

치과 파노라마 엑스레이에서 YOLOv8 기반 사랑니 탐지 성능 향상을 위한 라벨링 전략 비교 연구

김준수*, 김진세*, 서원호*, 양윤기*

*수원대학교,

okjs6577@naver.com , wlstpsla19@gmail.com , syh3942@naver.com , ygyang@suwon.ac.kr

A Comparative Study of Labeling Strategies for Improving YOLO v8-based Wisdom Tooth Detection Performance in Dental Panoramic X-rays

Jun-Su Kim, Jin-Se Kim, Won-Ho Seo, Yoon-Gi Yang

The University of SUWON

요약

본 논문은 파노라마 엑스레이 영상에서 YOLOv8 모델을 이용한 제3대구치(사랑니) 자동 탐지 성능 향상을 위한 데이터 라벨링 전략 비교 연구를 다룬다. 파노라마 영상은 치과 진단에 필수적이거나, 사랑니는 높은 변이성과 인접 치아 유사성으로 인해 자동 탐지가 어렵고 미탐지(False Negative) 시 임상적 문제를 야기할 수 있다. 기존 연구는 주로 일반적인 치아 탐지 성능 개선에 집중하여, 탐지가 어려운 사랑니에 대한 최적의 라벨링 전략 연구는 부족했다. 이에 본 연구는 사랑니 탐지 정확도, 특히 미탐지를 감소를 목표로, 사랑니만 라벨링하는 타겟 중심 전략과 사랑니 및 인접 제2대구치를 함께 라벨링하여 주변 맥락 정보를 제공하는 전략을 설계하고, 공개 데이터셋 기반으로 YOLOv8s 모델을 학습시켜 성능(AP@0.5, Recall@0.5)을 비교 평가하였다. 실험 결과 '맥락 정보 포함' 전략이 '타겟 중심' 전략 대비 Recall 값 11.62%, AP@0.5 값 4.43%의 유의미한 향상을 보여, 라벨링 시 주변 맥락 정보 활용이 사랑니 탐지 신뢰도 향상에 효과적임을 확인하였다. 본 연구 결과는 탐지가 어려운 객체에 대한 AI 모델 학습 시 효과적인 데이터셋 구축 및 라벨링 전략 수립에 방향성을 제시하고 향후 다량의 치과 영상 분석, 자동 진단 보조 시스템 등 임상 현장에서 활용될 AI 시스템의 신뢰성 향상에 기여할 수 있을 것으로 기대한다.

1. 연구 배경

파노라마 엑스레이 영상은 구강 전체의 구조를 한눈에 파악하게 해주어 치과 진단 및 치료 계획에 필수적인 도구이다. 최근 인공지능 특히 YOLO와 같은 딥러닝 객체 탐지 기술은 이러한 의료 영상 분석의 자동화 가능성을 열었으며[1], 이는 진료 효율성 증대에 기여할 수 있다.

그러나 모든 치아, 특히 제3대구치(사랑니)의 자동 탐지는 여전히 기술적 난제이다. 사랑니는 높은 형태와 위치 변이성, 인접 제2대구치와의 유사성, 영상 후방부 위치 및 구조물 중첩[4] 등의 복합적인 요인으로 인해 자동 탐지가 매우 까다롭다. 여기서 말하는 3대구치란 사랑니를 의미하고 2대구치는 사랑니 바로 앞의 어금니, 1대구치는 그 앞의 어금니를 뜻한다.

따라서 본 연구의 목표는 최신 객체 탐지 모델인 YOLOv8을 활용하여, 파노라마 영상에서 사랑니 탐지 성능을 개선하기 위한 효과적인 데이터 라벨링 전략을 비교 분석하는 것이다.

2. 기존 연구와의 비교 및 본 연구의 차별성

딥러닝 기반 자동 치아 탐지 연구는 활발히 진행되어 왔으며, YOLO와 같은 모델은 우수한 성능을 보여주었다. 기존 라벨링 방식[4]은 YOLOv8을 이용해 전체 치아 탐지 및 번호 지정에서 높은 성능을 입증하며 일반적인 치아 탐지에 대한 모델의 잠재력을 확인했다.

그러나 기존 연구들은 주로 전체 치아 대상의 일반적인 성능 향상이나 모델 아키텍처 비교에 초점을 맞추었다. 기존[3,4] 연구에서도 미탐지가 관찰되었으나, 이를 데이터 준비 관점에서 해결하려는 접근은 미흡했다.

본 연구는 다음과 같은 차별성을 갖는다. 첫째, 탐지가 어려운 '사랑니' 탐지 성능 개선에 명확히 초점을 맞춘다. 둘째, 모델 구조가 아닌 데이터 라벨링 전략이 성능, 특히 Recall 향상에 미치는 영향을 데이터 중심적으

로 탐구한다. 셋째, '타겟 중심' 전략과 '맥락 정보 포함' 전략이라는 구체적인 라벨링 방법론을 설계하고 직접 비교 평가한다. 요약하면 본 연구는 '모델'이 아닌 '라벨링' 방법에 따라 특정 어려운 문제를 더 잘 해결할 수 있는지에 대한 결과를 제공한다는 점에서 기존 연구와 차별화를 둔다.

