

맥락 정보 조절 및 위양성 분류기를 통한 MRE 영상에서 염증성 장질환 검출

이채린¹ 이한상¹ 서니은² 임준석² 홍헬렌^{1*}

¹서울여자대학교 소프트웨어융합학과

²연세대학교 의과대학 영상의학교실

*교신저자

hlhong@swu.ac.kr

Detection of Inflammatory Bowel Disease in MRE Images Using Contextual Information Modulation and Detection Refinement Classifier

Chaelin Lee¹ Hansang Lee¹ Nieun Seo² Joonseok Lim² Helen Hong^{1*}

¹Department of Software Convergence, Seoul Women's University

²Department of Radiology, Yonsei University College of Medicine

*Corresponding author

요약

크론병(Crohn's disease)은 만성 염증성 장 질환으로 크론병의 정확한 진단과 중증도 판정을 위해 MRE 영상에서 염증성 부위를 정확하게 검출하는 것이 필수적이다. MRE 영상에서 염증 검출은 다양한 크기와 위치, 형태와 신호강도를 가지고 있어 검출이 어렵고 높은 밝기 값으로 인해 신호 강도가 유사한 대변, 액체, 혈관 등과 혼동되어 위양성 검출이 나타나는 한계가 있다. 본 연구는 맥락 정보를 조절하는 Focal Modulation Network를 활용한 Mask R-CNN과 염증과 비염증을 분류하는 Detection Refinement Classifier를 제안한다. 공간적 맥락 정보를 반영해 염증 영역을 강조하고 염증과 비염증의 차이를 학습해 위양성을 줄인다. 실험 결과, 제안 방법은 기존 Mask R-CNN mAP(37.9%) 대비 높은 mAP(52.8%)를 보였으며 위양성 검출이 감소됨을 확인하였다.

Abstract

Crohn's disease is a chronic inflammatory bowel disease, and accurately detecting inflammatory regions in MRE images is essential for its diagnosis and severity assessment. Detecting inflammation in MRE images is challenging due to variations in size, location, shape, and signal intensity, as well as similarities with structures such as stool, fluid, and vessels, which lead to a high false positive rate. This study proposes a Mask R-CNN integrated with a context-aware Focal Modulation Network and a Detection Refinement Classifier to classify inflammatory and non-inflammatory regions. The proposed method enhances the detection of inflammatory regions by incorporating spatial context and learning the distinctions between inflammatory and non-inflammatory areas, effectively reducing false positives. Experimental results demonstrate that the proposed method achieves a higher mAP (52.8%) compared to the Mask R-CNN (37.9%) and decreases false positive detection.

키워드: 의료 영상 분석, 객체 검출, 딥러닝, 크론병, 염증성 장 질환

Keywords: Medical image analysis, Object detection, Deep learning, Crohn's disease, Inflammatory bowel disease

*corresponding author: Helen Hong / Seoul Women's University (hlhong@swu.ac.kr)

1. 서론

크론병은 소화기관의 모든 부위에서 발생할 수 있는 만성 염증성 장질환으로 복통, 설사, 체중 감소 등의 증상을 유발하며 재발이 잦고 완치가 어렵다 [1]. 현재 크론병 진단에 주로 사용되는 방법은 대장내시경으로, 장 내부 염증을 시각적으로 확인하고 점막 표면의 염증을 평가하는데 집중한다. 하지만 대장내시경은 침습적인 검사로 반복적인 검사가 필요한 크론병 환자에게 불편함을 준다 [2]. 이에 대안으로 제시되는 자기공명 장조영술(Magnetic Resonance Enterography, MRE) [3]은 비침습적 검사로 크론병으로 인한 장벽 두께의 증가, 조영제를 통한 염증의 강도나 크기, 형태, 위치 등을 포괄적으로 평가할 수 있어 장염증 활성도를 계산할 수 있다 [4]. 이때 사용되는 sMARIA(simplified MR Index of Activity) 점수 [5]는 기존 MARIA(MR index of activity) 점수를 간단화한 점수체계로 환자별 맞춤형 치료 전략을 세울 수 있도록 한다. 그러나 현재 MRE 영상에서 sMARIA 점수를 산출하는 과정은 의사가 수작업으로 염증 부위를 식별하고 평가해야 하기 때문에 많은 시간과 노력이 소요되고 고도의 전문성이 필요하다. 이는 진단의 일관성을 떨어트릴 수 있으며 환자 맞춤형 치료 전략을 세울 때 부정적인 영향을 미칠 수 있다. 이러한 문제를 해결하기 위해 딥러닝 기반의 자동화된 장염증 탐지 모델이 요구된다.

그림 1은 MRE 영상에서 관찰되는 다양한 형태의 장염증 양상을 보여준다. 해당 그림처럼 MRE 영상에서의 장염증은 염증의 부위가 장벽의 두께, 밝기, 위치에 따라 다양한 형태와 크기, 위치에서 나타나기 때문에 정확한 염증 검출이 어렵다. 특히 밝기, 형태 및 구조가 비슷한 장기(혈관)나 장 내 염증이 아닌 영역(대변, 액체)을 염증으로 혼동할 가능성이 있다. 그림 2는 장염증 검출에서의 위양성 검출 예시를 보여준다. 염증(inflammation)은 주로 장벽의 비후와 조영증강 영상 및 T2 강조영상에서의 높은 신호강도로 나타난다. 대변(stool)은 불규칙한 모양과 이질적인 내부 신호 패턴을 가지며, 조영증강 전 T1 강조영상에서의 고신호강도로 인해 염증과 혼동될 수 있다. 장관 내 액체(fluid)는 그 성분에 따라 다양한 신호강도를 보일 수 있다. 혈관(vessel) 역시 선형의 강한 조영증강을 보이는 구조로 보이게 되어, 염증 부위와 혼동될 수 있다. 이는 위양성의 원인이 되어 전체적인 진단의 정확성을 떨어트리는 문제를 발생시킨다. 이를 해결하기 위해 염증과 비염증 영역을 효과적으로 구분하는 접근이 필요하다.

