

딥러닝과 데이터 증강을 활용한 췌장 감지 및 크기 모니터링

강성민¹ ○ 송재하² 김지희¹ *

¹ 동국대학교 인공지능학과 ² 동국대학교 컴퓨터공학과
okangsungmin@gmail.com, wychrodjxz691@gmail.com, jihie.kim@dgu.edu

Deep Learning-Based Pancreas Detection and Size Monitoring with Data Augmentation for Medical Imaging Analysis

Sungmin Kang¹ ○ Jaeha Song² Jihie Kim¹ *

¹ Department of Artificial Intelligence, Dongguk University

² Department of Computer Science and Engineering, Dongguk University

요 약

국제 의료 가이드라인에 따라 췌장 크기 추적 및 관리의 중요성¹⁾이 높아짐에 따라 YOLOv8 기반 딥러닝 모델을 사용하여 췌장 크기를 측정하는 방법을 제시한다. 본 연구는 전문가의 높은 라벨링 비용을 줄이고 췌장 크기 추적에 대한 가능성을 시사한다. 컴퓨터 단층 촬영(CT)의 복셀(voxel) 수준의 데이터 셋에서 췌장만으로 MONAI¹⁾ 프레임워크 [1]를 통해 26,172개의 이미지(pixel)를 증강하여 데이터를 학습하였으며, 기존의 췌장 크기 추적 방법을 소프트웨어로 대체하고, 미세한 췌장 크기 측정을 비용 효율적으로 수행할 수 있도록 한다. 췌장의 객체 탐지에서 75.31%의 평균 정밀도를 보이며, 크기 추정이 가능한 모델로 숙련된 의료진의 의존도를 최소화하는 솔루션을 제공한다. 이러한 연구는 딥러닝을 활용하여, 보다 효율적인 능동적 환자 감시를 통해 조기 진단을 원활하게 지원한다.

1. 서 론

국제 의료 지침에 따라 췌장 크기의 관리 및 추적은 의학 분야에서 가장 중요한 의미를 지닌다. 복잡한 해부학적 특성을 가진 췌장은 혈당 수치와 소화 과정을 조절하는 데 중추적인 역할을 하고, 췌장 종양, 췌장염, 낭포성 병변과 같은 다양한 병변에 의심이 되는 췌장 상태를 진단하고 모니터링하는 방법으로 크기와 형태에 대한 정확한 추적 감시가 필수적이기 때문에 특히 컴퓨터 단층 촬영(CT)은 췌장과 그 크기를 시각적으로 표현하는 데 중요한 역할을 한다. 의료 영상에서 췌장 크기를 추적하는 기존 방법은 일반적으로 의료 전문가의 수동 레이블링에 의존한다. 그러나 이러한 접근 방식은 시간이 많이 걸릴 뿐만 아니라 숙련된 전문가가 필요하기 때문에 관련된 의료진의 상당한 시간과 노력이 요구되는 노동 집약적인 프로세스이다. 이처럼 높은 비용이 소모되고, 궁극적으로는 환자 대기 시간이 길어지고 의료 비용이 증가한다.

더욱이, 특히 CT 스캔에서 췌장 크기를 예측하는 정확성은 상당한 어려움에 직면한다. 췌장은 복부 내에 자리잡은 상대적으로 작은 기관인 췌장은 입력 볼륨의 상당한 부분을 차지하는 복잡하고 가변적인 배경 영역과 주변 조직과의 경계가 모호함을 초래하여 예측의 정확도는 만족스럽지 않기 때문에 데이터 기반의 접근 방식이 필요하다.

따라서 본 연구에 대한 기여는 다음과 같이 요약된다.

- YOLOv8 기반 객체 탐지 모델을 사용하여 대규모 복부 CT(컴퓨터 단층 촬영) 데이터 세트에서 췌장을 정확하고 자율적으로 감지하고 이를 통해 췌장 크기를 측정하는 새로운 접근 방식을 제시한다.
- MONAI 프레임워크를 통하여 의료 도메인에 맞는 데이터 증강(Data augmentation)²⁾ 방식의 적용으로 Spatial과 Intensity 기반의 transform을 비교 연구하여 성능을 검증한다.



