캡슐내시경의 위치추적을 위한 CNN 기반 위장관 랜드마크 분류기 설계1)

장현웅*, 임창남*, 박예슬*, 이광재**, 이정원*
*아주대학교 전자공학과, **아주대학교 의과대학 소화기내과
e-mail:jhw3169@gmail.com, chn0714@naver.com, yeseuly777@gmail.com, kjl@ajou.ac.kr, jungwony@ajou.ac.kr

Design of CNN-based Gastrointestinal Landmark Classifier for Tracking the Gastrointestinal Location

Hyeon-Woong Jang*, Chang-Nam Lim*, Ye-Seul Park*, Kwang-Jae Lee**, and Jung-Won Lee*
Department of Electrical and Computer Engineering, Ajou University*
Department of Gastroenterology, Ajou University Hospital**

요 약

최근의 영상 처리 분야는 딥러닝 기법들의 성능이 입증됨에 따라 다양한 분야에서 이와 같은 기법들을 활용해 영상에 대한 분류, 분석, 검출 등을 수행하려는 시도가 활발하다. 그중에서도 의료 진단 보조 역할을 할 수 있는 의료 영상 분석 소프트웨어에 대한 기대가 증가하고 있는데, 본 연구에서는 캡슐내시경 영상에 주목하였다. 캡슐내시경은 주로 소장 촬영을 목표로 하며 식도부터 대장까지 약 8~10시간 동안 촬영된다. 이로 인해 CT, MR, X-ray와 같은 다른 의료 영상과 다르게 하나의 데이터 셋이 10~15만 장의 이미지를 갖는다. 일반적으로 캡슐내시경 영상을 판독하는 순서는 위장관 교차점(Z-Line, 유문판, 회맹판)을 기준으로 위장관 랜드마크(식도, 위, 소장, 대장)를 구분한 뒤, 각 랜드마크 별로 병변 정보를 찾아내는 방식이다. 그러나 워낙 방대한 영상 데이터를 가지기 때문에 의사 혹은 의료 전문가가 영상을 판독하는데 많은 시간과 노력이 소모되고 있다. 본 논문의 목적은 캡슐내시경 영상의 판독에서 모든 환자에 대해 공통으로 수행되고, 판독하는 데 많은 시간을 차지하는 위장관 랜드마크를 찾는 것에 있다. 이를 위해, 위장관 랜드마크를 식별할 수 있는 CNN 학습 모델을 설계하였으며, 더욱 효과적인 학습을 위해 전처리 과정으로 학습에 방해가 되는 학습 노이즈 영상들을 제거하고 위장관 랜드마크 별 특징 분석을 진행하였다. 총 8명의 환자 데이터를 가지고 학습된 모델에 대해 평가 및 검증을 진행하였는데, 무작위로 환자 데이터를 샘플링하여 학습한 모델을 평가한 결과, 평균 정확도가 95% 가 확인되었으며 개별 환자별로 교차 검증 방식을 진행한 결과 평균 정확도 67% 가 확인되었다.

1. 서론

최근의 영상 처리 분야에서는 딥러닝과 머신러닝 기법이 우수한 성능을 보임에 따라, 영상에 대한 분류, 분석, 검출 등의 목적을 가지고 다양한 분야에서 활용되고 있다. 그중에서도 최근 주목받고 있는 분야로는 의료 분야가 있다. 의료 영상들을 이용하여 진단을 보조할 수 있는 정보를 생성하기 위한 연구들이 활발해지고 있는 추세이며[1], 최근에는 Vuno의 BoneAge[2], Lunit의 Insight[3] 등과 같은 기계학습 기반의 의료 진단 보조 소프트웨어도 많이 개발되고 있는 추세이다. 컴퓨터단(층)촬영영상(CT: Computed Tomography), 자기공명영상(MR: Magnetic Resonance), X-ray 등 다양한 목적의 의료 영상이 존재하는데, 본 연구에서는 그 중 캡슐내시경 영상에 주목하였다.

