한 균형이 필요하였다. 최소 150장 정도의 학습데이터가 필요하다고 판단되어 80%를 학습데이터에 사용하였으며 나머지 20%를 최종 모델을 검증하는 테스트(Test) 데이터 세트로 구성하였다. 또한 과적합(Overfitting)을 피하기 위해 같은 방법으로 5번 반복하여 검증하였다[23].

분류 시스템 개발은 우분투(ubuntu)14.04 환경에서 파이썬(Python)2.7 언어로 진행되었으며, 텐서플로우(Tensorflow) 1.4.1을 백엔드(backend)로 하는 케라스(Keras)2.1.3 라이브러리를 활용하여 딥러닝 구조를 구성하였다. 엔비디아 타이탄 엑스피(NVIDIA Titan XP)를 활용하여 딥러닝 학습이 진행되었다.

전체 연구 과정은 그림 2와 같이 크게 이미지 전처리 과정(Image preprocessing)과 딥러닝 학습 과정(Deep learning process)으로 나뉜다. 이미지 전처리 과정은 전문의의 판독(Expert Annotation), 데이터 세트 분할(Train/Test Split), 이미지 크기 조정(Image Resize), 이미지 정규화(Image Normalization)로 이루어졌으며, 딥러닝 학습 과정은 무작위 값을 초기값으로 학습하는 합성곱 신경망 모

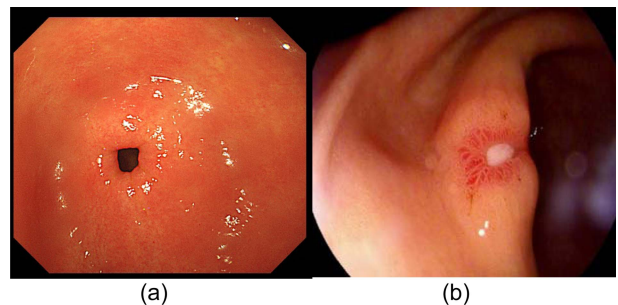


그림 1. 연구에 사용된 이미지. (a)정상 이미지, (b)비정상 이미지.
Fig. 1. Image Used in Research. (a)Normal Image, (b)Abnormal Image.

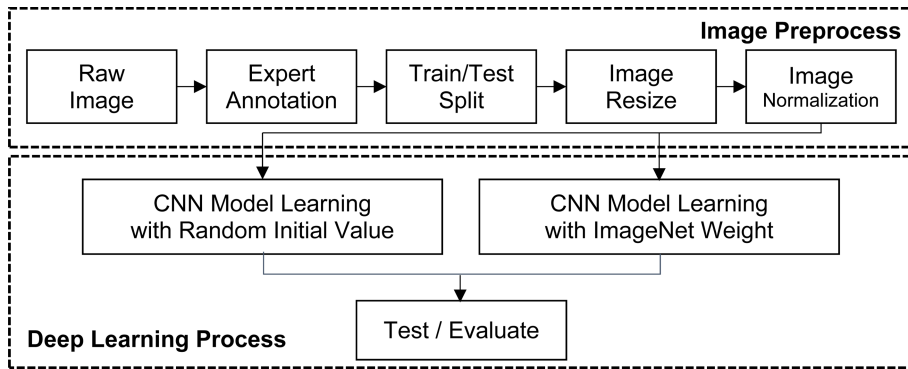
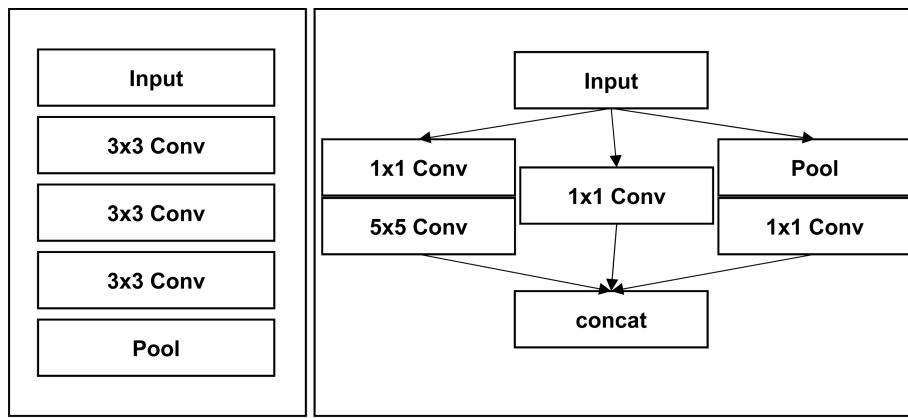
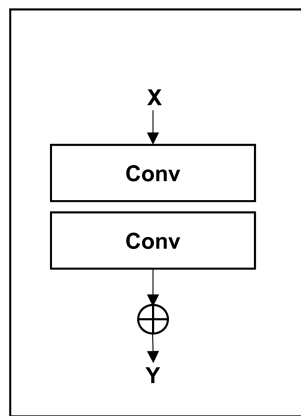