3. 실험 내용

3.1 실험 개요

본 연구는 치과 파노라마 엑스레이에서 YOLOv8s 모델을 이용하여 제3대구치 탐지 성능을 향상시키기 위한 최적의 데이터 라벨링 전략을 비교 분석하는 것을 목표로 한다. 이를 위해 네 가지 다른 라벨링 전략을 적용한 데이터셋을 구축하고, 동일한 조건 하에서 모델을 학습시킨 후 성능을 평가하였다. 2가지는 타겟 중심의 기존 연구방법을 일부 수정해서 사용하고 2가지는 맥락 정보를 포함시킨 새로운 라벨링 방법으로 진행한다.

3.2 데이터셋 구축

본 연구에 사용된 데이터셋은 Kaggle에서 얻은 공개 데이터 300장의 치과 파노라마 엑스레이 기반으로 구축되었다[5]. 데이터셋 구축, 라벨링, 증강 및 관리는 Roboflow 플랫폼을 활용하였다. 각 라벨링 전략별로 구축된 데이터셋은 학습(Train), 검증(Validation), 테스트(Test) 세트로 [80%/10%/10%] 비율로 분할하여 사용하였다.

3.3 라벨링 방법

전략 A : '1,2대구치 통합' 전략 (맥락 정보 포함)

이 전략은 제3대구치 탐지에 있어 가장 폭넓은 맥락 정보를 제공하는 것을 목표로 한다. 제3대구치뿐만 아니라, 제1대구치, 제2대구치 그리고 추

가로 1,2대구치를 합쳐서 라벨링한 총 4개의 클래스를 라벨링 하였다.

전략 B : '2,3대구치 통합' 전략 (맥락 정보 포함)[2]

이 전략은 제3대구치와 직접적으로 인접하여 형태 및 위치 구분에 영향을 줄 수 있는 제2대구치와 2,3대구치를 동시에 통합해서 함께 3가지 클래스로 라벨링하였다. 제3대구치 주변의 핵심적인 맥락 정보 활용에 초점을 맞춘 방식이다.

전략 C : '개별 라벨링' 전략 (타겟 중심)[3]

이 전략은 파노라마 영상에 나타나는 모든 대구치를 개별적으로 구분하는 것을 목표로 한다. 제1대구치, 제2대구치, 제3대구치를 각각 독립된 클래스 3가지로 라벨링 하였다. 기존 라벨링 방식이 이 방식이다.

전략 D : '3대구치 단독' 전략 (타겟 중심)

이 전략은 오직 목표 객체인 제3대구치만을 라벨링하는 가장 단순한 접근 방식으로 오직 1개의 클래스로 라벨링 했다.

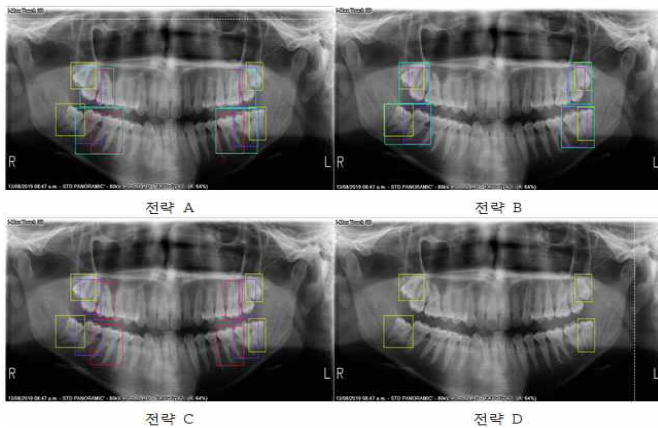


그림 1. 각 전략별 라벨링 예시

3.4 학습 방법

네 가지 라벨링 전략 간의 공정한 성능 비교를 위해 다음과 같은 주요 학습 파라미터를 모든 실험에 동일하게 적용하였다. 사용한 파라미터는 Epochs 는 100, Batch 사이즈 16, 이미지 크기는 640x640 픽셀, 옵티마이저는 AdamW와 Learning Rate는 0.001로 사용하고 진행하였다. 이러한 파라미터와 동일한 데이터 수치를 가지고 YOLOv8s 모델 아키텍처로 학습을 Google Colab의 NVIDIA T4 GPU를 기반으로 학습을 수행하였다.

3.5 실험 결과

각 라벨링 전략의 성능은 주로 제3대구치(Trd_Molar) 클래스의 탐지 정확도에 초점을 맞추어 평가하였다. 주요 평가지표는 다음과 같다.