캡슐내시경은 식도에서 항문까지 긴 위장관을 촬영하기 위한 장비로, 특별한 마취 없이 행해질 수 있고 다른 내시경과는 다르게 이물감이나 거부감이 덜하기 때문에 점차 사용량이 늘어나고 있다. 캡슐내시경의 가장 큰 장점은 대장 내시경으로는 판독이 어려운 소장을 촬영할 수 있다는 것이다. 보통, 캡슐내시경은 약 8시간에서 10시간의 측정 시간을 가지고 1초에 약 2장 이상의 사진을 찍는다. 이 과정에서 방대한 양의 데이터가 쌓이게 되며 이를 판독하는 의사는 보통 1시간에서 2시간의 판독시간을 필요로 하는데 그 이유는 다양하다.

우선, 의사 혹은 판독관은 각각의 프레임이 위장관 내에 어느 위치에 존재하는지 확인하기 위해, 위장관 교차점(Z-Line, 유문판, 회맹판)을 기반으로 위장관 랜드마크(식도, 위, 장(십이지장, 공장, 회장), 결장)를 관찰한다. 이때, 영상에서 발견되는 위장관 교차점은 환자마다 다르다. 이로 인해, 교차점을 발견하는데 많은 시간이 소모된다. 또한, 병변을 진단하기 위한 판독 시간 역시 많이 소요된다. 많은 연구에서 이처럼 병변을 판독하기 위한 시간을 줄이기 위해서 다양한 연구를 진행해왔다. [4]의 연구에서는 HSV 색공간을 활용해 폴립과 종양을 자동으로 감지하

1) "본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터지원사업의 연구결과로 수행되었음" (IITP-2019-2016-0-00309)

는 연구를 진행하였다. 이와 같은 연구는 판독자로 하여금 진단을 보조함으로써 판독하면서 발생하는 피로감을 줄일 수 있다.

하지만, 아직까지 위장관 교차점 혹은 랜드마크를 자동으로 구별해 줄 수 있는 연구는 미비한 실정이다. 이는 과거 캡슐내시경이 많이 이루어지지 않았고 또한 수집 자체도 어려운 문제점이 있었다. 또한, 영상 데이터 셋을 위장관 교차점의 데이터는 수 만장의 프레임 중 굉장히 작은 비율을 차지하기 때문에, 학습 데이터의 불균형 문제로 인해 학습이 어렵다는 문제점도 있었다. 따라서 본 연구에서는 아주대병원과 함께 연구를 진행하여, 의료 전문가에 의해 레이블링된 위장관 교차점을 기반으로 위장관 랜드마크를 학습하기 위한 네트워크 모델(분류기)을 제안하고자 한다.

위장관 랜드마크와 교차점을 구별해 내고자 하는 연구 목표는 다음 두 가지 중요한 응용에 사용될 수 있다. 첫째, 위장관 랜드마크를 구별하는 판독시간을 줄일 수 있다. 둘째, 위장관 랜드마크를 구별하면 자연스럽게 교차점을 인식할 수 있게 된다. 따라서 병원 현장에서 캡슐내시경 시술에 가장 어려움을 겪고 있는 정체(retention) 현상을 해결할 수 있다. 정체 현상이란 캡슐이 특정 위장관에 정체되어 모든 구간을 통과하지 못하는 현상을 의미한다. 이를 해결하기 위해서는 정체되어 있다고 판단될 때 환자의 자세를 바꿔주거나 걷는 등 조금의 움직임이 필요한데, 현재는 캡슐의 위치 추적이 이루어지지 않고 있어 이와 같은 정체 현상을 해결하기 어려운 상황이다. 본 논문에서는 이러한 두 가지 목표를 가지고 현재 캡슐내시경의 랜드마크 위치를 파악할 수 있는 CNN 학습 모델을 설계하였고, 샘플 환자 8명에 대한 성능을 분석하였다. 무작위로 환자 데이터를 샘플링 하여 학습한 모델을 평가한 결과, 평균 정확도가 95% 가 확인되었으며 개별 환자별로 교차검증 방식을 진행한 결과 평균 정확도 67% 가 확인되었다. 개별 환자에 대한 성능을 개선시키기 위해, 추후 더 많은 환자에 대한 데이터를 반영하여 연구를 확장할 계획이다.