기존 연구들은 내시경 영상과 MR 영상에서 슬라이스의 염증 유무를 분류해 염증을 검출하는 방식으로 접근한다. Klang et al. [6]는 Xception 모델을 활용해 캡슐 내시

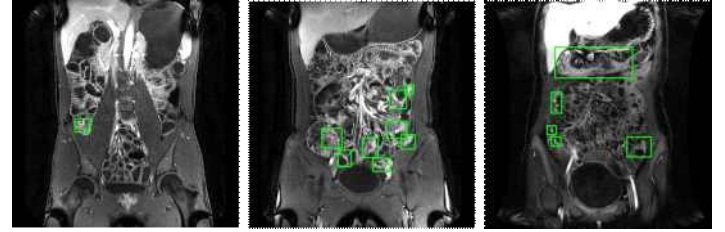


Figure 1: Variability and Detection Challenges of Inflammation in MRE.

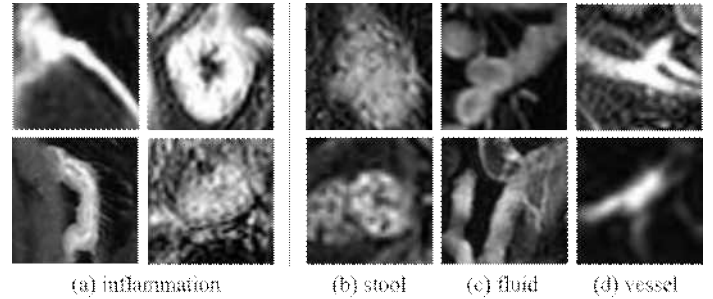


Figure 2: Examples of false positive causes in inflammation detection.

경 영상에서 점막 궤양과 정상 점막 슬라이드를 분류해 염증을 검출했다. Lakovidis et al. [7] 또한 캡슐 내시경 영상에서 Xception 모델을 활용하고 정상 점막과 침식, 궤양 뿐만 아닌 혈관 병변 등의 클래스를 추가하여 슬라이스별 분류를 수행해 염증을 검출했다. Arkko et al. [8]은 3D cine MRE 영상에서 얻어낸 운동성 맵에서 U-net [9]을 사용해 분할된 장 영역에 한정하여 운동성 지수를 계산해 로지스틱 회귀를 통해 크론병 확률을 예측했다. 이러한 접근 방식은 분류와 회귀를 통해 염증이냐 병변의 존재 가능성의 예측에 그치고, 실질적으로 이를 정확히 검출하는 데는 한계가 있다. 이를 위해 병변의 크기와 위치를 예측하는 검출 모델을 통한 접근 방식이 필요하다.

본 논문에서는 정확한 염증 탐지를 위해 다양한 스케일에서 필요에 따라 선택적으로 맥락 정보를 수집하는 Focal Modulation Network(FocalNet) [10]을 Mask R-CNN [11]에 적용하여 맥락 정보를 반영하고, 추가적인 위양성 검출을 감소시키기 위해 염증과 비염증을 분류하는 Detection Refinement Classifier(DRC)를 사용해 위양성을 줄이는 방법을 제안한다. 기존 MRE 영상의 Mask R-CNN을 통한 검출 방법의 위양성 검출 빈도가 높다는 문제를 해결하기 위해 맥락 정보의 중요도를 동적으로 반영해 염증이 있을 가능성이 높은 영역을 강조하고, 염증과 비염증 영역 간의 세밀한 차이를 학습하도록 해 위양성을 줄일 수 있을 것으로 기대된다. 또한 기존 Mask R-CNN과 비교해 위양성 검출 빈도를 유의미하게 줄였음을 검증하고자 한다.

2. 제안 방법

2.1 개요

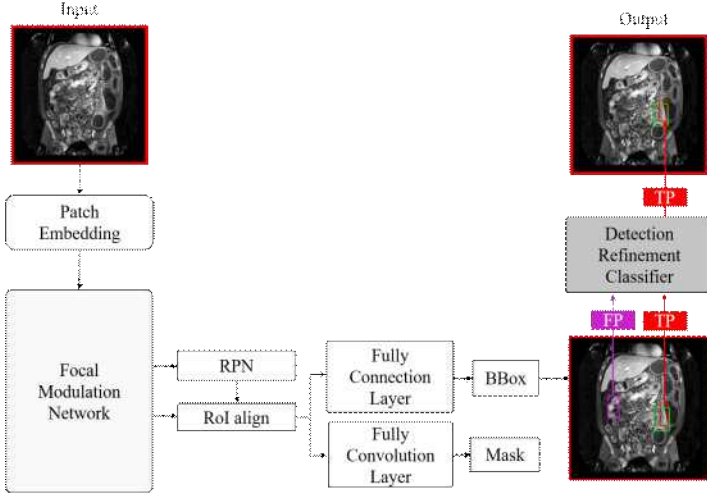


Figure 3: Overview of Proposed Model.

