

인공지능 기반 위 병변 검출 알고리즘 개발

Development of Gastric Lesion Detection Algorithm based on AI

김재승¹, 박동균^{2*}

Jae-Seoung Kim¹, Dong Kyun Park^{2*}

요약

위암은 1999년 이후 우리나라에서 가장 많이 발생하는 암으로 1위를 차지하고 있다. 위암은 내시경 검사를 통해 일차적으로 판단되고 조직검사를 통해 정확히 진단되기 전까지는 특징적인 증상이 없으며, 실제로 위 내시경 검사를 받은 환자는 받지 않은 환자에 비해 생존율이 2.24배 높다는 연구 결과가 발표된 바 있다. 이러한 문제를 해결하기 위하여 본 논문은 위암 환자의 위 내시경 시행 시 임상 의에게 실시간으로 보조적인 정보를 제공해 주고자 제안되었다. 본 논문에서는 Faster R-CNN을 위 병변 검출에 적합한 모델로 개선하여 보다 빠르고 정확한 검출 결과를 임상 의에게 제공하는 방법을 제안한다. 기존 알고리즘과의 비교 평가 결과 평균 91%의 정확도를 도출하여 제안 방법이 보다 효과적임을 증명하였으며, 영상 처리 속도 또한 0.1sec/frame을 도출하여 실시간 처리에 적합함을 증명하였다. 향후 연구로 다양한 환경에서의 내시경 영상 수집을 통한 학습 데이터의 개선을 통해 정확도를 개선하고자 한다.

핵심어 : 위 내시경, 합성곱 신경망, 영역기반 합성곱 신경망 모델

Abstract

Gastric cancer is the most common cancer and has been the number one incidence since 1999 in Korea. Gastric cancer is primarily judged through endoscopy and has no characteristic symptoms until accurately diagnosed through biopsy then In fact, a study found that patients who underwent gastroscopy had a 2.24 times higher survival rate than those who did not. to solve these problems, This paper was proposed to provide real-time ancillary information to clinicians when performing gastroscopy for gastric cancer patients. In this paper, we propose a method to provide faster and more accurate detection results to clinicians by improving Faster R-CNN as a model suitable for gastric lesion detection. As a result of comparative evaluation with existing algorithms, an average accuracy of 91% was derived, proving that the proposed method is more effective. and, The image processing speed was also proven to be suitable for real-time processing by deriving 0.1sec/frame. As a future study, we intend to improve the accuracy by improving the learning data through the collection of endoscopic images in various environments.

Keyword : Gastric Endoscopy, CNN, R-CNN Model

1 Health IT Research center, Gachon University Gil Medical Center, Incheon, Korea [Researcher]

e-mail: mpmkjs@gilhospital.com

2 Health IT Research center, Gachon University Gil Medical Center, Incheon, Korea [Professor]

e-mail: pdk66@gilhospital.com (Corresponding author)

* This work was supported by National IT Industry Promotion Agency(NIPA) grant funded by the Korea government(MSIT) (No.S0252-21-1001, Development of AI Precision Medical Solution(Doctor Answer 2.0))

Received(November 8, 2021), Review Result(1st: November 26, 2021), Accepted(December 10, 2021), Published(December 31, 2021)



© 2021 The Authors. Published by NCISS.
This is an open access article licensed under the Creative Commons Attribution-NonCommercial 4.0 International License.
To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc/4.0/>.

1. 서론

위장 질환은 우리가 흔하게 가질 수 있는 질환이며, 보건복지부에서 발표한 자료에 따르면 우리나라의 암 발생자 수 1위를 차지하였다 [1]. 위암의 발생 원인을 파악하고자 위유암종, 만성위염, 위출혈, 위궤양 등의 위 병변이 전암성 병변으로 검토되고 있으며 [2], 위암은 발병 후 내시경 검사를 통해 육안으로 일차적으로 판단되고 조직검사를 통해 정확히 진단되기 전까지는 특징적인 증상이 없으므로 이상 병변을 조기에 발견함으로써 위암의 발생을 예방하는 것이 가장 최선의 방법이라고 할 수 있다. 실제로 위 내시경 검사를 받은 환자는 받지 않은 환자에 비해 생존율이 2.24배 높다는 연구 결과가 발표된 바 있다 [3].