2. 위장관 랜드마크 별 특징 분석

2.1 캡슐내시경 이동 경로 및 특징

캡슐내시경은 음식물의 이동경로인 위장관을 따라 이동하며 촬영한다. 위장관은 위장관 교차점을을 기준으로 위치를 나누며 해당 지점을 위장관 랜드마크라 한다. 식도의 경우에는 워낙 빠른 시간 만에 통과를 하고 데이터의 양이 적기 때문에 크게 위, 소장, 대장 3가지로 분류하였다. 캡슐내시경은 촬영시간의 대부분을 소장에서 보내며, 대장의 일부분을 촬영하다 배터리가 소진되면 촬영은 끝난다. 만약 너무 오랜 시간동안 위에 머물게 된다면 환자의 자세를 바꿔주거나 다른 조치를 통해서 이를 해결해야 한다.

그림 1은 각각 위, 소장, 대장의 영상이다. 각 영상을 살펴보면 위의 경우에는 다른 두 위장관에 비해 큰 주름을 가지고 있고 탄력적인 것처럼 보인다. 소장의 경우에는 영양소를 흡수하기 위해서 융털을 가지고 있다. 이러한 특징으로 인하여 주름의 겉의 부분은 오돌토돌한 특징을 보인다. 대장의 경우에는 주름이 탄력이 없어 보이며, 소장처럼 융털을 가지고 있지 않고 대부분의 영상에 이물질이 껴있는 경우가 많다.

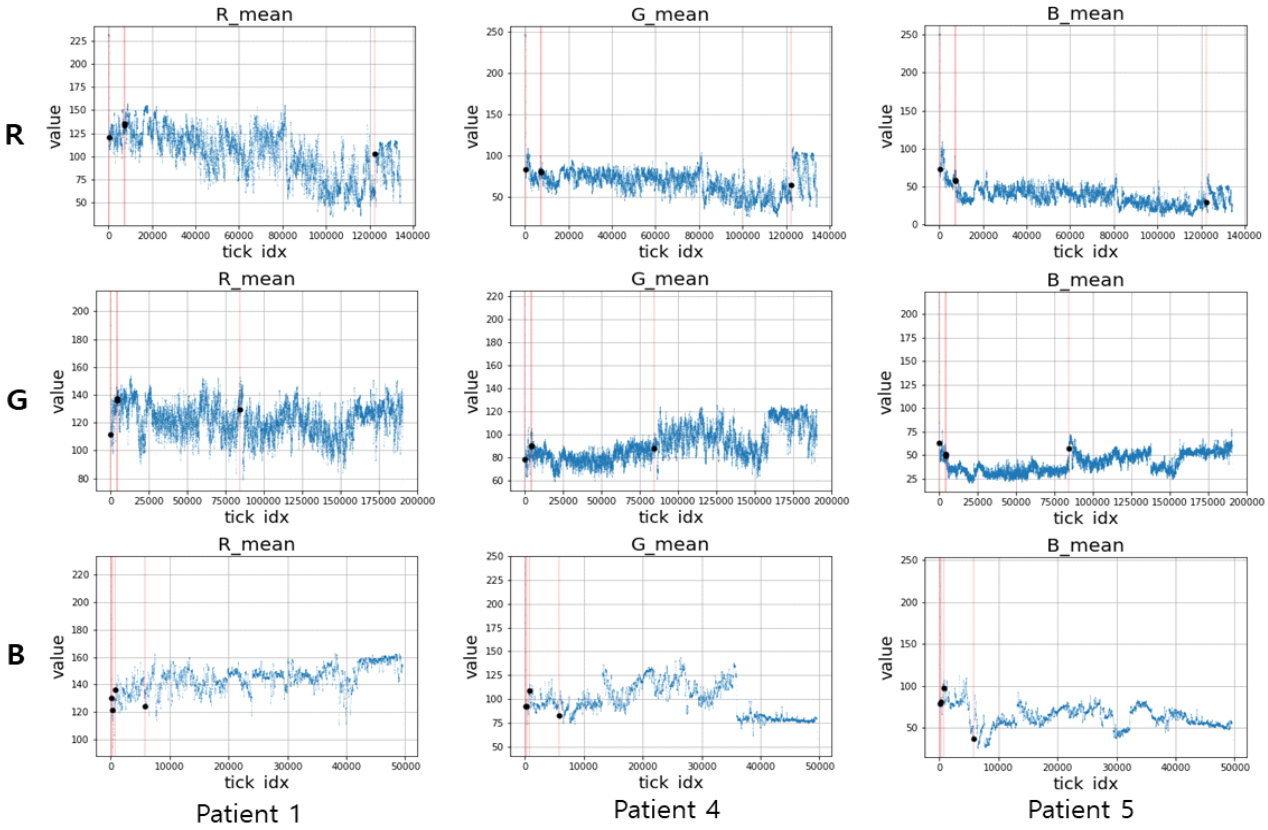
(그림 1) 대표적인 위장관 랜드마크 영상(위, 소장, 대장)