제안된 방법은 FocalNet를 특징 추출 네트워크로 사용하는 Mask R-CNN 기반의 염증 검출과 DRC 학습의 두 단계로 구성된다. 그림 3는 제안 방법을 나타내는 개요도다. 첫 번째 단계에서는 주어진 MRE 영상에서 Mask R-CNN을 활용하여 염증 부위를 검출하고, 바운딩 박스와 마스크를 생성한다. 이 과정에서 FocalNet를 특징 추출 단계에 적용하여 공간적 맥락 정보를 효과적으로 반영한다. 두 번째 단계에서는 생성된 바운딩 박스를 DRC를 통해 위양성 검출을 분류해 제거하고 최종적인 염증 검출 결과를 도출한다.

제안 방법의 기반이 되는 Mask R-CNN은 2-stage 객체 탐지 및 분할 모델로, 입력 영상에서 후보 영역을 생성하고 이를 기반으로 객체의 바운딩 박스를 예측한다. 백본을 거쳐 추출된 특징 맵에서 RPN(Region Proposal Network)을 사용해 객체가 존재할 가능성이 높은 후보 영역(Region of Interest, RoI)을 생성하고, 생성된 RoI에 대해 RoIAlign을 사용해 후보 영역의 특징을 고정된 사이즈로 정규화한다. 이후 Fully Connected Layer를 통해 최종 바운딩 박스 좌표를 예측하는 방식으로 동작한다.

2.2 Focal Modulation Network를 활용한 맥락 기반 특징 추출

기존 ResNet50 [12]은 인접한 영역의 지역적 특징 추출에만 집중하여 주변 영역과의 상호작용 및 관계에 대한 정보를 가지고 있는 전체적인 맥락을 반영하는 데 한계가 있다. 이를 해결하기 위해, 본 연구에서는 ResNet50을 FocalNet로 대체하여 지역적 및 전역적 정보를 효과적으로 학습하도록 설계하였다. FocalNet은 패치 단위로

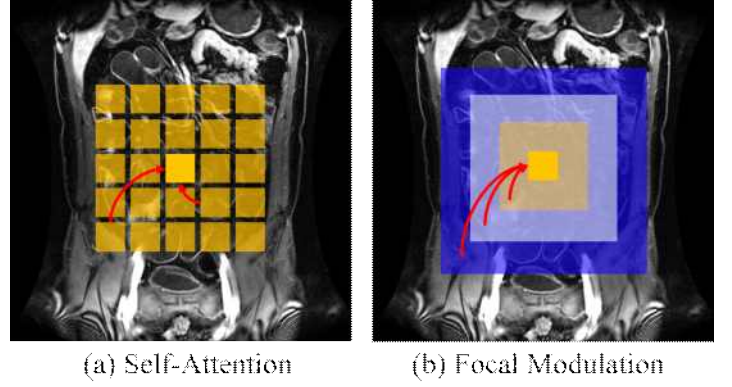


Figure 4: Self Attention vs Focal Modulation.

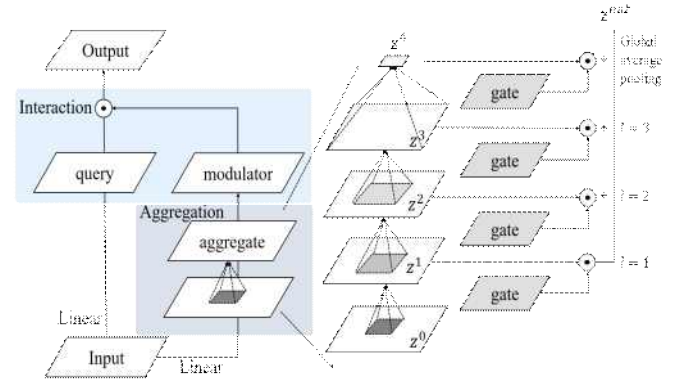


Figure 5: Architecture of Focal Modulation.

입력 영상을 임베딩하고, Focal Modulation Block을 반복 적용하는 구조로 이루어져 있다. 이 과정에서 공간 해상도를 점진적으로 줄이며, 계층적 학습 구조를 통해 지역적 특징과 전역적 맥락을 동시에 반영하여 더욱 정교한 특징 표현을 가능하게 한다.

Focal Modulation Block에서 핵심이 되는 부분은 Focal Modulation으로 염증 영역을 강조하는 역할을 한다. 그림 4는 Self-Attention과 Focal Modulation을 비교한 그림이다. Self-Attention [13]은 모든 위치 간의 상호작용에 동일한 가중치를 부여한다. 반면, Focal Modulation은 거리에 따라 가중치를 다르게 적용하여 지역적 정보와 전역적 정보를 통합한다. 이를 통해 중요한 정보에 집중하면서도 계산 효율성을 높이고 맥락 정보를 조절하여 반영할 수 있다.

그림 5는 Focal Modulation의 자세한 구조를 나타낸다. Focal Modulation은 패치 내에서 염증과 관련된 맥락 정보를 선택적으로 강조하고 불필요한 정보는 억제하며 총 3단계로 구성되어 있다. 첫번째 생성(Generation) 단계에서는 입력된 영상을 선형 레이어를 거쳐 채널별로 query($Q(x_i)$), gate(G), 초기 특징 맵(Z)으로 분리한다. 두번째 집계(Aggregate) 단계에서는 아래 수식을 통해 각 레벨 $l \in \{1, 2, \dots, L\}$ 의 현재 레벨의 특징 맵 Z^l 을 구한다.

