한국컴퓨터그래픽스학회
Korea Computer Graphics Society Vol. 28, No. 3, P. 55~65

어텐션 모듈과 기하학적 데이터 증강을 통한 X-ray 영상 내 해부학적 랜드마크 검출 성능 향상

이효정[1○] 마세리[2] 최장환[2*]

[1]이화여자대학교 컴퓨터의학과
[2]이화여자대학교 휴먼기계바이오공학부
{hyojoy, serie}@ewhain.net, choij@ewha.ac.kr

Improved Anatomical Landmark Detection Using Attention Modules and Geometric Data Augmentation in X-ray Images

Hyo-Jeong Lee[1○] Se-Rie Ma[2] Jang-Hwan Choi[2*]

[1]Department of Computational Medicine, Graduate Program in System Health Science and Engineering
[2]Division of Mechanical and Biomedical Engineering, Graduate Program in System Health Science and Engineering
Ewha Womans University

요 약

X-ray 두개골 영상에서 주요 해부학적 부위들 간의 거리를 계측하는 것은 진단과 치료 등 임상적 의미에서 매우 중요하다. 최근에는 딥러닝 기술의 발전을 바탕으로 랜드마크를 식별 및 검출하는 자동화 시스템들이 제시되고 있다. 이러한 딥러닝 기반 모델을 과적합 없이 학습 시키기 위해서는 대량의 영상과 라벨링 데이터가 필요하다. 기존에는 숙련된 판독의가 환자의 영상에서 랜드마크를 수동으로 식별하여 라벨링하는 방식으로 계측이 이루어져 왔다. 그러나 이러한 계측 방식은 많은 비용이 소요될 뿐만 아니라, 재현성이 떨어지기 때문에 자동화된 라벨링 방법에 대한 필요성이 제기되고 있다. 또한, X-ray 영상에는 광자가 통과하는 경로 상의 여러 인체조직들이 표시되기 때문에 랜드마크 식별이 일반 자연 이미지 또는 삼차원 모달리티 영상에 비해 어렵다. 본 연구에서는 X-ray 영상 내에 대량의 라벨링 데이터 생성을 가능하게 하는 기하학적 데이터 증강 기법을 제안하고 있다. 또한, 두개골 내 주요한 16개 랜드마크들의 검출 성능을 향상시키기 위해 다양한 어텐션 기법들의 구현 및 적용을 통해 랜드마크 검출을 위한 최적의 어텐션 메커니즘을 제시하였다. 마지막으로 주요 두개골 랜드마크들 중 안정적인 검출이 보장되는 마커들을 도출하였으며, 이러한 마커들은 임상적인 활용 가능성이 높을 것으로 기대된다.

Abstract

Recently, deep learning-based automated systems for identifying and detecting landmarks have been proposed. In order to train such a deep learning-based model without overfitting, a large amount of image and labeling data is required. Conventionally, an experienced reader manually identifies and labels landmarks in a patient's image. However, such measurement is not only expensive, but also has poor reproducibility, so the need for an automated labeling method has been raised. In addition, in the X-ray image, since various human tissues on the path through which the photons pass are displayed, it is difficult to identify the landmark compared to a general natural image or a 3D image modality image. In this study, we propose a geometric data augmentation technique that enables the generation of a large amount of labeling data in X-ray images. In addition, the optimal attention mechanism for landmark detection was presented through the implementation and application of various attention techniques to improve the detection performance of 16 major landmarks in the skull. Finally, among the major cranial landmarks, markers that ensure stable detection are derived, and these markers are expected to have high clinical application potential.

키워드: 랜드마크 검출, 의료 영상, 심층 학습, 데이터 증강, 어텐션
Keywords: Landmark Detection, Medical Image, Deep Learning, Data Augmentation, Attention

*corresponding author: Jang-Hwan Choi/Ewha Womans University(choij@ewha.ac.kr)

Received : 2022.06.10./ Review completed : 1st 2022.06.29. / Accepted : 2022.07.05.
DOI : 10.15701/kcgs.2022.28.3.55
ISSN : 1975-7883(Print)/2383-529X(Online)