2.2 캡슐내시경 학습 노이즈 영상

캡슐내시경은 음식물이 소화되는 경로를 통과하며 촬영하기 때문에, 위장관 랜드마크 별 학습에 방해가 되는 다양한 노이즈 영상이 존재한다. 예를 들어, 대장의 경우 노란 액체로 가득 찬 영상도 있으며, 배설물만 보이는 영상 등이 존재한다. 또한, 위나 소장의 경우, 장액에 의해 거품으로 가득 낀 영상 등 많은 영상이 존재 하는데, 이는 위장관의 특성을 파악하기 어렵도록 만들어 학습에 방해가 된다. 따라서 이와 같은 데이터를 미리 걸러낸다면 판독시간을 줄일 수 있을 뿐만 아니라 의미 없는 영상을 학습하지 않기 때문에 학습에 있어서도 좋은 결과를 낼 수 있다. [5]의 연구에서는 유사도와 엔트로피, 명암도를 이용하여 장시간의 판독을 요구하는 캡슐내시경 영상 중 중복된 영상을 제거하였다. [6]의 연구도 마찬가지로 판독자의 판독 시간을 줄이기 위하여 노이즈영상과 장기 사진을 분류한 후에 은닉마크로프 모델을 이용해서 영상을 분류하여 판독에 있어 쉽게 수행할 수 있게 했다. [7, 8]의 연구에서는 SVM과 K-평균 군집화를 이용하여 판독에 있어서 의미없는 노이즈 영상을 제거하여 판독시간을 단축시키려 한다. 본 연구에서는 3장에서 학습 노이즈 영상이 학습에 어느 정도의 영향을 끼치는지 비교하기 위해, 노이즈가 있는 데이터 셋과 없는 데이터 셋을 이용하여 학습을 수행하였으며 이에 대한 결과를 비교하였다.

2.3 위장관 랜드마크 별 색공간 특징 분석

본 절에서는 위장관 랜드마크 및 환자 별로 색공간 특징 분석 결과를 소개한다. 총 8명의 환자 데이터를 가지고 수행하였으며, 표 1과 같이 각 환자의 데이터를 위, 소장, 대장 별로 분류한 후에 해당 영역의 이미지의 평균을 R, G, B 별로 구했다. 같은 위장과 랜드마크지만 환자 별로 R, G, B 채널의 평균값을 비교해보면, 위장의 경우에 R의 값은 비슷하지만 G, B의 값은 많은 차이가 났다. 마찬가지로 소장과 대장에서 결과를 살펴보면, 환자 별로 차이가 나는 것을 확인할 수 있다. 그림 2는 개별 환자 영상 데이



(그림 2) 임의의 3명 환자에 대한 프레임 별 R, G, B 통계 정보 비교

터 셋을 프레임 별로 R, G, B 통계 정보를 살펴본 것이다. 빨간 선이 순차적으로 위에서 소장, 소장에서 대장으로 바뀌는 시점을 나타낸다. 임의의 3명 환자에 대해 살펴보았으며, B 채널에 대해서 위장관 랜드마크 별로 차이가 많은 것을 확인할 수 있었다.