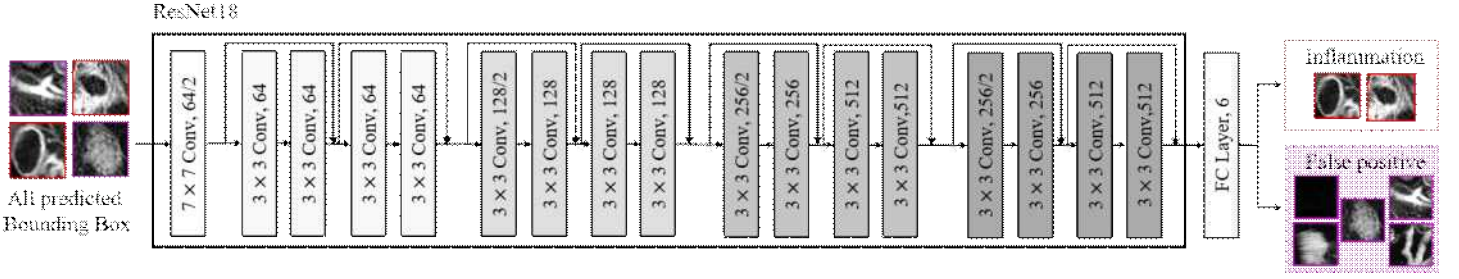


Figure 6: Detection Refinement Classifier.

$$Z^l = f_a^l(Z^{(l-1)}) \equiv \text{GeLU}(\text{DWConv}(Z^{(l-1)})) \quad (1)$$

이때, $Z^{(l-1)}$ 는 이전 레벨의 출력 특징 맵이고, f_a^l 은 각 레벨에서의 맥락화(contextualization) 함수이다. DWConv는 깊이별 합성곱(Depth-Wise Convolution)을 의미하며, GeLU(Gaussian Error Linear Unit)는 비선형 활성화 함수이다. 깊이별 합성곱에서 사용하는 커널(kernel)의 크기는 레벨이 증가할수록 점진적으로 증가함으로써 작은 커널은 인접한 영역에서 세밀한 지역적 맥락 정보를 얻어내고 큰 커널은 더 넓은 영역에서 대략적인 맥락 정보를 얻어낸다. 마지막 $L+1$ 번째 Z^{L+1} 에 대해선 아래 수식 (2)을 통해 전역 평균 풀링(Global Average Pooling)을 수행해 전역 정보를 얻는다.

$$Z^{L+1} = \text{AvgPool}(Z^L) \quad (2)$$

최종적으로 이 단계에서는 수식 (3)을 통해 통합된 맥락 정보 Z_{out} 를 생성한다.

$$Z_{out} = \sum_{l=1}^{L+1} G^l \odot Z^l \quad (3)$$

이때 G^l 은 추출된 특징 맵 Z^l 에 대응하는 게이트 맵이고, \odot 는 요소별 곱셈(element-wise multiplication) 연산을 나타낸다. 최종적으로 모든 레벨에서 추출된 맥락 정보를 더해지게 된다. 이렇게 얻은 통합된 맥락 정보를 활용해 세 번째 상호작용(Interaction) 단계에서는 수식 (4)를 통해 최종 특징 맵 y_i 을 생성한다.

$$y_i = Q(x_i) \odot (\sum_{l=1}^{L+1} (g_i^l \cdot z_i^l)) \quad (4)$$

이때 $\sum_{l=1}^{L+1} (g_i^l \cdot z_i^l)$ 는 수식 (3)에서 집계된 맥락 정보를 의미하며 $Q(x_i)$ 는 첫번째 생성 단계에서 만들어진 query를 의미한다. 이러한 과정을 거친 특징 맵은 염증 가능성이 높은 영역이 강조되어 있어 이후 Mask R-CNN의 RPN 단계에서 염증 가능성이 높은 영역에 바운딩 박스를 제안할 수 있도록 하는 효과를 가진다.

2.3 Detection Refinement Classifier

FocalNet을 통해 맥락 정보를 고려한 Mask R-CNN에서 예측된 바운딩 박스에는 여전히 혈관, 대변, 액체 등 염증과 유사한 비염증 요소로 인해 위양성이 발생하는 한계가 있다. 이를 해결하기 위해, 예측된 바운딩 박스의 염증 유무를 재확인하고 위양성을 제거하는 DRC를 제안한다.

그림 6은 DRC의 구조를 나타낸다. 이 모델은 염증 및 비염증 영역의 특징을 학습하여 정확성을 높이는 데 중점을 둔다. DRC는 전체 MR 영상 대신 Mask R-CNN에서 추출된 바운딩 박스에 기반한 정방향 패치 영상과 정답 값에 기반한 패치 영상을 입력으로 받는다. ResNet18을 특징 추출기로 사용하여 염증과 5개의 위양성 클래스(대변, 액체, 혈관 및 MRE 영상 관상면에서 장 외부 영역(outlier-bright, outlier-dark))로 구성된 총 6개의 클래스 패치 영상 특징을 학습하고, 완전 연결 계층을 통해 각 클래스로 분류하도록 훈련된다. 분류 결과 위양성 클래스로 판단된 패치는 제거하고, 염증으로 분류된 패치만 최종 결과로 유지한다. 이 과정은 염증과 유사한 비염증 요소로 인해 발생하던 위양성을 효과적으로 제거하며, 최종적으로 탐지 결과의 정확성을 향상시킨다.