그러나 위암으로 발전할 수 있는 조기 위암 및 위병변의 경우 발견이 어려운 경우가 많고, 병변의 형태만으로 위염이나 위궤양과의 구분이 쉽지 않아 진단을 놓치거나 오진하는 사례가 발생한다. 또한 종양의 침윤 깊이에 따라 치료 방법이 다를 수 있어 조기위암의 판단 및 위 병변에 대한 분류의 내시경적 병기 결정이 무엇보다 중요하다.

이러한 문제를 해결하기 위하여 의료분야에 IT기술을 접목하는 연구들이 활발히 이루어지고 있으며, 특히 인공지능 기술을 활용한 의료영상 진단, 검출 분야에 높은 성과를 보이고 있다. 현재 인공지능 기술을 의료분야에 적용한 기술들은 실제 의료현장에 적용됨으로써 기존 의료프로세스의 시간 및 비용의 절감 뿐 아니라, 의료진의 컨디션이나 숙련도에 따른 차이를 줄여주는 보조적인 수단으로 그 효과성이 검증되고 있다.

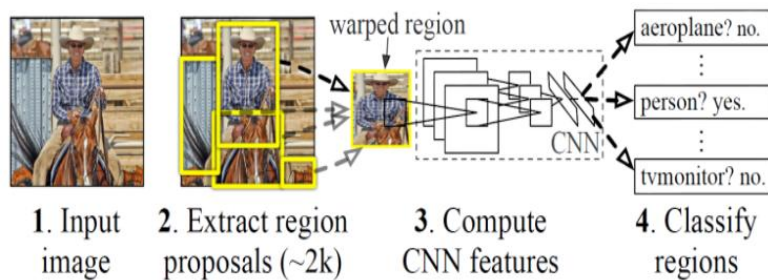
본 논문은 주요 암 질환 중 우리나라에서 가장 발병도가 높고 국가적으로 치료 및 예방에 대한 중요성이 높은 위암에 대한 조기 예방에 도움이 될 수 있는 인공지능 기반의 위 병변에 대한 검출 알고리즘을 제안하고자 한다. 실제 임상이가 위 내시경 검사 시에 위암으로 발전할 수 있는 위 병변에 대한 검출을 자동으로 수행하여 알려줌으로써 진단에 보조적인 수단으로 활용할 수 있다. 위 병변에 대한 내시경 영상 및 임상 소견 정보를 Faster R-CNN(Region Convolutional Neural Network) [4]를 통해 학습하여 실시간 환경에서도 적용 가능한 알고리즘을 제안한다.

본 논문의 구성은 다음과 같다. 2장에서는 R-CNN 기반 알고리즘에 대해 살펴보고 실제 내시경 진단 환경에 적용 시킬 수 있는 알고리즘에 대해 분석한다. 3장에서는 본 논문에서 제안하는 방법에 대해 기술한다. 4장에서는 제안한 알고리즘의 성능을 평가하고 결과에 대한 고찰을 한다. 마지막으로 5장에서는 본 논문의 내용과 결과를 요약하여 결론지으며, 이를 발전시킬 향후 연구과제에 대해 설명한다.

2. 관련 연구

2.1 R-CNN기반 객체 검출 알고리즘

R-CNN은 Girshick 등 [5]에 의해서 2014년 발표되었으며 영상 분류를 수행하는 CNN과 영상 내에 객체가 존재할 영역을 제안해 주는 region proposal 알고리즘을 조합하여 복잡한 자연 영상에서도 높은 성능의 객체 인식을 할 수 있는 가능성을 마련하였다. [그림 1]은 R-CNN의 구조를 나타내며[5] 크게 3단계로 구성된다. 첫 번째 단계는 region proposal 단계로 입력 영상에 대해 약 2,000개의 후보 영역을 찾아 바운딩 박스(bounding box) 형태의 지역적 카테고리 영역을 반환하는 동작을 수행한다. region proposal 알고리즘은 기존 슬라이딩 윈도우(sliding window) 탐색 방식의 비효율성을 개선하기 위하여 검출하고자 하는 객체가 있을 법한 영역을 빠른 속도로 찾아내는 알고리즘으로 다양한 방법들이 제안되었고, 그 중에서 selective search [6]와 edge box [7]알고리즘이 보편적으로 좋은 성능과 속도를 보여주었으며 [8] R-CNN은 selective search를 사용하였다.