1. 서론

의료 영상에서 객체 검출, 분류, 분할 등의 자동화 알고리즘은 진단의 효율성을 높일 수 있다. 특히 두부 계측(cephalometric) 분석 분야의 자동화는 부정교합 진단이나 두부 치료 계획 수립 등 임상적인 응용에 널리 사용될 수 있기 때문에 딥러닝(deep-learning) 기술이 유용하게 활용될 수 있다. 두부 계측 분석을 자동화하기 위한 초기 연구들은 주로 템플릿 매칭 또는 지식 기반 시스템 등 전통적인 방법을 통해 발전되어 왔다[1, 2]. 그러나 템플릿 매칭 방식을 에지 향상 및 객체 검출에 활용하기 위해서는 선명도와 X선 대비, 노이즈 등에 대한 고려가 필수적이므로 전처리 기술이 정확도에 큰 영향을 미쳤으며 이 방식은 시간 소모적이라는 문제점이 있었다. 또한 검출 대상을 지정하는 시스템이 유연하지 못했기 때문에 새로운 랜드마크를 추가하는 것에 어려움이 있었다. 지식 기반 시스템의 경우 숙련된 교정 전문의의 시간과 노력을 필요로 한다. 특히 동일한 물체에 대해 다양한 각도에서 획득된 이미지마다 매번 수동으로 랜드마크 라벨링을 수행해야 했기 때문에 많은 비용이 들고 재현의 어려움 등의 한계점들이 존재했다.

위와 같은 제한들은 이미지에서 특성맵을 추출하는 필터까지도 스스로 학습하여 객체 탐지, 얼굴 인식, 자세 추정 등 이미지에서 패턴을 찾아내는 분야에 유용하게 활용 가능한 합성곱 신경망(convolutional neural networks)의 발전으로 극복될 수 있었다. 합성곱 신경망을 활용하면 역전파를 통해 계층적인 단계를 거쳐 이미지를 높은 수준으로 추상화하며 학습할 수 있다는 장점을 갖는다. 최근에는 U자 형태의 YOLO(You Only Look Once)를 기반으로 하는 등 합성곱 신경망의 종단 간(end-to-end) 학습 방식을 의료 영상에 활용하여 랜드마크를 더욱 효율적으로 검출하고자 하는 연구들이 제안된 바 있다[3]. 이때 적절한 학습 데이터를 사용할 경우 학습 시간 또한 절약 할 수 있다. 계산 비용을 줄이기 위해 지도학습에서 확률론적 해석으로 적절한 학습 데이터를 수집하는 방법과 심층 강화 학습을 통해 차원을 줄여 학습 속도를 감소 시키는 방법 등이 연구 되고 있다[4, 5]. 그러나 의료 영상의 경우에는 자연 영상과 달리 학습에 사용될 수 있는 데이터를 수집하기가 어렵고 데이터의 절대적인 양이 부족하기 때문에 기존의 우수한 모델들에서도 여전히 제한된 정확도를 보인다는 한계점이 있었다. 아래의 Figure 1은 본 연구에서 제안하는 방법의 개요도를 나타낸다. 본 연구에서는 이러한 한계점을 극복하기 위해 기하학적 데이터 증강 기법을 수행한다. 또한 최근 검출 부분 연구들에서 우수한 성능의 SOTA(State-Of-The-Art) 모델에 다수 활용되었던 ResNet을 기반으로 다양한 종류의 어텐션을 적용한 뒤 실험 결과를 통해 개선된 검출 성능을 제시한다.