3. 실험 및 결과

3.1 데이터

본 연구에서는 2016년 3월부터 2018년 12월까지 수집된 크론병 환자 중 활동성 염증이 관찰된 180명과 활동성 염증이 없는 28명으로 구성된 MR 소장 조영술(MRE) 영상을 사용하였다. 활동성 염증이 관찰된 데이터셋은 sMARIA 점수 기준으로 경증(1-8 사이) 환자 105명, 중증(8 이상) 환자 75명에 대해 다양한 상태의 환자들이 포함되어 있다. 모든 MRE 영상은 정맥기(portal phase)에서 촬영된 조영증강 영상으로 해상도는 512×512 화소이며 화소 크기는 각각 $0.78\text{mm} \times 0.98\text{mm}$ 이고 절편 두께는 3-4mm 범위이다. 염증은 3명의 영상의학과 전문의에 의해 개별적으로 수동 분할하여 레이블링하였으며, 총 16,542장의 영상에 대해 7810개의 바운딩 박스 좌표 데이터를 얻었다. 데이터셋은 학습용

Table 1: Slice-by-slice performance comparison between the baseline and proposed methods

Methods	mAP (%)		Precision (%)		Recall (%)	
	@10	@50	@10	@50	@10	@50
ResNet50(Baseline)	37.9	29.7	32.2	24.6	42.8	34.6
FocalNet	39.7	30.7	35.0	26.7	43.2	34.4
ResNet50+DRC	49.9	38.3	42.5	35.1	50.2	41.6
FocalNet+DRC	52.8	44.3	50.5	41.9	55.4	47.0

8038장, 검증용 4474장, 테스트용 4030장으로 분할하여 사용했다. 이때 염증과 비염증 슬라이스의 비율은 각 데이터셋 모두 약 1:3로 구성되어 있다.

Mask R-CNN에서 탐지된 바운딩 박스를 활용해 DRC를 위한 별도의 데이터셋을 구축하였다. 바운딩 박스의 긴 변을 기준으로 정사각형 형태의 패치를 추출한 후 224×224 해상도로 조정하여 사용하였다. 학습 데이터는 정답 바운딩 박스와 IoU 점수가 0점인 위양성 박스로 구성되었으며 염증(ground truth, GT) 3937장, 대변 1464장, 혈관 174장, 액체 220장, 장 외부 영역 중 밝은 영역 584장, 장 외부 영역 중 어두운 영역 86장의 총 6개의 클래스로 구별되었다. 추가적으로 테스트 데이터셋 또한 DRC의 성능을 확인하기 위해 학습 데이터와 동일하게 1명의 영상의학과 전문의의 검토하여 클래스 라벨링을 수행하여 별도의 데이터셋을 구성하였다.

3.2 실험 환경 및 계획

본 연구에서는 NVIDIA GeForce RTX 4070 Ti와 Python 언어를 사용하여 학습과 평가를 진행했다. FocalNet을 활용한 Mask R-CNN 학습은 mmDetection [14] 프레임워크를 사용하여 MS COCO 데이터 [15]로 사전 훈련된 모델을 초기 가중치로 설정하였다. 최적화를 위해 AdamW 옵티마이저를 사용하였으며 배치 사이즈는 4로 설정하였다. 학습률은 $3e-5$ 에서 $4e-5$ 로 조정하고 4,5 epoch에서 기존 학습률(learning rate)을 기존의 10분의 1로 감소시켰다. FocalNet의 초기 커널 크기는 9×9 로 설정했고, 맥락 추출 단계(Focal level)는 3으로 설정하여 레벨이 증가할 때마다 커널 크기의 가로 및 세로 길이가 2씩 증가하도록 하였다. DRC 학습은 주피터 노트북 환경에서 진행되었으며 ImageNet 1K 데이터셋을 사용하여 사전 훈련된 ResNet18을 사용하였다. 모델을 최적화 하기 위해 AdamW 옵티마이저를 사용했고 배치 사이즈는 8, 학습률은 $3e-7$ 로 설정하였다. 또한, 과대적합을 방지하기 위해 조기 종료 방법을 사용하여 검증 손실값이 이전 epoch 검증 손실값보다 10번 낮아지지 않으면 학습을 종료하도록 했다.

DRC 학습 시 혈관과 액체 클래스의 경우 데이터셋의 규모

가 상대적으로 작고 형태가 다양하여 학습이 어려운 문제가 있을 수 있다. 이를 해결하기 위해 상하 Flip 기법을 활용한 데이터 증강을 수행하였으며 클래스 간 불균형 문제를 보완하기 위해 클래스별 데이터셋 수에 반비례하는 클래스 가중치를 적용한 크로스 엔트로피 손실 함수를 사용하였다. 추론 과정에서는 검증 및 테스트 데이터셋에서 Mask R-CNN이 예측한 모든 바운딩 박스를 입력으로 사용하였다.

실험에서는 제안된 방법의 효과를 검증하기 위해 Baseline을 설정하였다. Baseline은 ResNet50을 특징 추출기로 사용하는 Mask R-CNN을 설정하여 기존 성능과 비교할 기준을 마련했다. 또한 DRC의 효과를 평가하기 위해 DRC 적용 여부에 따른 성능을 비교했다. 각 설정에 대해 모델을 학습한 후, 슬라이스 단위 mAP(mean Average Precision), 정밀도(Precision), 재현율(Recall) 등의 지표를 통해 성능을 측정하고 비교하였다. mAP의 경우 위양성 감소를 보다 정확하게 반영하여 평가할 수 있도록 임계값을 초과한 모든 바운딩 박스를 정답으로 인정하도록 설정했다. 또한 정답 바운딩 박스가 없는 슬라이스의 경우는 모든 예측된 바운딩 박스가 위양성으로 AP를 0%로 설정하고, 예측된 바운딩 박스가 없을 경우 AP를 100%로 설정해 처리하였다.