그림 1: 췌장에 대한 CT 이미지(좌측)와 레이블링(중간)

2. 관련 연구

기존의 췌장과 췌장의 양성 병변을 감지하기 위한 U-Net 기반의 신경망 [2],[3]은 제안되었다. 그러나 이러한 알고리즘은 임상적 의의 높은 레이블링 비용 및 추가적인 계산 리소스 요구 등으로 실제 임상 환경에서 쉽게 적용하기 어려운 제약 사항이 있다. 또한, 이러한 알고리즘은 췌장 크기를 정량적으로 측정하는 기능을 제공하지 않는 한계가 있다. 췌장 크기 측정은 의료 진단 및 환자

1) <https://monai.io/>

2) <https://docs.monai.io/en/latest/transforms.html>

모니터링에서 중요한 역할을 하며, 정확한 크기 측정은 환자 상태 평가 및 치료 결정에 영향을 미친다. 따라서 체장 크기를 정량적으로 측정할 수 있는 신경망 기반의 알고리즘의 필요성이 대두된다.

3. Yolo v8: 객체 탐지 및 크기 알고리즘

모델은 객체 탐지 알고리즘으로서 2023년 기준 제안된 YOLO (You Only Look Once)의 v8 [4]을 사용하였다. 아래 그림 2에서 모델의 성능이 이전 버전과의 차이를 보여준다. 좌측 parameter에 따른 정확도(mAP), 우측: 최적화 프레임워크에 처리 속도에 따른 정확도(mAP)이다. 이전 버전과의 개선된 점은 Backbone network, Anchor-Free head 변화와 새로운 Loss function의 추가로 성능 개선을 보였다고 한다.

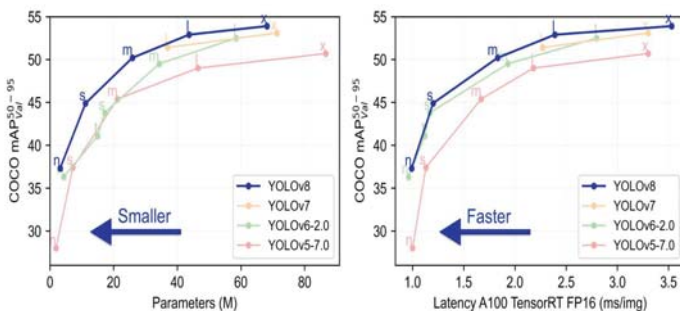


그림 2: YOLO v8 이전 버전과의 성능 차이 비교

이러한 객체 감지 기술을 활용하여 predicted bbox의 크기를 측정하면 이를 통해 nibabel의 get_zooms() 메소드를 활용하여 복셀 수준에서의 데이터 축에 대한 간격 (millimeters) x, y, z을 얻을 수 있고 아래와 같은 계산식으로 최종적으로 체장의 실제 크기를 유추할 수 있게 하였다.

$$\text{Actual length(mm)} = \max(\alpha(w), \beta(h))$$

* w, h = width, height
* α, β = voxel size x, y

3.1. 데이터 증강

데이터 증강은 데이터의 다양성을 향상시켜 모델의 일반화 능력을 향상시킨다. 데이터 수집이 어려운 경우 소규모 데이터셋을 확장하는 하거나 데이터의 변화를 통해 노이즈를 모방하여 강건한 모델의 성능을 얻도록 한다.

일반적인 YOLOv8에 내장된 Augmentation의 경우 이미지 회전, 뒤집기 등의 일반적인 방식에서 보다 전문적인 오픈소스 의료이미지 프레임워크인 MONAI의 데이터증강을 활용하여 Spatial과 Intensity 기반의 transform을 비교 실험하였다. 이미지 내의 객체 중심으로 Spatial에서는 아핀(Affine) 변환, 랜덤 비율로의 확대, 그리드 왜곡, 탄력적 변환이고 Intensity에서는 무작위 임의의 제거, 밝기 또는 대비 변화, 가우시안 스무딩, 임계값으로 intensity 조정 하여 비교하고자 했다.