표 1. 성능 비교표 [Trd_Molar(사랑니) 클래스 기준]

전략	라벨링된 클래스	Trd_Molar AP@0.5	Trd_Molar Precision	Trd_Molar Recall	전체 mAP@0.5
전략 A	Fst_Molar, Non_Molar3, Sec_Molar, Trd_Molar	0.9622	0.9017	0.9605	0.8826
전략 B	Sec_Molar, Trd_Molar, With_Molar2	0.9416	0.9025	0.8953	0.9349
전략 C	Fst_Molar, Sec_Molar, Trd_Molar	0.9233	0.9293	0.8605	0.9238
전략 D	Trd_Molar	0.9214	0.9254	0.8837	0.9214

AP@0.5 (Average Precision at IoU=0.5) : IoU(Intersection over Union) 임계값을 0.5로 설정했을 때의 평균 정밀도. 객체 탐지 모델의 전반적인 성능을 나타내는 핵심 지표이다.

Recall@0.5 : IoU 임계값 0.5에서 실제 제3대구치 중 모델이 올바르게 탐지한 비율. 미탐지율과 직접적으로 연관된다.

Precision@0.5 : IoU 임계값 0.5에서 모델이 제3대구치라고 예측한 것

중 실제 제3대구치인 비율. 오탐지율과 관련된다.

mAP@0.5 : 데이터셋에 포함된 모든 클래스에 대한 AP@0.5의 평균값. 모델의 전반적인 성능을 참고하기 위해 기록하였다.

분석 결과 전략 A가 사랑니 탐지에 있어 전반적으로 가장 우수한 성능을 보였다. 이는 AP@0.5 (0.9622) 와 Recall (0.9605) 모두 네 가지 전략 중 가장 높게 나타났기 때문이다. AP@0.5의 지표는 가장 낮은 전략 D(0.9214) 대비 가장 높은 전략 A(0.9622)에서 약 4.43% 향상 됐다. 전략 A의 Recall (0.9605) 은 다른 모든 전략, 즉 전략 B (0.8953), 전략 C (0.8605), 전략 D (0.8837)와 비교했을 때 월등히 높은 것을 확인 할 수 있다.

전략 A는 전략 B 대비 Recall 값이 6.52%, 전략 C 대비 11.62%, 전략 D 대비 7.68% 가 높게 나타나 가장 사랑니 탐지를 잘하는 라벨링 전략으로 꼽을 수 있다.

다만 Precision값은 기존 방식들과 같은 전략 C,D가 타겟 중심으로 한 전략 A,B 라벨링 방식이 월등히 높은 것을 확인 할 수 있어, Recall값과 Precision값을 동시에 높일 수 있는 추가적인 개선 방법을 모색할 필요가 있다.

4. 연구 결과 요약 및 향후 연구 방향

본 연구의 시스템은 YOLOv8 모델을 기반으로 하며, 탐지가 어려운 사랑니의 정확하고 신뢰성 있는 식별 및 대량의 영상 데이터를 효율적으로 처리하는 것을 목표로 하였다. 성능 평가 결과, 맥락 정보를 포함한 전략을 적용한 시스템이 타겟 중심 전략 대비 전반적으로 높은 탐지율(Recall)을 보여 실험 목적에 맞는 효과를 입증하였다. 그러나 예측(Precision)값은 기존 타겟 중심 전략이 더 높게 나와 여전히 개선이 필요해 보인다. 이를 보완하기 위해, 더욱 다양하고 방대한 데이터셋 확보 및 정교한 라벨링과 모델 구조 최적화가 필요할 것 같다. 이러한 연구들을 통해 사랑니 자동 탐지의 정확성과 신뢰도를 높여, 대량의 치과 영상을 효율적으로 분석하고 임상 현장에서 활용 가능한 AI 기반 진단 보조 시스템 개발에 기여할 수 있을 것을 기대한다.

참 고 문 헌

- [1] Hwang JJ, Jung YH, Cho BH, Heo MS. An overview of deep learning in the field of dentistry. Imaging Sci Dent. 2019;49(1):1-7. doi:10.5624/isd.2019.49.1.1
- [2] Dong-Hyeok NamGung, & Dong-Hoi Kim (2023). Proposal of New Labeling Method for Detection Improvement of Two-People Riding on E-Scooter in YOLOv8n. Journal of Digital Contents Society, 24(7), 1575-1581. 10.9728/dcs.2023.24.7.1575
- [3] Mendes, Alison & Quintanilha, Darlan & Pessoa, Alexandre & Paiva, Anselmo & Neto, Pedro. (2025). Automated Tooth Detection and Numbering in Panoramic Radiographs Using YOLO. Procedia Computer Science. 256. 1318-1325. 10.1016/j.procs.2025.02.244.
- [4] Hamamci, Ibrahim Ethem & Er, Sezgin & Simsar, Enis & Yuksel, Atif & Gultekin, Sadullah & Ozdemir, Serife & Yang, Kaiyuan & Li, Hongwei & Pati, Sarthak & Stadlinger, Bernd & Mehl, Albert & Gundogar, Mustafa & Menze, Bjoern. (2023). DENTEX: An Abnormal Tooth Detection with Dental Enumeration and Diagnosis Benchmark for Panoramic X-rays. 10.48550/arXiv.2305.19112.
- [5] Humans in the Loop. (Accessed [2025-05-05]). Teeth Segmentation on dental X-ray images. Kaggle. Retrieved from <https://www.kaggle.com/datasets/humansintheloop/teeth-segmentation-on-dental-x-ray-images>