그림 2. 위내시경 이미지 분류 블록도.
Fig. 2. Block Chart of Gastroscopy Image Classification.

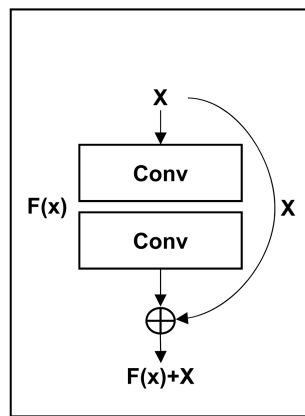


(a)

(b)



(c)



(d)

그림 3. 연구에 사용된 VGG16, Inceptionv3, Resnet50의 주요 구조. (a)3x3 컨볼루션 레이어, (b)1x1 컨볼루션 레이어, (c)기존의 학습 레이어, (d)잔여 학습 레이어.

Fig. 3. The main structure of VGG16, InceptionV3, ResNet50, used in the study. (a) 3x3 Convolution layer, (b) 1x1 Convolution layer, (c) Conventional Learning layer, (d) Residual Learning layer.

델(CNN Model Learning with Random Initial Value) 과 이미지넷 에서 학습된 가중치[22]를 초기값으로 하는 합

성곱 신경망 모델(CNN Model Learning with ImageNet Weight)을 비교 및 평가한다.

이미지넷에서 학습된 조건과 똑같이 만들기 위해서, 이미지 크기는 가로 224 픽셀, 세로 224 픽셀로 재구성하였으며, RGB 각 채널(channel)에 대해 스케일링(scaling) 없이 0을 중심으로 픽셀 분포가 이루어지도록 정규화 하였다. CNN 구조로 이미지 인식 분야에서 좋은 성능을 보이는 VGG16, Inceptionv3, Resnet50을 적용하였다.

1. VGG16

VGG16에 구조적 특징은 그림 3(a)와 같이 3x3 컨볼루션 레이어를 3개 이어 붙이는 것에 있다. 3개의 컨볼루션 레이어를 이어 붙임으로써 7x7 컨볼루션 레이어의 효과를 얻을 수 있으며 더욱이 파라미터의 수가 줄어들고, 학습속도가 향상되는 효과가 있다.

2. Inceptionv3

그림 3(b)와 같이 Inception의 가장 큰 특징은 1x1 컨볼루션 레이어이다. 기본적으로 딥러닝은 망이 깊어질수록 성능이 비교적 좋아진다. 하지만 그만큼 연산량이 증가하여 효율적인 학습이 어렵다. Inception은 1x1 컨볼루션 레이어를 배치하여 파라미터 수를 기존의 CNN 모델에서 1/12만큼 줄였다. 줄어든 연산량 만큼 깊어진 망을 통해 좋은 성능을 보였다.

3. ResNet50

잔차 학습 블록(Residual learning block)은 ResNet의 핵심이다. 기존의 CNN은 그림 3(c)처럼 레이어를 거치고 학습되면서 최적의 Y값을 찾는다. 하지만 그림 3(d)의 잔차 학습 블록은 입력 X를 학습 레이어 이후에 더함으로써 최적의 $F(x)+X$ 를 목표로 학습된다. 즉, 레이어의 학습과정은 입력 값과 출력 값의 차이인 잔차(residual) $F(x)$ 에만 집중하게 된다. 기존의 레이어 학습과정에서 출력된 Y값 보다 잔차 $F(x)$ 는 그 크기가 작기 때문에 망의 깊이가 급격히 증가하여 좋은 성능을 보였다. 또한 컨볼루션 레이어의 학습이 잘 이루어졌을 경우 잔차 $F(x)$ 값은 0으로 수렴한다. 그 결과 입력의 작은 부분을 쉽게 검출 할 수 있게 되었다.