2. 관련 연구

2.1 특징점 검출 관련 연구

랜드마크 검출과 관련된 연구들은 대부분 얼굴 특징점 검출(facial keypoint detection) 또는 사람의 신체 관절점 예측(human pose estimation) 문제를 위주로 발전되어 왔다. 기존에 연구되어 온 얼굴 특징점 검출 기법은 크게 회귀(regression) 기반의 접근 방식과 딥러닝 기반 접근 방식으로 나뉜다. 회귀 기반 검출 기법 중 하나인 직접 회귀(direct regression) 방식은 입력 영상에 대한 랜드마크 좌표(X, Y)를 직접 매핑하여 한번에 예측하는 방식을 말한다. 관련 연구로 Wing Loss는 직접 회귀 기법을 사용하여 특징점 검출에 최적화 된 새로운 손실 함수를 제안하였다[6]. 제안된 Wing Loss 손실 함수는 두 가지 손실 함수가 결합된 것으로, 랜드마크의 분산 정도에 따라 둘 중 적합한 손실 함수를 선택하여 사용하는 방식을 활용한다. 이를 통해 기존에 랜드마크 검출 분야에서 널리 사용되던 L2 손실 함수와 비교해 보았을 때 AFLW 데이터셋에서 향상된 랜드마크 검출 성능이 확인되었다. 또 다른 직접 회귀 방식인 DAG(Deep Adaptive Graph)는 보간법을 통해 랜드마크들을 하나의 그래프 G로 표현한 뒤 GCN(Graph Convolutional Network)을 통과하며 학습하는 방식을 제안한 바 있다[7]. 훈련 셋에 대한 평균 가중치로 계산된 초기 그래프는 먼저 GCN-Global을 거치며 이때 초기 그래프에 대한 투시 변환 예측이 수행된다. 이후 DAG 모델은 정확한 예측 그래프를 도출하기 위해 GCN-Local을 통과하며 각 랜드마크의 오프셋을 예측한다. 그러나 이러한 직접 회귀 기반의 방식들은 점진적으로 정답값을 찾아나가는 것이 아니라 단 한번의 예측만으로 결과를 도출하기 때문에 검출의 정확도가 낮다는 한계점을 갖는다. 또 다른 회귀 기반 검출 기법인 계단식 회귀(cascade regression) 방식은 랜드마크의 위치를 예측하기 위해 다양한 회귀 함수들을 활용하여 예측할 랜드마크 좌표의 정답값을 단계적으로 갱신해나간다. Q.Liu 등은 PCSR(Pose-based Cascade Shape Regression)회귀 모델을 제안하여 얼굴 포즈 변형으로 인해 발생하는 랜드마크의 분산에 강력히 반응할 수 있는 모델에 관한 연구를 진행하였다[8]. 그러나 계단식 회귀 기법은 입력 영상의 국소적(local) 특성에만 의존하며 랜드마크 주변의 영역만이 특성 추출기를 통과하게 되기 때문에 결과적으로 랜드마크들 간의 관계에 대한 학습은 간과된다는 단점이 있다. 또한 계단식 회귀 모델들은 수작업(hand-crafted)으로 추출된 특징을 사용하기 때문에 합성곱 신경망 등의 딥러닝 방식보다 매우 시간 소모적이며 종단 간 학습을 위한 하나의 구조로 통합하기 어렵다는 한계가 있다.

최근에는 딥러닝 기법의 발전과 회귀 방식의 한계점에 착안하여 합성곱 신경망에 기반한 검출 방식들이 다수 연구되고 있는 추세이다. 합성곱 신경망을 활용하면 회귀 방식에서 보다 더욱 차별적인(discriminative) 정보들을 학습할 수 있다. CPM(Convolutional Pose Machine)은 히트맵(heatmap) 기반으로 합성곱 신경망에서 각 단계의 필터가 한번에 학습할 수 있는 수