3.3 실험 결과

표 1은 전체 염증 부위에 대해 제안 방법과 기존 방법의 장염증 검출 성능을 비교한 결과를 나타낸다. 기존의 Mask R-CNN은 염증과 유사한 밝기와 형태의 장기나 장 내 염증이 아닌 비염증 영역이 혼동되면서 mAP@10 37.9%, Precision@10 32.2%의 낮은 성능을 기록했다. 반면 제안된 FocalNet은 기존 방법 대비 mAP@10에서 1.7%, Precision@10에서 2.8% 성능 향상을 기록하였다. 기존 방법보다 FocalNet에서 성능이 개선되는 것은 지역적 특징 정보뿐만 아니라 픽셀 간 장거리 의존성과 전역적 맥락 정보를 통합적으로 반영했기 때문으로 분석된다. 이러한 맥락 정보의 반영은 염증과 비염증 요소 간의 복잡한 차이를 효과적으로 학습할 수 있도록 하여, 기존 ResNet50 기반 방법에서 발생했던 위양성 탐지를 감소시키는 데 기여하였다. 다음으로 DRC를 적용한 모델(ResNet50+DRC)은 mAP@10 49.9%,

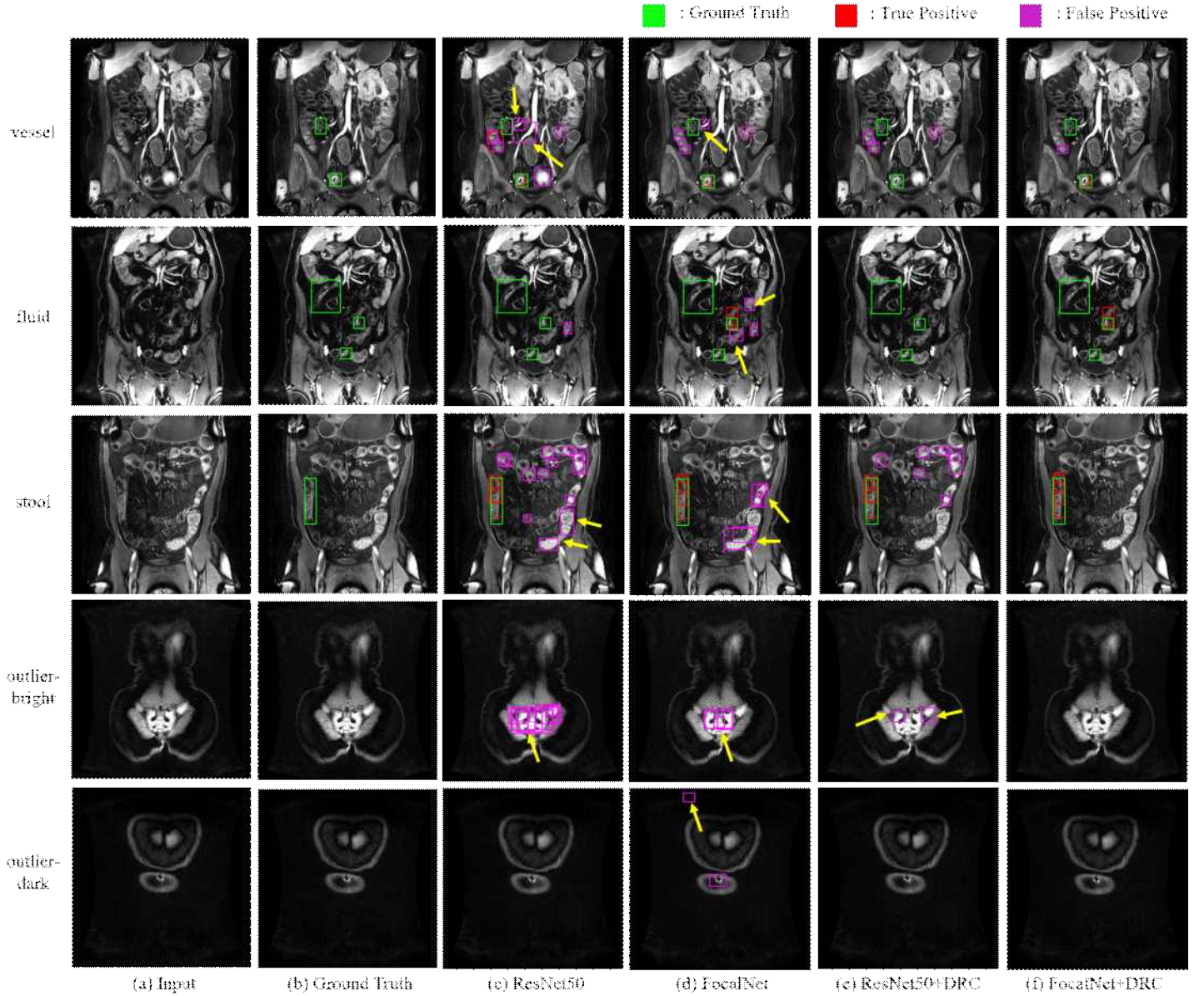


Figure 7: False Positive Reduction in Inflammation Detection

Precision@10 42.5%, Recall@10 50.2%를 기록하여, ResNet50(Baseline) 대비 각각 12.0%, 10.3%, 7.4% 향상된 성능을 보였다. FocalNet에 DRC를 적용한 최종 제안 방법 (FocalNet+DRC)은 mAP@10 52.8%, Precision@10 50.5%, Recall@10 55.4%를 기록하여, FocalNet 단독 모델 대비 각각 13.1%, 15.5%, 12.2% 향상된 성능을 달성하였다. 이러한 성능 향상은 DRC가 바운딩 박스 단위에서 염증과 신호 강도나 형태가 유사한 비염증 영역을 분류하고, 특히 비염증 슬라이스에서 발생하는 위양성을 효과적으로 제거했기 때문이다.