CNN 모델에서 기존에 활용 되었던 확률 기울기 하강(Stochastic Gradient Descent, SGD)을 손실함수(loss function)로 사용하였으며, 모멘텀(momentum)은 0.9로 설정하였다[19-21]. 전이학습되는 가중치가 크게 변화되지 않게 하기 위해 학습율(learning rate)은 0.0001로 설정하였다. 컴퓨팅 파워에 맞춰서 에폭(epoch)수는 300, 배치 사이즈(batch size)는 16인 환경에서 학습하였다.

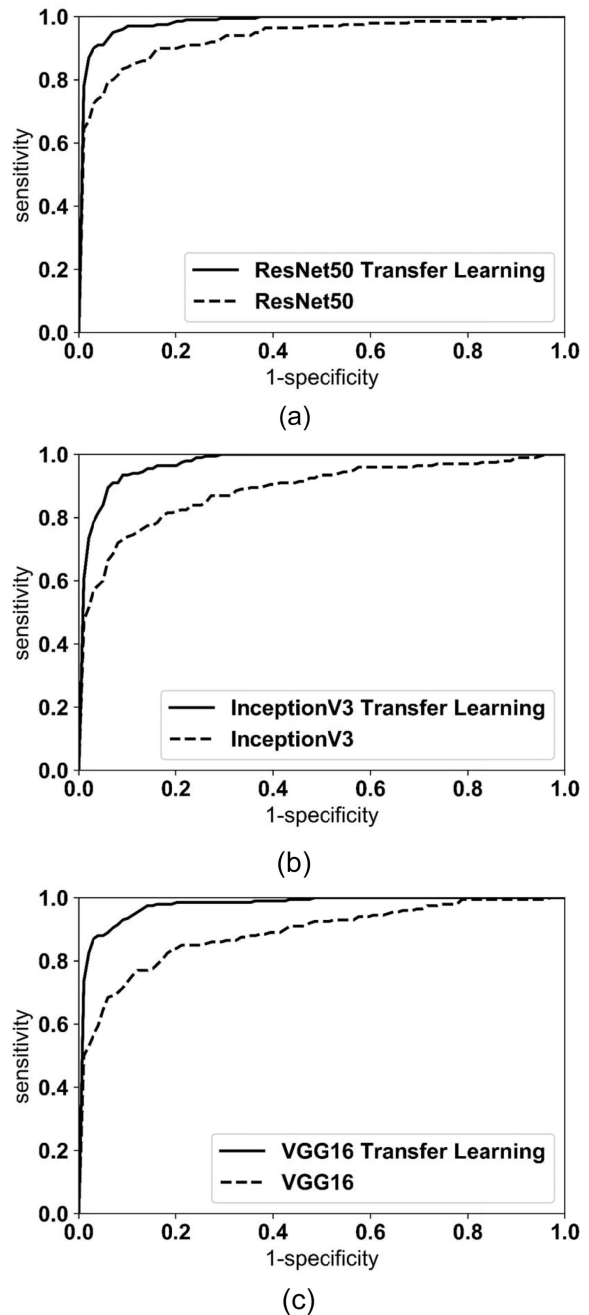


그림 4. 각 CNN 구조 별 전이학습 전, 후 ROC 그래프. (a) Resnet50 모델의 전이학습 전, 후 ROC 그래프, (b) InceptionV3 모델의 전이학습 전, 후 ROC 그래프, (c) VGG16 모델의 전이학습 전, 후 ROC 그래프. Fig. 4. ROC graphs before and after transfer learning for each CNN structure. (a) ROC graph before and after transfer learning for ResNet50, (b) ROC graph before and after transfer learning for InceptionV3, (c) ROC graph before and after transfer learning for VGG16.

III. 연구 결과

정상 이미지 40장과 비정상 이미지 118장, 총 158장의

표 1. 모델 별 분류 정확도(95% 신뢰구간) 및 AUC(95% 신뢰구간)
Table 1. Classification Accuracy (95% confidence interval) and AUC (95% confidence interval) by each model

Model	Accuracy	AUC
Resnet50 Transfer Learning	0.947 (0.919~0.974)	0.98 (0.97~0.99)
Inceptionv3 Transfer Learning	0.924 (0.908~0.94)	0.97 (0.96~0.98)
VGG16 Transfer Learning	0.938 (0.929~0.947)	0.98 (0.97~0.99)
Resnet50	0.9 (0.864~0.936)	0.94 (0.91~0.97)
Inceptionv3	0.862 (0.831~0.893)	0.89 (0.85~0.93)
VGG16	0.87 (0.843~0.897)	0.89 (0.87~0.91)