<표 1> 데이터 증강 구분

Spatial	Intensity
RandAffine	RandCoarseDropout
RandZoom	RandBiasField
GridDistortion	GaussianSmooth
Rand3DElastic	ThresholdIntensity

3.2. 성능 평가

본 연구에서는 객체 탐지에서 사용되는 평가 지표인 평균 정밀도(Mean Average Precision, mAP)를 통해 모델을 평가하였다. 이는 객체 감지 모델의 성능을 정량화하고 평가하는 중요한 지표로서 여러 객체 클래스에 대한 개별 평균 정밀도(AP)의 평균값으로 나타낸다. 정밀도(Precision), 재현율(Recall), IoU(Intersection over Union)와 같은 핵심 개념을 포함한다. IoU는 객체 감지 모델의 정확도를 측정하는 지표로, 계산식은 아래와 같다.

$$\text{IoU} = \text{Area of overlap} / \text{Area of Union}$$

높은 IoU는 모델이 정확한 객체 위치를 감지하는 것을 의미하며, 임계값(threshold)에 따라 정밀도와 재현율을 계산하게 된다. 모델이 예측을 수행할 때, 임계값을 변화시키면서 각 임계값에 대한 정밀도와 재현율을 그래프 형태로 그린 PR(Precision-Recall) 곡선을 얻을 수 있는데 이때, PR 곡선의 아래 면적이 바로 AP로서, 모델의 성능을 종합적으로 나타낸다. 높은 AP는 모델이 다양한 임계값에서 높은 정밀도와 재현율을 유지하고, 다양한 객체 클래스에 대해 우수한 성능을 가지고 있다는 것을 의미한다.

4. 데이터셋 구성

모델 학습의 경우 복부 CT에서 다기관 분할(Multi-Organ segmentation) 및 체장 분할 등 총 4가지 benchmark dataset에서 체장 mask만을 사용하였다. AMOS [5], BTCV, MSD [6]의 경우 다기관 분할에 초점을 맞추어 강건한 분할 알고리즘을 연구하기 위해 제안된 데이터 셋으로 복부 기관에 대한 각각 15개, 13개, 8개로 레이블링 되었다. 그리고 NIH(The National Institutes of Health Clinical Center) 또한 체장만 레이블링 되어있어 연구에 사용하였다. 모든 데이터는 NIfTI 파일 형식으로 제공되었으며 전처리 과정에서 복셀 수준에서 체장만 레이블링된 부분을 추출하여 이미지 jpg 형식으로 변환하였다. 데이터 증강은 표 1에서의 2, 4개의 조합으로 훈련 및 검증은 8:2로 나누어서 학습하였다.

<표 2> 데이터셋 별 활용 이미지 수 (장)

데이터셋	AMOS	BTCV	MSD	NIH
갯 수	9,760	746	8,784	6,882
총 계	26,172			
2개 조합	52,344			
4개 조합	104,688			

4.1. 실험 결과

표 3의 정량적인 결과로 Spatial 기반의 데이터 증강이 상대적으로 좋은 성능을 보였다. 그중 RandZoom와 GridDistortion의 조합이 가장 좋은 성능을 보이는 결과를 보여주었다. 연구에서 Intensity가 상대적으로 낮은 이유는 일반적으로 큰 객체의 경우에는 Feature를 학습하기에 충분한 정보를 담고 있지만 작은 객체에서의 Feature는 부족하기 때문에 Intensity로 이미지를 변형하게 되면 객체 탐지의 낮은 평균 정밀도를 보이는 경향을 보였다.