[그림 1] R-CNN의 구조

[Fig. 1] Structure of R-CNN

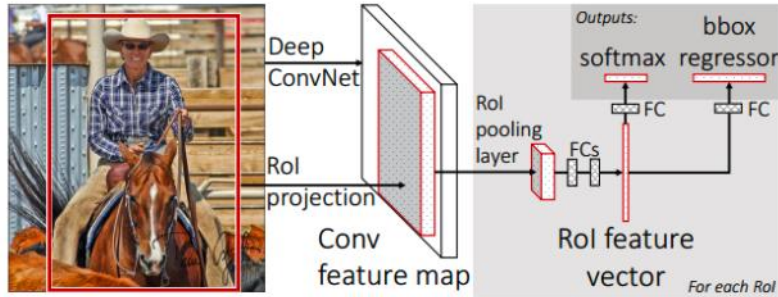
하지만 이러한 결과는 이미지가 특정 크기의 규격에 맞춰져 있고 영상 내에 하나의 객체만 존재하는 데이터에 대한 인식 성능을 평가한 것이기 때문에 많은 연구자들은 이에 더 나아가서 실제 복잡한 자연영상에서의 객체를 분류하고 인식하기 위한 방법들을 제안하였다.

이러한 노력의 결과로, R-CNN(Region-based CNN)의 발견과 R-CNN을 기반으로 한 알고리즘의 개선은 객체 인식 분야를 혁신적으로 발전시켰다. R-CNN을 기반으로 한 알고리즘은 Fast R-CNN, Faster R-CNN 순으로 개선되었으며 현재 실시간 객체 인식이 가능한 수준까지 이르렀다

2.2 Fast R-CNN

R-CNN은 2.2.1에서 기술한 3개의 단계 별로 독립적인 학습이 필요하기 때문에 학습이 어렵고,

입력 영상 1장 당 region proposal의 개수만큼 CNN의 연산이 필요하므로 연산량이 매우 많다는 단점이 있다. 이러한 문제를 해결하기 위하여 R-CNN의 제1저자는 2015년에 이를 개선한 Fast R-CNN [9]을 발표했다. [그림 2]는 R-CNN의 구조를 나타 낸다 [9].



[그림 2] Fast R-CNN 구조

[Fig. 2] Structure of Fast R-CNN

Fast R-CNN은 기존 R-CNN을 크게 2가지 부분에서 개선시켰다. 첫째, R-CNN의 1)CNN 영상 특징 추출, 2)SVM을 사용한 클래스 분류, 3)바운딩박스 보정의 3가지 모듈이 서로 분리되어 있는 구조를 [그림 2]와 같이 하나의 네트워크로 구성하여 학습 속도 및 정확도를 개선시켰다. softmax 분류기와 바운딩박스 보정에 대한 두 함수의 loss를 더한 multi-task loss 기반으로 동시에 두 가지 task를 학습하도록 하였고, R-CNN의 ROI(Region of Interest) 샘플링 방식 대신, mini-batch를 사용하여 학습에 필요한 CNN 연산량을 효율적으로 줄였다.

둘째, region proposal의 개수만큼 CNN연산을 하는 대신, 영상 전체에 한번의 CNN연산을 통하여 특징맵을 추출하고, 해당 특징 맵에서 region proposal의 계산된 값들을 공유하는 방법을 제안하여 연산량을 대폭 감소시켰다. R-CNN은 region proposal들을 고정된 크기로 warping하여 동일한 크기로 변환 시켰는데, 단순히 warping을 통한 크기의 변환은 정보의 손실이 발생하기 때문에 Fast R-CNN에서는 정보의 손실 없이 특징맵의 다양한 크기의 영역으로부터 일정한 크기의 특징을 추출하기 위하여 RoI 풀링(Region of Interest Pooling)기법을 적용하였다. Fast R-CNN에서 사용한 RoI 풀링 기법은 K. He 등 [10]이 제안한 SPP(Spatial Pyramid Pooling)기법으로 SPP를 CNN 특징맵에 적용할 수 있음을 보여줬다.