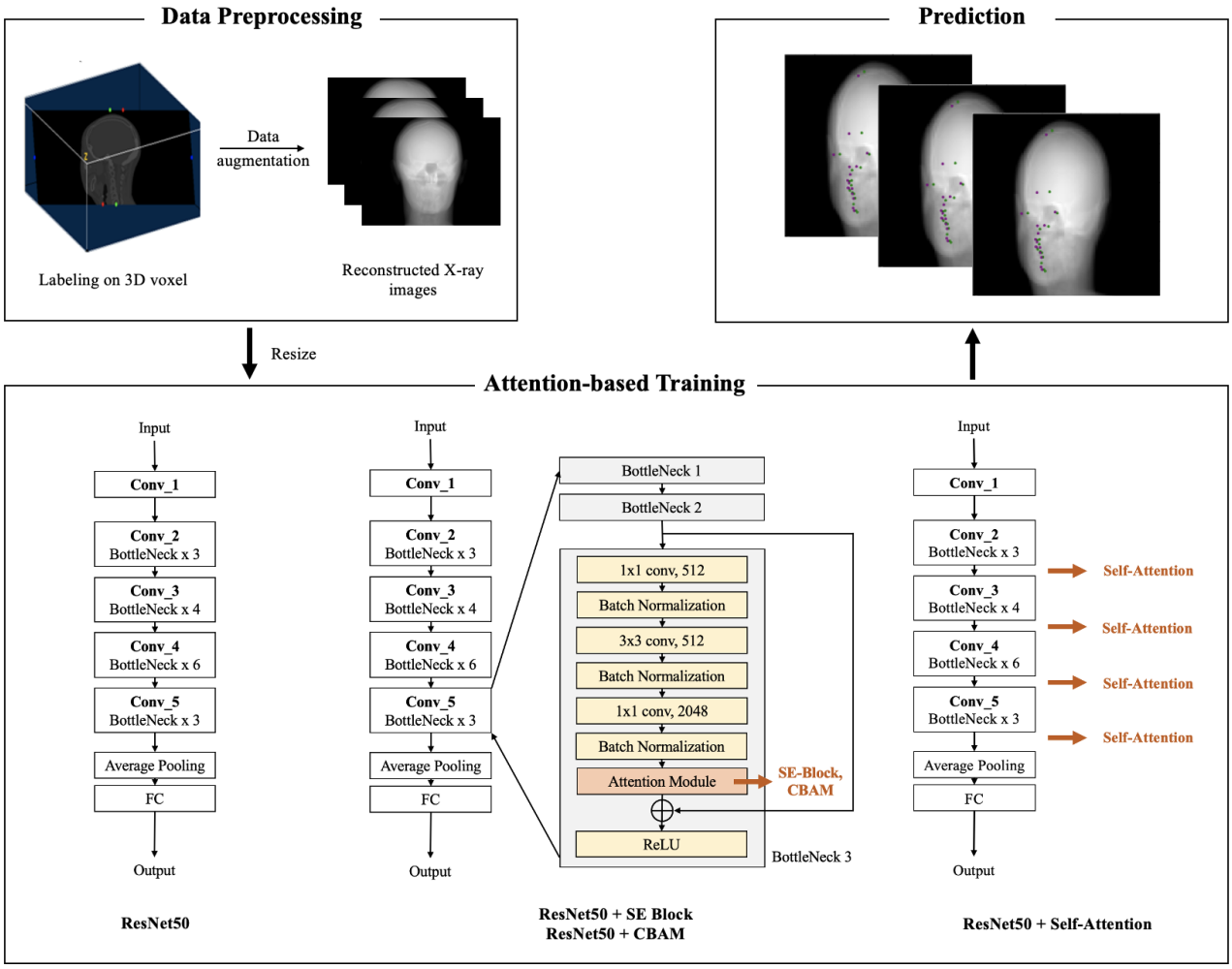


Figure 1: Overall process of the proposed framework. SE-block and CBAM modules are inserted into the position prior to the last ReLU activation function in ResNet50. Since the Self-Attention layer could be used in subsequent to every convolutional blocks in ResNet50, the best position for Self-Attention layer was determined through experiments on the entire number of possible cases.

용 영역(receptive field)을 국소적인 영역에서 전역적인 영역으로 확장해나가며 랜드마크의 예상 좌표를 서서히 좁혀나가는 검출 방식을 제안하였다[9]. CPM의 가장 큰 특징은 입력 영상 내 각 픽셀들에 랜드마크가 위치할 확률이 가우시안 분포(gaussian distribution)로 표현된 히트맵, 혹은 신뢰맵(belief map)을 활용한다는 점이다. 여러 단계(multi-stage)로 이뤄진 CPM은 각 단계마다 형성된 신뢰맵을 다음 단계로 전달하며 학습을 수행한다. 이 과정에서 층이 깊어질수록 수용 영역을 넓혀가며 특정 랜드마크에 대한 신뢰맵 형성 시 주변의 다른 랜드마크들에 대한 공간 정보를 함께 참조하여 더욱 정교한 신뢰맵을 얻는 방식으로 기존의 회귀 모델보다 안정적인 성능을 보였다. 또 다른 히트맵 기반 합성곱 신경망 검출 모델인 HRNet(High-Resolution Network)은 층이 거듭될수록 해상도가 낮아지는 합성곱 신경망의 한계점을 해결하여 더욱 정확한 랜드마크 좌표를 검출할 수 있는 방법을 개발하였다[10]. HRNet은 높은 해상도를 전체 학습에 걸쳐 유지하기 위해 다양한 해상도 별 서브 네트워크(sub network)들을 병렬적인 구조로 활용하며, 결과적으로 높은 해상도로 도출된 히트맵을 통해 랜드마크를 검출하는 기법으로 COCO 데이터셋에서 SOTA의 성능을 보였다.

2.2 의료 영상에서 랜드마크 검출 관련 연구

의료 영상에서 합성곱 신경망을 활용한 랜드마크 검출의 초기 연구 중 하나로, H.Lee 등은 두개골 엑스레이 영상에서 총 19개의 랜드마크들을 검출하기 위해 랜드마크 별 X, Y 좌표를 각각 독립적인 개별 변수로 여기는 접근 방식을 사용하였다[11]. 이로써 랜드마크 검출 문제를 개별 변수로 표현된 좌표값 예측 문제로 재구성한 것이다. 또한 각 좌표 변수마다 하나의 개별적인 합성곱 신경망을 통과하도록 하여 총 38개의 합성곱 신경망의 학습 결과가 종합되어 최종적인 예측을 수행하도록 하였다. 제안되었던 방식은 랜덤 포레스트 등의 전통적인 머신러닝 기법으로부터 벗어나 심층 학습을 활용한다는 점에서 이전 연구들과의 차별성을 가졌지만 얕은 모델 구성으로 인해 안정적인 학습 성능에 이르지 못했다는 한계점을 남겼다. 이후에 B.Bier 등은 골반 엑스레이 영상에서 23개의 해부학적 랜드마크를 자동으로 검출하기 위해