표 2. T검정을 통한 전이학습 전, 후 모델의 성능 비교
Table 2. Comparison of performance before and after Transfer learning by T-test

Model	T-test p value (Accuracy)	T-test p value (AUC)
Resnet50	0.031	0.021
Inceptionv3	0.009	0.006
VGG16	0.003	0.001

위내시경 이미지로 학습된 모델을 검증하였으며, 평가 방법으로 혼동 행렬(Confusion Matrix)에서 파생된 분류 정확도 (Accuracy)와 ROC(Receiver Operating Characteristic)커브의 아래 면적인 AUC(Area Under Curve)를 사용하였다. 학습 데이터와 검증 데이터를 서로 다르게 구성하여 5번 반복하였으며, 각각의 분류정확도와 AUC의 평균을 구하였다.

위내시경 영상의 정상 이미지와 비정상 이미지를 분류한 결과는 표 1과 같다. ResNet50 구조의 경우 전이학습 적용 전, 후 분류정확도가 0.047 향상된 0.947를 보였으며, AUC 또한 0.04 향상되어 0.98를 기록하였다. Inceptionv3 구조의 경우 전이학습 적용 전, 후 0.062 향상된 0.924의 분류정확도와 0.08 향상된 0.97의 AUC를 기록하였다. VGG16 구조의 경우 0.068 향상된 0.938의 분류정확도와 0.09 향상된 0.98의 AUC 기록하였다. 전이학습하기 전, 후 CNN 모델의 성능 차이를 T검정(T-test)으로 확인하면 표 2와 같다. 연구에 적용한 모든 CNN 모델에서 통계적으로 유의한 차이를 보였다($p < 0.05$). 비정상 이미지의 비정상 부위가 아닌 다른 불필요한 곳에서 이미지를 판단하는지 확인하기 위해 Class Activation Map(CAM)[24] 을 통해 합성곱 신경망의 마지막 레이어에서 높은 가중치를 두는 지점

을 확인 하였다. 그 결과 표 3에 내용처럼 용종 근처에 집중되는 것을 확인 할 수 있었다.

IV. 결 론

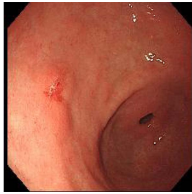
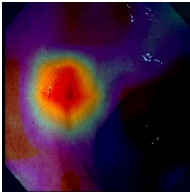

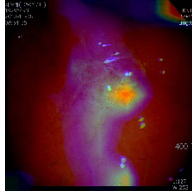
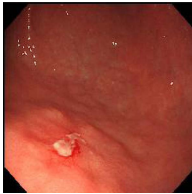
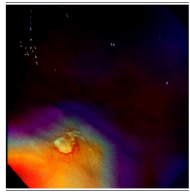

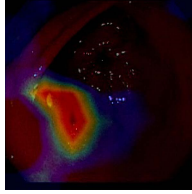
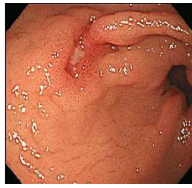
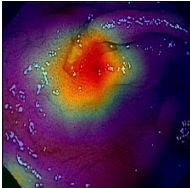

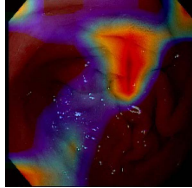

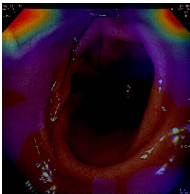

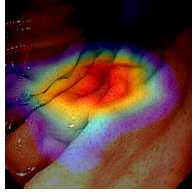
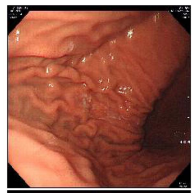
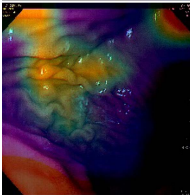
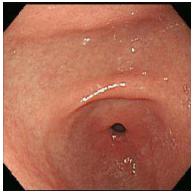
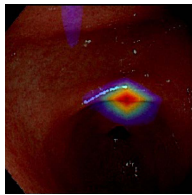
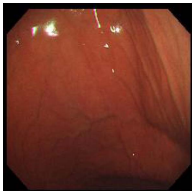
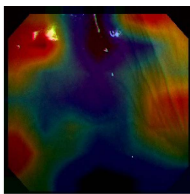