3. 인공지능 기반 위 병변 검출 알고리즘

본 논문에서는 위 내시경 시 의료 영상에서 병변을 검출하고 그 특성을 분석 함으로써 의학 전문의의 임상적 진단을 보조하는 수단을 제공하기 위하여 Faster R-CNN [4]기반의 실시간 위 병변

검출 알고리즘을 제안한다.

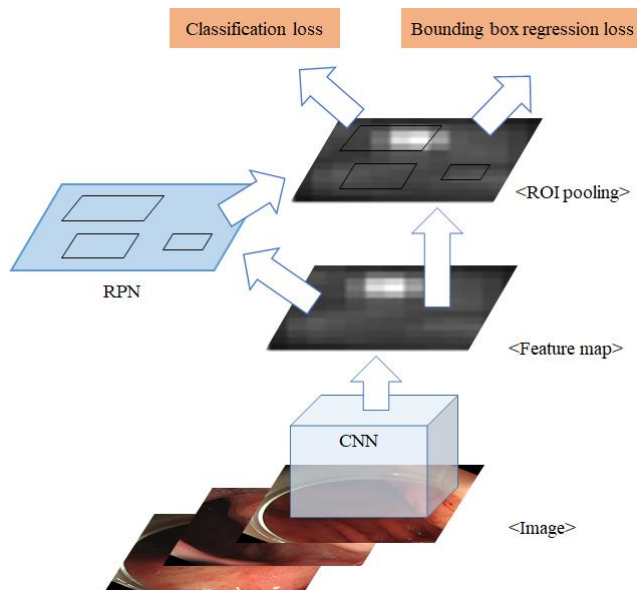
Faster R-CNN은 기존 R-CNN기반의 알고리즘보다 매우 빠른 연산 속도를 제공하지만, 실제 의료 현장에서의 적용을 위해 실시간으로 복잡한 영상을 분석하기 위해서는 검출하고자 하는 대상에 따른 하이퍼 파라미터의 수정 및 모델의 개선이 필요하다.

이러한 문제점을 개선하기 위하여 본 논문에서는 Faster R-CNN에서 가장 연산량이 많이 필요로 하는 RPN 구조를 개선하는 방법을 제안한다.

3.1 Faster R-CNN 기반의 위 병변 검출 기술

Fast R-CNN은 R-CNN의 속도 문제를 해결하고자 하였으며 R-CNN에 비해 10배 이상 빨라졌으나, CPU에서 연산되는 selective search 방식의 비효율적인 region proposal 방법은 Fast R-CNN의 전체 파이프라인 중 상당히 시간을 많이 차지하기 때문에 비효율적인 구조를 가진다. 이러한 문제를 해결하기 위하여 2015년 중반, Shaoqing Ren, Kaiming He, Ross Girshick, Jian Sun로 구성된 Microsoft Research 팀은 Faster R-CNN [4]을 제안하였다.

Faster R-CNN의 핵심은 CPU에서 처리되었던 CNN 외부에서의 selective search 기반의 region proposal을 CNN 내부에서 region proposal을 생성 할 수 있는 네트워크를 추가했다는 것이며 이를 RPN(Region Proposal Network)라고 한다.



[그림 3] Fast R-CNN 기반 위 병변 검출 구조

[Fig. 3] Faster R-CNN-based gastric lesion detection structure

Faster R-CNN의 RPN구조는 미리 정의된 여러 크기의 비율의 k개의 anchor box를 CNN의 특징맵에서 sliding window방식으로 선회하여 sliding window 위치마다 k개의 box를 출력하도록 하였다. RPN은 모든 anchor의 위치에 대해 객체와 배경을 구분하는 2k개의 classification 출력과 바운딩 박스의 크기와 위치를 보정하기 위한 4k개의 regression 출력을 가진다. 여기서 anchor를 사용하는 이유는 현실 세계에서의 객체들은 다양한 크기와 형태를 가지고 있기 때문에 이를 반영하기 위함이며, 논문에서는 3가지의 크기(128, 256, 512)와 3가지의 비율(1:1, 2:1, 1:2)을 사용해 총 k=9로 사용하였다. anchor의 개수와 크기, 비율은 하이퍼 파라미터(hyper-parameter)의 특성을 띄며 영상 내의 검출하고자 하는 객체의 형태에 따라 최적화 하여 사용된다. [그림 3]은 본 논문에서 제안하는 Faster R-CNN기반의 알고리즘 구조를 나타낸다.