CPM의 아이디어로부터 발전시킨 네트워크를 제안하였다[12]. 해당 연구에서는 엑스레이 영상이 획득되는 방향과 무관하게 높은 검출 정확도를 도출해내기 위해 히트맵에 기반한 순차적인 예측 방식을 전개하였다. 이 기법은 CPM이 갖던 장점과 마찬가지로 랜드마크의 예상 위치를 점차 좁혀나가며 점진적인 과정을 통해 예측 결과를 도출해내므로 직접 회귀 방식이나 얕은 합성곱 신경망이 사용된 경우보다 더욱 정교하고 안정적인 결과에 도달할 수 있다는 장점을 가졌다. 이 외에도 엑스레이 영상에서 관심영역 ROI(Region Of Interest)를 추출하여 패치 기반의 학습 방식을 제안한 연구 사례가 있다. Y.Song 등은 두개골 엑스레이 영상에서 영상의 스케일이 너무 큰 경우 여러 개의 랜드마크를 단번에 검출하는 것에 무리가 된다고 판단하였다. 따라서 랜드마크별로 패치를 생성한 뒤 작아진 입력 영상에 대한 랜드마크 검출을 시도하였다[13]. 이 과정에서 데이터 증강을 위해 랜드마크 별로 총 400장의 크롭된 ROI 패치를 생성하였으며, 패치 기반 학습을 통해 ISBI Challenge 데이터 셋에서의 테스트 결과 2.5mm를 기준으로 하는 SDR(Successful Detection Rate) 측면에서 91.7%의 우수한 성능을 보였다.

위와 같은 관련 연구 사례들에서도 언급되었듯이, 의료 영상에서의 랜드마크 검출 문제는 학습에 사용할 수 있는 데이터의 양이 적다는 근본적인 제한점을 갖는다. 따라서 본 연구에서는 의료 영상의 특성을 고려하여 랜드마크 라벨링의 효율성을 높일 수 있는 기하학적 데이터 증강 기법을 소개한다. 또한 랜드마크 검출 성능을 향상시키기 위해 최근 컴퓨터비전 분야에서 다양하게 응용되고 있는 합성곱 신경망 기반 어텐션 모델들을 구성한다. 마지막으로 최종적인 실험 결과를 제시하여 기하학적으로 증강된 데이터 셋에 대한 각 어텐션 모델들 간의 성능 차이를 비교하고 예측에 영향을 미친 요인을 분석한다.

3. 데이터

3.1 XCAT 팬텀 데이터

실험에 사용된 데이터는 두개골 XCAT 팬텀(phantom) 엑스레이 영상이다[14]. 해당 데이터 셋의 구성은 13명의 남성과 15명의 여성으로 이루어져 있었으며, 표준 남성 및 여성 성인의 상세한 두개골 구조 정보를 포함하는 총 28개의 version 7 매트랩 배열(matlab array) 형태로 제공되었다. 모델에 입력으로 주어질 데이터 셋을 구축하기 위해 라벨링과 투영을 통한 기하학적 데이터 증강 과정을 거쳐 모든 랜드마크의 좌표 정보가 포함된 데이터 셋을 구축한다.

3.2 라벨링 (Labeling)

랜드마크를 라벨링하는 과정은 초기에 데이터가 제공된 형태인 매트랩 배열에서 수행될 수 없다. 따라서 라벨링을 위해 데이터를 800×800×200에서 800×800×250 사이의 크기를 갖는 3차원 복셀(voxel)로 변환하여 사용하였다. 검출의 대상이 될 단일 랜드마크는 총 16개의 지점을 채택하였다. 랜드마크 선정의 기준으로는 IEEE ISBI 2015 Challenge에서 사용되었던 19개의 랜드마크와 표준 방사선학 용어(radiology terminology)에서 해부학적으로 의미가 있다고 알려진 주요 랜드마크들을 우선적으로 고려하여 선택하였다[15]. 각 랜드마크를 지칭하는 용어는 대한의사협회와 대한해부학회, 그리고 자연어 처리에 주로 활용되는 어휘 데이터베이스인 WordNet 등에서 제공하는 의학용어사전을 기준으로 통일하였다. Table 1은 본 연구에서 선정하여 라벨링 한 16개 랜드마크들에 부여한 랜드마크 인덱스와 명칭, 그리고 각각에 대한 정의를 나타낸다. 선정된 랜드마크들은 Menton, Gnathion, Pogonia, Bpoint, Lowerlip, Infradentale, Lower incisal_incision, Upper incisal_incision, Upper lip, Subnasale, ANS, PNS, Orbitale(left), Orbitale(right), Nasion, 그리고 Bregma이다. 라벨링을 통해 3차원 복셀 인덱스로부터 획득된 (X, Y, Z) 좌표는 이후 투영(projection) 단계에서 활용된다.