표 2는 DRC를 통해 제거된 위양성 클래스의 비율을 나타냈다. 클래스별 위양성 제거 비율은 각 클래스에 해당하는 바운딩 박스 중 위양성으로 분류되어 제거된 비율로 계산되었다. 대변의 경우 전체 제거율 81.0%와 유사한 비율로 제거

되었으나 혈관, 액체와 같은 영역은 60%대로 낮은 제거율을 보였다. 이는 혈관과 액체의 패치 영상 수가 적고 형태가 다양하여 변동성이 심해 DRC가 학습하는데 어려움이 있었던 것으로 분석된다. 그 외 장 외부 영역에서는 90% 이

Table 2: Reduction Rate from DRC

False Positive Class	Reduction Rate (%)
False Positive (All)	81.0
stool	84.9
vessel	61.9
fluid	67.7
outlier-dark	100
outlier-bright	91.0

상 높은 제거 성능을 기록하였다. 이는 DRC가 기존 Mask R-CNN에서 발생한 위양성을 감소시키며 최종 탐지 성능을 향상 시키는데 기여했음을 알 수 있다.

그림 7은 각 방법의 결과를 시각화해서 비교한 것이다. 노란색 화살표는 각각 혈관, 액체, 대변, 장 외부 영역 중 밝은 영역, 장 외부 영역 중 어두운 영역을 나타내며 Baseline은 이러한 비염증 요소를 염증으로 잘못 탐지한 바운딩 박스가 많다. FocalNet은 일부 위양성을 줄였지만 여전히 비슷한 신호의 위양성이 검출되었다. 반면 제안된 DRC를 제안한 방법은 노란색 화살표로 표시된 타 장기 위양성을 제거하고 정답 값과 유사한 탐지 결과를 보이는 걸 확인할 수 있다. 이는 DRC가 염증과 유사한 신호를 가진 비염증 요소를 구별하는 학습 과정을 추가적으로 진행하여 모델이 보다 정확하게 염증 부위를 탐지하도록 유도했음을 보여준다. 정답 값이 존재하지 않는 비염증 슬라이스에서도 Baseline와 FocalNet은 다수의 위양성 박스를 생성하며, 신호 강도가 높은 영역을 염증으로 잘못 탐지하는 경향을 보인다. 반면 DRC를 추가한 경우 위양성이 대부분 제거되었으며 불필요한 탐지가 최소화되었다. 이는 DRC가 정답 값이 존재하지 않는 슬라이스에서도 위양성을 효과적으로 제거하여 성능을 개선했음을 보여준다.

4. 결론

본 연구에서는 만성 염증성 장 질환의 진단 및 중증도 판정을 위해 MRE 영상에서 염증 부위를 정확히 검출하고 위양성을 줄이기 위해 FocalNet과 DRC를 활용한 방법을 제안하였다. 제안된 방법은 지역적인 정보에만 의존한 기존 Mask R-CNN에 공간적 맥락 정보를 반영하는 FocalNet를 적용하고, 염증과 비염증 영역을 분류하는 DRC를 도입했다. 실험 결과 제안된 방법은 기존 방법 대비 mAP와 정밀도에서 성능 향상을 보였으며, 특히 위양성 검출이 감소하여 정확도가 개선됨을 확인하였다. 또한 염증과 비염증 영역을 구분함으로써 염증 탐지의 신뢰성을 높이는데 기여하였다. 향후 연구에서는 Self Attention을 사용한 Mask R-CNN과 비교를 통해 맥락 정보를 조절하는 FocalNet의 명확한 효과를 파악하는 등의 정량적 비교를 수행하고, 검출된 염증 부위의 정량적 지표와 sMARIA 점수 간의 상관관계를 분석하여 임상적 유용성을 검증할 예정이다.

감사의 글

본 연구는 정부(과학기술정보통신부)의 재원으로 한국연구재단의 지원(No. RS-2024-00336063, No. RS-2023-00207947) 및 서울여자대학교 학술연구비(2025-0213)의 지원을 받아 수행된 연구임.