3.2 Faster R-CNN의 RPN 개선

RPN은 입력 영상의 CNN을 통하여 추출한 feature map을 입력으로 하여 객체인지 배경인지에 대한 score를 가진 바운딩 박스들을 출력한다. region proposal을 생성하기 위해 feature map 위로 작은 네트워크를 슬라이드 하는데, 이때 window는 3x3의 크기로 설정한다. 각각의 sliding window의 위치에 동시에 여러 개의 region proposal을 예측하는데 이때 미리 정의해 놓은 크기와 비율을 가진 박스들을 사용하며 이를 anchor라고 정의한다.

[표 1] Faster R-CNN과 제안하는 anchor의 비교

[Table 1] Comparison between Faster R-CNN and the proposed anchor

	scale	aspect ratios
Faster R-CNN	128	2:1
	128	1:1
	128	1:2
	256	2:1
	256	1:1
	256	1:2
	512	2:1
	512	1:1
	512	1:2
Proposed	128	1:1.5
	128	2:1
	256	1:1.5
	256	2:1
	512	1:1.5
	512	2:1

Faster R-CNN에서는 3개의 크기(128,256,512), 3개의 비율(2:1, 1:1, 1:2)의 anchor를 사용한다. 이는 자연 영상에서의 다양한 객체의 형태를 포함하기 위함이다. 이러한 anchor의 개수가 많으면 많을수록 연산량이 증가하며, 본 논문에서는 내시경 진단 시 실시간 환경에 적합한 속도를 확보하기 위하여 위 병변 검출에 적합한 anchor를 재정의함으로써 불필요한 region proposal의 연산을 줄이고, 바운딩 박스 regression의 정확도를 향상시킨다. [표 1]은 기존 Faster R-CNN에서 사용한 anchor와 본 논문에서 제안하는 anchor를 비교한다.

Faster R-CNN의 경우 불필요한 region proposal의 개수가 많을수록 배경을 객체 영역으로 인식하게 되는 문제점이 발생하며, 연산량 또한 기하급수적으로 증가한다. 이러한 문제점을 해결하고자 본 논문에서는 위 병변에 대한 형태학적 분석을 통해 위 병변 검출 시 최적의 anchor를 적용함으로써 속도와 정확도를 개선시킨다.

RPN의 classification 레이어에서는 모든 anchor들에 대해 전달받은 anchor들과 ground-truth들을 비교하여 classifier를 학습시킨다. anchor에 대한 위 병변의 해당 유무는 ground-truth와 얼마나 오버랩 되는지에 대한 수치로 결정하며 이러한 수치를 판단하기 위하여 IoU(Intersection over Union)을 사용하며 다음 (식 1) 같이 정의한다.

$$IoU = \frac{A \cap GT}{A \cup GT} \quad (1)$$

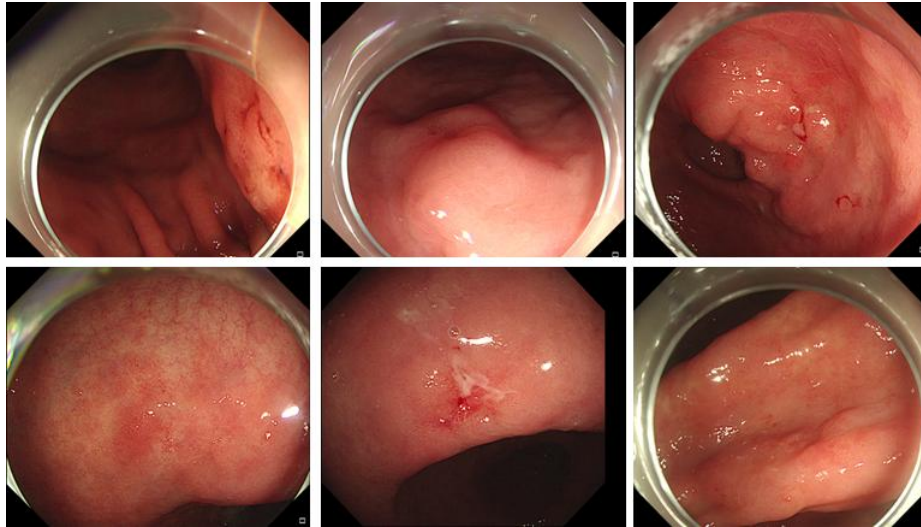
A 는 anchor의 좌표를 의미하며 GT 는 ground-truth box의 좌표를 의미한다. IoU를 사용한 anchor의 위 병변 영역 대한 판단은 다음 (식 2)와 같다.