Table 1: Description of 16 landmarks used in the experiments.

랜드마크 (Landmark)		정의 (Definition)
L1	**Menton**	하악골의 가장 아래쪽 지점
L2	**Gnathion**	하악골 턱선 아래쪽 테두리의 중간 지점
L3	**Pogonion**	턱의 앞쪽 중간 지점
L4	**Bpoint**	후방 중심 지점, Pogonion과 Infradentale 사이 하악골
L5	**Lower lip**	아래 입술의 중앙 지점
L6	**Infradentale**	아래턱의 두 중앙 앞니 사이에 있는 잇몸의 가장 높은 지점
L7	**Lower incisal_incision**	순측 상악 앞니 가장자리 중심 지점
L8	**Upper incisal_incision**	순측 하악 앞니 가장자리 중심 지점
L9	**Upper lip**	윗 입술의 중앙 지점
L10	**Subnasale**	비중격이 중간 시상면에서 윗입술과 합쳐지는 지점
L11	**ANS**	두개 상악골이 융합돼 형성된 돌출부
L12	**PNS**	연구개 구개뼈 뒤쪽 척추의 정점
L13	**Orbitale (left)**	눈이 위치하는 안와의 가장 낮은 왼쪽 뼈
L14	**Orbitale (right)**	눈이 위치하는 안와의 가장 낮은 오른쪽 뼈
L15	**Nasion**	코뼈와 이마뼈 사이 교차점의 앞쪽 지점
L16	**Bregma**	두개골 관상 봉합사와 시상 봉합사의 접합 지점

3.3 투영(Projection)을 통한 기하학적 데이터 증강

카메라의 최종적인 영상은 3차원 공간 상의 좌표들을 2차원 공간으로 투사하는 원리를 통해 획득된다. 합성곱 신경망의 입력 데이터가 될 2차원의 엑스레이 영상을 생성하는 과정도 이와 같다. 먼저 3차원 복셀 상에서 라벨링 된 (X, Y, Z) 좌표를 2차원 좌표로 변환한 뒤, 데이터 증강 및 투영을 위한 카메라 행렬(camera matrix)을 도출하는 것이다. 이를 위해 일반적인 핀홀 카메라 모델에서 차원 간 좌표 변환 시 사용되는 관계식을 활용하였으며 이는 수식 1와 같다.

$$P = \begin{bmatrix} f & v & P_x \\ 0 & f & P_y \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} r_1 & r_2 & r_3 & \vdots & t_1 \\ r_4 & r_5 & r_6 & \vdots & t_2 \\ r_7 & r_8 & r_9 & \vdots & t_3 \end{bmatrix} = [\textit{camera matrix}] \tag{1}$$

투영 과정에서는 의료 영상 처리를 위한 투영 및 영상 재건 알고리즘 시뮬레이션 기능들을 제공하는 CONRAD 소프트웨어를 이용하였다[16]. 카메라의 원리에서 내부 매개변수(intrinsic parameter)는 초점 거리(focal length), 이미지 센서 포맷(image sensor format), 주점 거리(principal point), 그리고 x축과 y축의 뒤틀림 계수(skew coefficient) 등의 정보를 포함하며 외부 매개변수(extrinsic parameter)는 3차원과 2차원의 영상에서의 좌표 매칭 쌍 등의 정보를 포함한다. 본 연구에서는 이러한 내부 매개변수와 외부 매개변수 간의 관계를 활용하여 카메라 투영 행렬(camera projection matrix)을 도출하였다. 이렇게 획득한 투영 행렬 값에 대해 CONRAD에서 전방 투영(forward projection)을 적용하여 결과적인 2차원 좌표값을 얻었다. 투영 행렬을 이용하여 동차적(homogeneous)인 2차원 상의 (u, v) 좌표값을 얻기 위한 수학적 관계식은 수식 2와 같이 표현된다. Figure 2는 전방 투영을 통해 획득된 각도별 2차원 엑스레이 영상과 랜드마크 정답값을 나타낸다.