<표 1> 랜드마크 별 개별 환자의 R, G, B 통계 정보

label	patient No	R_mean	G_mean	B_mean
stomach	1	123.78	79.08	66.53
	2	119.52	80.90	68.39
	3	127.71	90.45	67.77
	4	127.71	83.01	50.94
	5	129.46	94.58	81.53
	6	109.21	75.24	51.49
	7	114.21	77.34	60.30
	8	127.20	82.34	51.10
intestine	1	123.03	84.25	43.33
	2	118.91	86.23	37.14
	3	115.75	76.94	31.05
	4	124.21	80.64	32.99
	5	135.75	97.20	78.54
	6	127.41	90.09	41.76
	7	130.23	91.22	45.36
	8	129.95	85.19	43.69
colon	1	94.93	80.94	37.91
	2	116.31	90.52	43.53
	3	118.35	81.02	57.44
	4	120.08	100.00	47.95
	5	146.13	99.11	61.82
	6	127.89	103.32	48.33
	7	118.93	91.99	50.24
	8	89.32	72.74	56.51

3. 위장관 랜드마크 분류기 설계 및 성능 비교

3.1 실험 환경

실험에 적용한 데이터는 총 8명의 데이터로 노이즈를 포함하면 위장사진 20,175장, 소장사진 114,046장, 대장사진 48,884장의 영상을 이용하였다. 실험은 크게 2가지로 이루어졌는데 첫 번째는 환자의 데이터를 무작위로 샘플링 하여 학습과 검증을 수행하였고, 두 번째는 환자 별로 데이터를 구분하여 7명 환자의 데이터를 학습하고 나머지 1명에 대한 데이터를 이용하여 검증하는 방식의 교차 검증을 진행하였다. 교차검증의 경우 1명의 데이터를 정확도 측정용으로 사용했다. 또한, 학습 데이터의 비율은 가장 적은 레이블의 데이터를 기준으로 위, 소장, 대장 1:1:1의 비율로 진행하였다. CNN 모델의 세부사항은 표 2와 같다.

3.2 모델 검증 결과

먼저, 무작위로 환자 데이터를 샘플링 하여 학습하고 평가 하였을 때의 결과는 표 3과 같다. 학습 노이즈의 유무에 따라 학습 성능이 변화하는데, 학습 노이즈가 존재하였을 때 약 93%였으며 학습 노이즈를 제거하였을 때는 2% 정도 성능이 개선된 95%였다. 다음은 환자 별로 구분하여 학습을 진행하였을 때의 결과이다. 결과는 표 4, 표5와 같으며, 1, 3, 5 환자의 경우에는 정확도가 매우 떨어지고, 나머지 환자의 경우에는 준수한 정확도가 나왔다. 이 결과를 보아 환자마다 데이터의 차이가 심하다는 것을 알

<표 2> 실험에 사용한 CNN 모델의 세부 사항

Layer	Layer comment	output
Input	Input Image	128 x 128 x 32
Conv1	5 x 5 x 32 (SAME)	128 x 128 x 32
Pool1	2 x 2 max pooling	64 x 64 x 32
Conv2	5 x 5 x 64 (SAME)	64 x 64 x 64
Pool2	2 x 2 max pooling	32 x 32 x 64
Conv3	3 x 3 x 128 (SAME)	32 x 32 x 128
Conv4	3 x 3 x 128 (SAME)	32 x 32 x 128
Conv5	3 x 3 x 128 (SAME)	32 x 32 x 128
Flatten	flatten conv5 output	131072
FC_input	131072 Input node	131072
FC_hidden	256 Hidden node	256
FC_output	3 Output node	3

<표 3> 랜덤 샘플링 후의 학습 결과

	with noise	without noise
accuracy	0.93	0.95

수 있다. 또한, 해당 환자의 특징은 다른 환자에게서 나타나지 않았다는 의미로 받아들여진다. 학습 노이즈를 제거한 후에는 1, 2, 4번의 환자의 경우에는 정확도가 증가하였고, 7번 환자의 경우에서만 정확도가 하락하였다. 5번 환자의 경우는 정확도가 매우 떨어지는데 표 1에서 값을 보면 다른 데이터들과 매우 큰 차이가 나는 것을 볼 수 있다. 따라서 이와 같이 큰 정확도 차이를 개선하기 위해서는 환자 데이터에 대한 일반화를 위해 보다 많은 데이터 수집이 필요하다는 것을 확인할 수 있다. 즉, 각 환자와 비슷한 데이터의 환자가 존재해야 하는데 이를 해결하기 위해서는 많은 환자, 증상이 비슷한 환자의 데이터를 수집해서 학습을 시킨다면 차이를 개선시킬 수 있다.