$$p^* = \begin{cases} 1 & \text{if } IoU > 0.7 \\ -1 & \text{if } IoU < 0.3 \\ 0 & \text{if otherwise} \end{cases} \quad (2)$$

p^* 은 위 병변 해당 여부를 구별하는 anchor의 값을 의미하며 IoU가 0.7 이상일 때는 객체로 인식하고 0.3 이하일 때는 배경으로 간주한다.

4. 성능 평가

위 병변에 대한 인공지능 기술 개발 연구를 위하여 임상 시험 심사 위원회(IRB, Institutional Review Board)의 허가를 받은 가천대 길병원의 위 내시경 영상을 사용하였다. 모든 영상에 대한 위 병변값 및 레이블링 영역은 소화기 내과 전문의의 검증을 통해 수집하였다. 제안 방법의 성능평가를 위해 총 500명의 환자를 대상으로 수집한 내시경 영상에 대해 기존 알고리즘과의 속도 및 정확도를 비교 평가하였다. 본 연구에서 수집한 데이터 샘플은 [그림 4]와 같다.



[그림 4] 수집 데이터 샘플

[Fig. 4] Collect data samples

수집 데이터의 경우 병리 결과를 LGD, HGD, Cancer로 구분하고, 종양의 위치(upper, middle, lower)에 따라 구분하였으며, 추가적인 학습 파라미터로 흡연, 음주 유무, 재발여부, 기저질환(HTN, DM, Dyslipidemia, Heart disease (CAD, CHF), Brain disease (CVA, hemorrhage)에 대한 정보를 추가로 학습하였다.

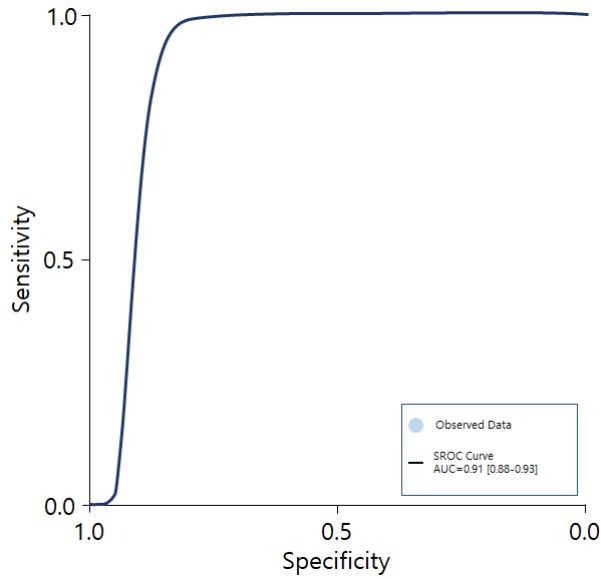
앞서 기술한 R-CNN, Fast R-CNN, Faster R-CNN과 본 논문에서 제안하는 알고리즘의 성능을 비교해 보면, [표 2]에서 보는 바와 같이 인식 성능도 증가하였지만, 처리 속도가 상당한 수준으로 개선된 결과를 확인할 수 있다. 이는 네트워크의 구조에 대한 개선도 있었지만 region proposal network의 개선이 가장 큰 역할을 했다는 의미로 볼 수 있다.

[표 2] R-CNN 기반 알고리즘 성능 비교

[Table 2] Comparison of R-CNN-based Algorithm Performance

	R-CNN	Fast R-CNN	Faster R-CNN	Proposed
영상 1개당 처리 속도	50sec	2sec	0.2sec	0.1sec
상대적 속도	1x	25x	250x	500x
mAP(VOC 2007)	66.0	66.9	69.9	70.1

본 논문에서 제안하는 알고리즘의 성능은 [그림 5]에 나타내었다. 수집된 데이터의 정상군과 위 병변을 가지고 있는 실험군으로 분류한 결과 AUC값은 0.91을 도출하였다.