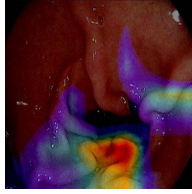
본 논문은 합성곱 신경망 구조를 활용한 위내시경 이미지 분류에서 전이학습의 효용성에 대하여 비교 및 평가하였다. 실생활 이미지에서 학습된 모델의 가중치를 위내시경 분류 모델의 초기값으로 전이하여 학습하였을 때와 임의의 값으로 학습하였을 때, 연구에 적용된 모든 합성곱 신경망 구조에서 전이학습을 하였을 경우에 더 높은 분류정확도와 ROC의 AUC를 보였다. 또한 그 차이는 통계적으로 유의미하였다($p < 0.05$). 다양한 모델에 적용하였을 때 분류정확도 및 AUC값이 가장 우수하였던 ResNet50 모델 기준으로 전이 학습을 하기 전, 후 평균 정확도는 0.047 상승하였으며, AUC값 또한 0.04 상승하였다. 보통의 의료 이미지와 실생활 이미지는 촬영 방법과 이루어진 비트(bit)가 다르지만, 내시경 이미지는 실생활 이미지와 유사한 방법으로 촬영되고 같은 비트를 갖는다. 그렇기 때문에 실생활 이미지에서 학습된 가중치가 전이되었을 때 더욱 효과적이었다고 생각된다.

기본적으로 획득한 데이터가 충분히 많다면, 합성곱 신경망 성능이 향상될 것이다. 하지만 전문의의 주석(Annotation)이 포함된 양질의 데이터를 모으기 어려운 의료 데이터 특성상 전이학습은 효과적인 방법으로 판단된다. 물론 이미지 회전, 확대, 시프트 등을 활용하여 데이터를 증강(Augmentation)하는 방법도 존재한다. 데이터 증강 기법을 통해서 부족한 학습 데이터를 늘리고, 최종적으로 딥러닝 모델의 성능이 개선되는 경우도 있다[25]. 이러한 기법은 전이학습과 트레이드 오프(trade-off) 관계가 아니기 때문에, 추후 성능 개선을 위해 추가적으로 적용될 수 있다고 생각된다.

본 연구에서 가장 분류 정확도 및 AUC가 높았던 전이학습이 적용된 ResNet50 모델을 Class Activation Map(CAM)으로 확인하면 표 3과 같다. CAM은 딥러닝 모델의 마지막 레이어인 전결합층(Fully connected layer)을 Global Average Pooling(GAP) 레이어로 교체 후, 딥러닝 모델이 어떤 부분에 집중하여 최종 결과를 산출 하였는지 시각화하는 기술이다. 표 3을 보면, 비정상 이미지의 경우 오분류된 이미지와 정분류된 이미지 모두 비정상 부위에 높은 비중을 두고 있다. 즉, 모델의 최종 확률 임계값에 변화를 주면 향후 비정상 이미지의 오분류율이 낮아질 가능성이 있다고 판단된다. 정상 이미지의 경우, 오분류된 이미지의 CAM을 확인하면 병변이 없는 부위에 높은 비중을 두고 있다. 특히 위장관벽에 불규칙한 주름들을 비정상 부분으로 판단하고 있다. 병변과 불규칙한 주름을 구별하기 위해서는 본 연

표 3. 전이학습 후 Resnet50의 적용된 원본 이미지와 CAM

Table 3. After the Transfer learning, Original image and Class Activation Map

		Resnet50-Transfer Learning Result			
		Correct Classification		Misclassification	
		Original Image	Class Activation Map	Original Image	Class Activation Map
Abnormal					
					
					
Normal					
					
					

구에서 활용한 이미지보다 더 다양하고 많은 정상 이미지가 필요하다. 비정상 이미지의 수만큼 정상 이미지의 수가 충족되어 균등한 데이터로 딥러닝 모델이 학습한다면, 보다 높은 분류정확도를 기록될 것이라고 생각된다.