References

- [1] K. Cushing and P. D. R. Higgins, "Management of Crohn disease: A review," *JAMA*, vol. 325, no. 1, pp. 69-80, 2021.
- [2] E. Braithwaite, J. Carbonell, J. S. Kane, D. Gracie, and C. P. Selinger, "Patients' perception of colonoscopy and acceptance of colonoscopy-based IBD-related colorectal cancer surveillance," *Expert Rev. Gastroenterol. Hepatol.*, vol. 15, no. 2, pp. 211-216, 2021.
- [3] A. Mantarro, P. Scalise, E. Guidi, and E. Neri, "Magnetic resonance enterography in Crohn's disease: How we do it and common imaging findings," *World J. Radiol.*, vol. 9, no. 2, pp. 46-54, 2017.
- [4] S. H. Park, S. H. Park, and B. D. Ye, "Interpretation of enterography in patients with Crohn's disease," *Korean J. Abdom. Radiol.*, vol. 5, no. 1, pp. 1-16, 2021.
- [5] J. Roseira, A. R. Ventosa, H. T. de Sousa, and J. Brito, "The new simplified MARIA score applies beyond clinical trials: A suitable clinical practice tool for Crohn's disease that parallels a simple endoscopic index and fecal calprotectin," *United Eur. Gastroenterol. J.*, vol. 8, no. 10, pp. 1208-1216, 2020.
- [6] Y. A. Klang, A. Lahat, A. Nevler, D. Yablecovitch, S. Neuman, N. Levhar, B. Avidan, R. Yanai, O. Ben-Bassat, U. Kopylov, S. Ben-Horin, and R. Eliakim, "Deep learning algorithms for automated detection of Crohn's disease ulcers by video capsule endoscopy," *Gastrointest. Endosc.*, vol. 92, no. 3, pp. 620-630, 2020.
- [7] D. K. Iakovidis, G. Dimas, A. Karargyris, F. Bianchi, G. Ciuti, and A. Koulaouzidis, "Deep learning and colon capsule endoscopy: Automatic detection of blood and colonic mucosal lesions using a convolutional neural network," *Comput. Biol. Med.*, vol. 116, p. 103570, 2020.
- [8] A. Arkko, T. Kaseva, E. Salli, T. Mäkelä, S. Savolainen, and M. Kangasniemi, "Automatic detection of Crohn's disease using quantified motility in magnetic resonance enterography: Initial experiences," *Clin. Radiol.*, vol. 77, no. 2, pp. 96-103, 2022.
- [9] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional networks for biomedical image segmentation," in *Proc. Int. Conf. Med. Image Comput. Comput.-Assist. Intervent. (MICCAI)*, pp. 234-241, 2015.
- [10] J. Yang, C. Li, X. Dai, and J. Gao, "Focal modulation networks," in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, vol. 35, 2022.
- [11] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R

- CNN,” in *Proc. IEEE Int. Conf. Comput. Vis. (ICCV)*, pp. 2961-2969, 2017.
- [12] K. He, X. Zhang, S. Ren, and J. Sun, “Deep residual learning for image recognition,” in *Proc. IEEE Conf. Comput. Vis. Pattern Recognit. (CVPR)*, pp. 770-778, 2016.
- [13] A. Vaswani, N. Shazeer, N. Parmar, J. Uszkoreit, L. Jones, A. N. Gomez, L. Kaiser, and I. Polosukhin, “Attention is all you need,” in *Proc. Adv. Neural Inf. Process. Syst. (NeurIPS)*, pp. 5998-6008, 2017.
- [14] K. Chen, J. Wang, J. Pang, Y. Cao, Y. Xiong, X. Li, S. Sun, W. Feng, Z. Liu, J. Xu, Z. Zhang, D. Cheng, C. Zhu, T. Cheng, Q. Zhao, B. Li, X. Lu, R. Zhu, Y. Wu, J. Dai, J. Wang, and D. Lin, “MMDetection: Open MMLab detection toolbox and benchmark,” *arXiv preprint arXiv:1906.07155*, 2019.
- [15] T. Lin, M. Maire, S. Belongie, J. Hays, P. Perona, D. Ramanan, P. Dollár, and C. L. Zitnick, “Microsoft COCO: Common objects in context,” in *Proc. Eur. Conf. Comput. Vis. (ECCV)*, pp. 740-755, 2014.

〈 저자 소개 〉

이 한 상

- 2011년 2월 한국과학기술원 전기 및 전자공학과 졸업(학사)
- 2013년 2월 한국과학기술원 전기 및 전자공학과 졸업(석사)
- 2019년 2월 한국과학기술원 전기 및 전자공학과 졸업(박사)
- 2019년 3월~2025년 2월 한국과학기술원 정보전자연구소 연수연구원
- 2025년 3월~현재 서울여자대학교 미래산업융합연구소 연구교수
- 관심분야 : 인공지능, 딥러닝, 컴퓨터 비전, 의료영상처리 및 분석
- <https://orcid.org/0000-0003-0917-1589>



이 채 린

- 2024년 2월 서울여자대학교 소프트웨어융합학과 졸업(학사)
- 관심분야 : 인공지능, 딥러닝, 의료영상처리 및 분석
- <https://orcid.org/0009-0004-9472-4225>



서 니 은

- 2008년 02월 울산대학교 의학과 졸업(학사)
- 2012년 02월 울산대학교 의학과 졸업(석사)
- 2016년 02월 울산대학교 의학과 졸업(박사)
- 2015년 03월~현재 연세대학교 의과대학 영상의학과 교수
- 관심분야 : 복부영상의학, 위장관질환, 자기공명영상, 영상분석
- <https://orcid.org/0000-0001-8745-6454>



임 준 석

- 1994년 02월 연세대학교 의과대학 의학과 졸업(학사)
- 1998년 02월 연세대학교 의과대학 진단방사선과 졸업(석사)
- 2007년 02월 연세대학교 의과대학 진단방사선과학 졸업(박사)
- 2005년 03월~현재 연세대학교 의과대학 영상의학교실 교수
- 관심분야 : 복부영상의학, 디지털헬스, 인공지능
- <https://orcid.org/0000-0002-0334-5042>





홍 헬 렌

- 1994년 2월 이화여자대학교 전자계산학과 졸업(학사)
- 1996년 2월 이화여자대학교 전자계산학과 졸업(석사)
- 2001년 8월 이화여자대학교 컴퓨터학과 졸업(박사)
- 2001년 9월~2003년 7월 서울대학교 컴퓨터공학부 BK 조교수
- 2006년 3월~현재 서울여자대학교 소프트웨어융합학과 교수
- 관심분야 : 인공지능, 딥러닝, 의료영상처리 및 분석
- <https://orcid.org/0000-0001-5044-7909>