[그림 5] ROC 커브

[Fig. 4] ROC curve

5. 결론

본 논문은 위암으로 발전할 수 있는 조기 위암 및 위 병변의 진단에 보조적인 수단으로 제공될 수 있는 인공지능 기술 기반 위 병변 검출 알고리즘을 제안하였다. 최근 R-CNN기반의 알고리즘을 활용한 영상분석 기술에 대한 연구가 좋은 성과를 보이고 있으며 특히, 의료 영상 분석 분야에 많은 시도가 이루어지고 있다. 이러한 인공지능 기술이 실제 의료현장에 도입되기 위해서는 높은 정확도 뿐 아니라, 실시간으로 신속히 임상이가 놓칠 수 있는 부분에 대해 분석 결과를 보조해 주어야 한다. 본 논문에서는 이러한 문제를 해결하고자 기존에 높은 평가를 받고 있는 Faster R-CNN의 정확도 및 속도를 개선하여 정확도 91% 연산 속도 0.1sec/frame 도출하여 위 병변 진단에 적합한 개선된 알고리즘을 제안하였다. 그러나, 특정 영상의 환경에 대해서는 인식률이 차이가 나는 점을 확인할 수 있었다. 이러한 문제점은 향후 더욱 다양한 환경의 영상 데이터 수집과 하이퍼 파라미터의 추가적인 조정을 통해 개선하고자 한다.

References

- [1] K. W. Jung, Y. J. Won, H. J. Kong, E. S. Lee, "Cancer Statistics in Korea: Incidence, Mortality, Survival, and Prevalence in 2016," *Cancer research and treatment: official journal of Korean Cancer Association*, vol. 51, no. 2, April 2019, pp. 417-430, doi: 10.4143/crt.2019.138.
- [2] Y. L. Kim, W. H. Kim, T. J. Kim, K. W. Choi, "Histotopographic Characterization of Chronic Gastritis Associated with Early Gastric Carcinomas", *The Korean Journal of Gastroenterology*, vol. 24, no. 2, April 1992, pp. 216-223.
- [3] H. Kim, Y. Hwang, H. Sung, J. Jang, C. Ahn, S. G. Kim, et al., "Effectiveness of gastric cancer screening on gastric cancer incidence and mortality in a community-based prospective cohort," *Cancer research and treatment: official journal of Korean Cancer Association*, vol. 50, no. 2, April 2018, pp. 582-589, doi: 10.4143/crt.2017.048.
- [4] S. Ren, K. He, R. Girshick, J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks", *Advances in Neural Information Processing Systems*, vol. 39, June 2015, pp. 1137-1149, doi: 10.1109/TPAMI.2016.2577031.
- [5] R. Girshick, J. Donahue, T. Darrell, J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation", *IEEE Conference on Computer Vision and Pattern Recognition*, June 23-28, 2014, Columbus, USA, pp. 580-587.
- [6] J. R. R. Uijlings, K. E. A. van de Sande, T. Gevers, A. W. M. Smeulders, "Selective search for object recognition", *International Journal of Computer Vision*, vol. 104, April 2013, pp. 154-171, doi: 10.1007/s11263-013-0620-5.
- [7] C. L. Zitnick, P. Dollár, "Edge boxes: Locating Object Proposals from Edges", *Computer - Vision ECCV 2014*, September 2014, pp. 391-405, doi: 10.1007/978-3-319-10602-1_26.
- [8] J. Hosang, R. Benenson B. Schiele, "How Good are Detection Proposals, really?", *British Machine Vision Conference*, September 1-5, 2014, Nottingham, UK, pp. 1-25, doi: 10.5244/C.28.24.
- [9] R. Girshick, "Fast R-CNN", *IEEE International Conference on Computer Vision*, December 7-13, 2015, Santiago, Chile, pp. 1440-1448.
- [10] K. He, X. Zhang, S. Ren, J. Sun, "Spatial pyramid pooling in deep convolutional networks for visual recognition", *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 37, september 2015, pp. 1904-1916, doi: 10.1109/TPAMI.2015.2389824.