<표 4> 학습 노이즈 영상이 있을 때의 학습 성능 분석

Patient No	Stomach	Intestine	Colon	Total Accuracy
1	0.15	0.47	0.36	0.36
2	0.80	0.75	0.65	0.72
3	0.56	0.77	0.31	0.62
4	0.65	0.93	0.88	0.90
5	0.00	0.10	0.46	0.30
6	0.47	0.97	0.75	0.91
7	0.82	0.94	0.24	0.90
8	0.80	0.52	0.55	0.68
* Measure of Accuracy : F1-score				

<표 5> 학습 노이즈 영상이 없을 때의 학습 성능 분석

Patient No	Stomach	Intestine	Colon	Total Accuracy
1	0.22	0.64	0.42	0.53
2	0.52	0.86	0.71	0.79
3	0.42	0.85	0.19	0.62
4	0.57	0.93	0.94	0.93
5	0.09	0.59	0.32	0.31
6	0.66	0.97	0.76	0.92
7	0.59	0.93	0.34	0.88
8	0.66	0.53	0.83	0.68
* Measure of Accuracy : F1-score				

4. 결론 및 향후 연구

본 논문에서는 캡슐내시경의 위치추적을 위한 위장관 랜드마크 분류기 설계하고, CNN을 이용한 방법을 제안하였다. 모든 환자에 대해 무작위로 선별한 데이터에 대한 학습 정확도는 약 95% 정도로 매우 높았지만, 환자별로 구분하였을 때는 정확도의 차이가 매우 컸다. 이에 대한 원인은 환자마다 데이터의 분포가 상이한 것으로 파악된다. 따라서 향후 연구에서는 아주대병원과의 협업을 통해 더 많은 환자 데이터에 대해 기존 모델을 적용 및 확장해 볼 계획이며, 이를 기반으로 위장관 내 위치 추적 및 실시간 진단 지원을 위한 연구를 진행할 계획이다.

참고문헌

[1] Litjens, Geert, et al. "A survey on deep learning in medical image analysis." *Medical image analysis,* Vol. 42, pp. 60-88, 2017

[2] Kim Jeong Rye, et al. "Computerized bone age estimation using deep learning based program: evaluation of the accuracy and efficiency." *American Journal of Roentgenology,* Vol. 209, No. 6, pp. 1374-1380, 2017

[3] Eui Jin Hwang, et al. “Development and Validation of a Deep Learning-Based Automatic Detection Algorithm for Active Pulmonary Tuberculosis on Chest Radiograpgs”, *Clinical Infectious Diseases*, doi: 10.1093/cid/ciy967, 2018.11

[4] .Karargyris, Alexandros, and Nikolaos Bourbakis. "Detection of small bowel polyps and ulcers in wireless capsule endoscopy videos", *IEEE transactions on BioMedical Engineering* Vol 58, No. 10, pp. 2777-2786, 2011

[5] 이현규, 최민국, 이돈행, 이상철, “영상 프레임 분석을 통한 대용량 캡슐내시경 영상의 지능형 판독보조 시스템”, *지능정보연구,* Vol 15, No. 2 pp. 33-48, 2009

[6] Chen, H., Wu, X., Tao, G., & Peng, Q. “Automatic content understanding with cascaded spatial-temporal deep framework for capsule endoscopy videos”, *Neurocomputing*, 229, pp. 77-87, 2017

[7] 유동연, 박예슬, 이정원, “캡슐내시경 영상의 학습 노이즈 제거를 위한 SVM 기반 과잔여물 이미지 분류 기법”, *한국소프트웨어종합학술대회*, pp 1865-1867, 2018

[8] 유동연, 박예슬, 이병정, 이정원, “의료 영상 데이터 중심 소프트웨어 개발을 위한 학습 노이즈 영상 분류 기법”, *한국소프트웨어공학학술대회*, pp 317-322, 2019