참고문헌

- [1] <http://www.ncc.re.kr>, accessed on Apr. 24, 2018.
- [2] H. Katai *et al.*, "Five-year survival analysis of surgically resected gastric cancer cases in Japan: a retrospective analysis of more than 100,000 patients from the nationwide registry of the Japanese Gastric Cancer Association (2001-2007)," *Gastric Cancer*, vol. 21, no. 1, pp. 144-154, 2018.
- [3] H. A. Park *et al.*, "The Korean guideline for gastric cancer screening," *J. Korean Med. Assoc.*, vol. 58, no. 5, pp. 373-384, 2015.
- [4] T. Hirasawa *et al.*, "Application of artificial intelligence using a convolutional neural network for detecting gastric cancer in endoscopic images," *Gastric Cancer*, no. 0123456789, pp. 1-8, 2018.
- [5] S. Menon and N. Trudgill, "How commonly is upper gastrointestinal cancer missed at endoscopy? A meta-analysis," *Endosc. Int. Open*, vol. 02, no. 02, pp. E46-E50, 2014.
- [6] Y. Shimodate *et al.*, "Gastric superficial neoplasia : high miss rate but slow progression," no. December 2014, pp. 722-726, 2017.
- [7] K. Y. Hosokawa O, Hattori M, Douden K, Hayashi H, Ohta K, "Difference in accuracy between gastroscopy and colonoscopy for detection of cancer," *Hepatogastroenterology*, vol. 54, pp. 442-444, 2007.
- [8] M. Häfner, A. Gangl, M. Liedlgruber, A. Uhl, A. Vécsei, and F. Wrba, "Combining Gaussian Markov random fields with the discretewavelet transform for endoscopic image classification," *DSP 2009 16th Int. Conf. Digit. Signal Process. Proc.*, pp. 1-6, 2009.
- [9] P. Wang, S. M. Krishnan, C. Kugean, and M. P. Tjoa, "Classification of endoscopic images based on texture and neural network," *Annu. Reports Res. React. Institute, Kyoto Univ.*, vol. 4, pp. 3691-3695, 2001.
- [10] G. H. Yann LeCun, Yoshua Bengio, "Deep learning," *Nature*, vol. 521, pp. 436-444, 2015.
- [11] M. I. Razzak, S. Naz, and A. Zaib, "Deep Learning for Medical Image Processing: Overview, Challenges and Future," *CoRR*, vol. 1704.06825, pp. 1-30, 2017.
- [12] A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Adv. Neural Inf. Process. Syst.*, pp. 1-9, 2012.
- [13] K. He, "Delving Deep into Rectifiers : Surpassing Human-Level Performance on ImageNet Classification," 2014.
- [14] V. Gulshan *et al.*, "Development and validation of a deep learning algorithm for detection of diabetic retinopathy in retinal fundus photographs," *JAMA-J. Am. Med. Assoc.*, vol. 316, no. 22, pp. 2402-2410, 2016.
- [15] A. Esteva *et al.*, "Dermatologist-level classification of skin cancer with deep neural networks," *Nature*, vol. 542, no. 7639, pp. 115-118, 2017.
- [16] H. C. Shin *et al.*, "Deep Convolutional Neural Networks for Computer-Aided Detection: CNN Architectures, Dataset Characteristics and Transfer Learning," *IEEE Trans. Med. Imaging*, vol. 35, no. 5, pp. 1285-1298, 2016.
- [17] G. Wimmer, A. Vécsei, and A. Uhl, "CNN Transfer Learning for the Automated Diagnosis of Celiac Disease," 2016.
- [18] F. Zhang, X. Xu, and Y. Qiao, "Deep classification of vehicle makers and models: The effectiveness of pre-training and data enhancement," *2015 IEEE Int. Conf. Robot. Biomimetics, IEEE-ROBIO 2015*, pp. 231-236, 2015.
- [19] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," pp. 1-14, 2014.
- [20] C. Szegedy, V. Vanhoucke, S. Ioffe, J. Shlens, and Z. Wojna, "Rethinking the Inception Architecture for Computer Vision," 2015.
- [21] K. He, X. Zhang, S. Ren, and J. Sun, "Deep Residual Learning for Image Recognition," *2016 IEEE Conf. Comput. Vis. Pattern Recognit.*, pp. 770-778, 2016.
- [22] L. F.-F. Olga Russakovsky, Jia Deng, Hao Su, Jonathan Krause, Sanjeev Satheesh, Sean Ma, Zhiheng Huang, Andrej Karpathy, Aditya Khosla, Michael Bernstein, Alexander C. Berg, "ImageNet Large Scale Visual Recognition Challenge," *Int. J. Comput. Vis.*, vol. 115, pp. 211-252, 2015.
- [23] A. Y. Ng, "Preventing 'Overfitting' of Cross-Validation data," *CEUR Workshop Proc.*, vol. 1542, pp. 33-36, 2015.
- [24] R. R. Selvaraju, A. Das, R. Vedantam, M. Cogswell, D. Parikh, and D. Batra, "Grad-CAM: Why did you say that?," pp. 1-4, 2016.
- [25] M. Lin, Q. Chen, and S. Yan, "Network In Network," pp. 1-10, 2013.