〈표 3〉 정량적 결과 (%)

Model	Precision	Recall	mAP50
Baseline	77.74	71.86	73.6
RandAffine RandZoom	79.87	74.17	74.91
RandAffine GridDistortion	79.46	74.02	74.7
RandAffine Rand3DElastic	76.3	70.62	72.77
RandCoarse Dropout RandBiasField	78.62	72.22	74.18
RandCoarse Dropout Gaussian Smooth	67.74	62.77	64.96
RandCoarse Dropout Threshold Intensity	63.23	59.92	64.0
+ w/ Spatial	78.16	72.47	73.7
+w/ Intensity	70.86	65.0	67.71
RandZoom GridDistortion	80.23	74.32	75.31

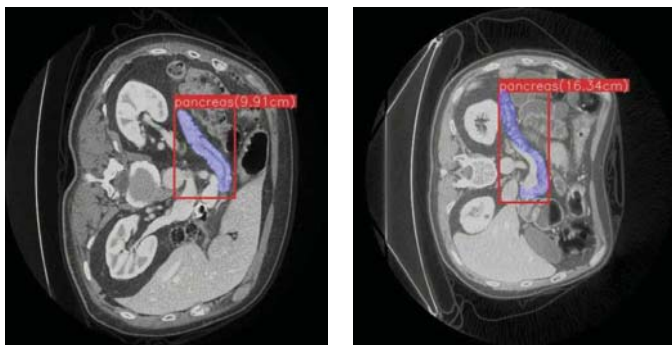


그림 3: YOLO 모델 정성적 결과

5. 결론 및 한계점

본 연구에서는 딥러닝 모델을 적용하여 객체 검출 및 크기 측정에 대한 평가를 진행했다. 본 연구를 통해 YOLO-v8에 의료 도메인에 맞는 데이터 증강 방식의 적용이 적합한 것을 알 수 있고, 연구 결과는 의료 분야, 특히 췌장 추적 분야에서 딥러닝의 잠재력을 보여준다.

자동화된 접근 방식은 전문 지식에 대한 의존도를 크게 줄이고 기존 의료 분야 워크플로우에 원활하게 통합할 수 있는 임상 의사 결정 지원 솔루션(Clinical Decision Support System, CDSS)에 제공하여 췌장 크기 추적을 더욱 확장 가능하고 비용 효율적으로 만들고자 한다.

그러나 모델의 한계점으로 첫 번째는 췌장 크기 추정에 대한 평가가 필요하다. 실제 췌장의 크기를 가진 임상 기록의 의료 데이터와 비교하여 크기가 유사한지에 대한 검증이 이루어져야 하고, 두 번째로는 췌장 추적 감시를 위해 Task-specific 모델로 설계되었으며 다른 기관에 일반화할 수 있는 다양성이 부족하다. 따라서 후향적 연구에서는 이러한 한계점을 극복하여 다기관에 대한 범용적인 모델로 연구를 할 것이다.

Acknowledgment

본 연구는 과학기술정보통신부 및 정보통신기획평가원의 대학ICT연구센터육성지원사업(IITP-2023-2020-0-01789)과 인공지능융합혁신인재양성사업(IITP-2023-RS-2023-00254592)의 연구결과로 수행되었음.

참고 문헌

- [1] Cardoso, M. Jorge, et al. "Monai: An open-source framework for deep learning in healthcare." arXiv preprint arXiv:2211.02701 (2022).
- [2] Abel, L. et al. Automated Detection of Pancreatic Cystic Lesions on CT Using Deep Learning. Diagnostics 2021, 11, 901.
- [3] Duh MM, et al. Deep Learning to Detect Pancreatic Cystic Lesions on Abdominal Computed Tomography Scans: Development and Validation Study JMIR AI 2023;2:e40702
- [4] Terven, Juan, et al. "A comprehensive review of YOLO: From YOLOv1 to YOLOv8 and beyond." arXiv preprint arXiv:2304.00501 (2023).
- [5] Ji, Yuanfeng, et al. "Amos: A large-scale abdominal multi-organ benchmark for versatile medical image segmentation." Advances in Neural Information Processing Systems 35 (2022): 36722-36732.
- [6] Antonelli, Michela, et al. "The medical segmentation decathlon." Nature communications 13.1 (2022): 4128.