$$\begin{bmatrix} \tilde{u}/\varepsilon \\ \tilde{v}/\varepsilon \\ \varepsilon \end{bmatrix} = [\textit{camera matrix}] \cdot \begin{bmatrix} x \\ y \\ z \\ 1 \end{bmatrix} \tag{2}$$

4. 어텐션(Attention) 모델

실험에서는 기존의 랜드마크 검출 연구들에서 우수한 성능의 백본(backbone) 모델로 수차례 언급된 ResNet50을 사용하였다[17]. 2015년도에 IILSVRC에서 발표되었던 ResNet은 잔차 블록(residual block)에서 지름길(skip connection)을 통해 모듈의 출력 값에 입력 값을 더함으로써 기울기 소실(gradient vanishing) 문제를 해결한다는 장점을 갖는 대표적인 합성곱 신경망 모델이다. ResNet은 가중치 층이 쌓인 깊이에 따라 ResNet18부터 ResNet152까지 다양한 변형된 모델들을 갖는데, 본 연구에서는 이 중 두개골 XCAT 데이터에 대해 가장 안정적인 학습 성능을 보

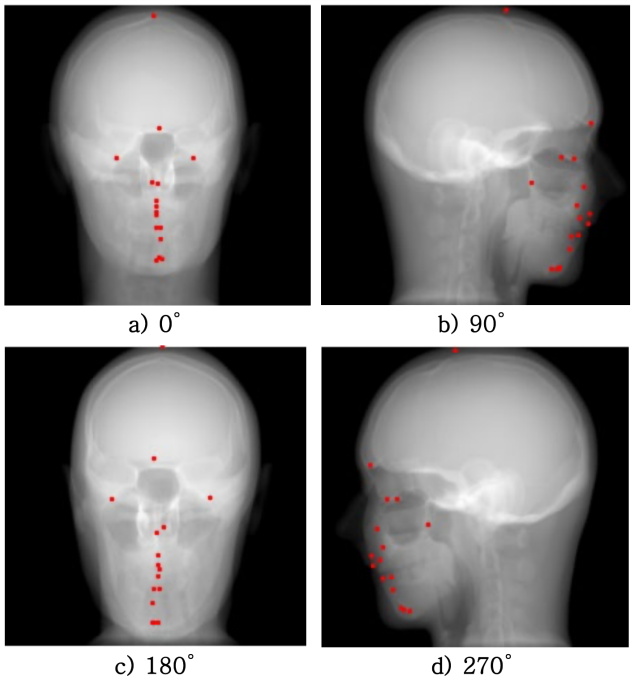


Figure 2: X-ray images and landmark labels identified after forward projection. Red dots are the label values for each of 16 landmarks. a), b), c), and d) represent images and labels viewed at 0 degrees, 90 degrees, 180 degrees, and 270 degrees, respectively.

인 ResNet50을 선택하여 사용하였다. 또한 변형이 가해지지 않은 기본 ResNet50 이외에도 검출 성능의 향상을 위해 다음으로 소개할 어텐션 모듈들이 사용된 어텐션 모델들을 추가로 구성하여 실험하였다. 본 연구에서 사용된 어텐션 기법은 크게 채널 어텐션(channel attention)과 공간 어텐션(spatial, or pixel-wise attention), 그리고 셀프 어텐션(self-attention)으로 나뉜다.

4.1 SE-block

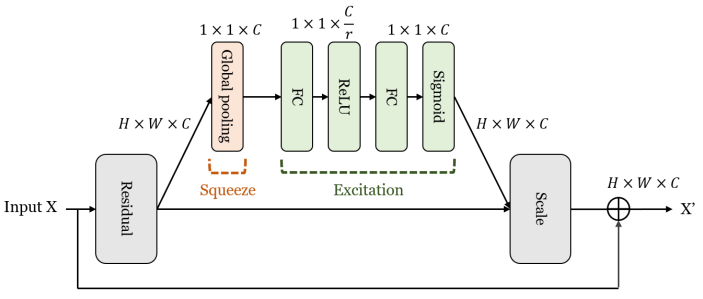


Figure 3: Structure of the SE-block[18].

SENet(Squeeze-and-Excitation Network, SENet)은 채널 간의 관계에 집중도를 부여하며 어텐션을 계산하는 채널 어텐션 기법이다[18]. SE-block은 채널 별 가중치를 계산한 뒤 가중치가 높은 채널은 다른 채널들보다 유의미한 특성을 포함할 가능성이 높으므로 더 큰 집중도를 부여하는 방식으로 학습한다. SE-block이 채널 가중치를 추출하는 단계는 크게 압축(squeeze)과 재조정

(excitation)으로 나뉜다. 먼저 합성곱 층을 통과한 특성맵의 각 채널은 global average pooling 연산을 통해 H×W의 크기를 갖는 1×1 크기의 벡터로 압축된다. 이렇게 도출된 1×1 크기의 채널 벡터는 재조정 단계의 입력값이 되며, 재조정 과정은 완전연결층(fully-connected layer)과 ReLU 활성화함수, 그리고 완전연결층과 시그모이드(sigmoid) 함수 순서대로 진행된다. 재조정 단계에서는 앞 단계에서 생성된 각 채널 별 1차원 벡터에 대한 정규화를 수행한 뒤 구체적인 가중치를 부여한다. SE-block의 동작 구조는 Figure 3과 같다. 본 연구에서는 ResNet50의 합성곱 블록(convolutional block) 뒤의 ReLU 활성화 함수가 적용되기 이전 위치에 SE-block을 삽입함으로써 첫 번째 어텐션 모델을 구성하였다.

4.2 CBAM

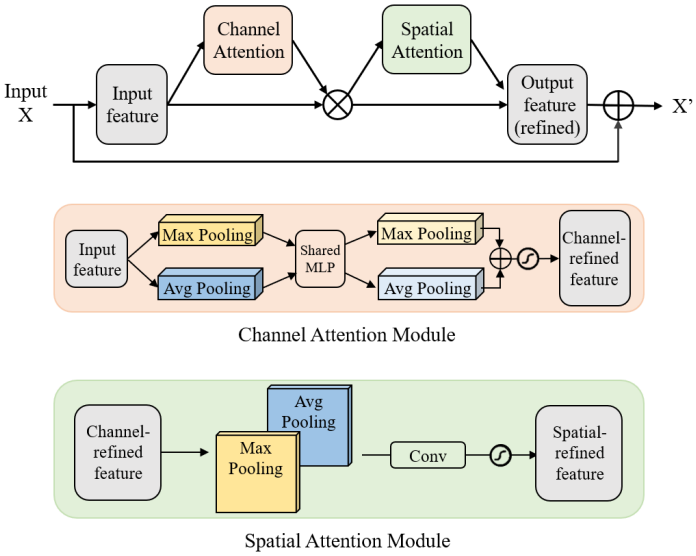


Figure 4: Structure of the CBAM module[19].

CBAM(Convolutional Block Attention Module)은 BAM으로부터 발전된 어텐션 기법이다[19]. CBAM은 SE-block과 같이 채널 간의 상호작용 및 중요도를 고려하는 채널 어텐션뿐만 아니라 입력 이미지에서 픽셀 공간에 대한 정보도 함께 고려하는 공간 어텐션을 동시에 수행한다. 또한 이 두 가지 어텐션 연산을 병렬적이 아니라 각각 순차적으로 수행한다는 점에서 앞서 발표되었던 BAM과의 차이점을 갖는다. CBAM에서는 채널 어텐션을 위해 average pooling과 max pooling의 두 가지 풀링 기법을 모두 사용하며 각각의 결과값을 shared MLP에 통과시킨 뒤 시그모이드 함수를 거쳐 최종적인 채널 중요도를 계산하는 구조를 갖는다. 공간 어텐션을 위하여는 max pooling과 average pooling을 순차적으로 이어 붙이는 방식을 통해 더욱 집중해서 학습할 중요도가 높은 픽셀의 공간 정보를 추출한다. CBAM 모듈의 구조는 Figure 4과 같다. 본 연구에서는 SE-block이 적용된 위치와 마찬가지로 ResNet50이 갖는 합성곱 블록의 ReLU 함수 이전에 CBAM 모듈을 추가하여 두 번째 어텐션 모델을 구현하였다.

4.3 Self-Attention

셀프 어텐션(Self-Attention)은 합성곱 신경망의 구성 요소 중 하나인 합성곱 연산 층에서 국소적인 수용 영역으로 인해 층이 점차 깊어질수록 발생하기 쉬운 장기 지속성 문제(long range dependency)를 해결하기 위한 방법으로 제안되었다[20]. 셀프 어텐션에서 중요도를 계산하는 방식은 수식 3과 같이 표현된다. 셀프 어텐션은 기존의 합성곱 신경망 모델들이 필터를 학습하던 방식과 달리, 총 세 개의 가중치를 학습하여 각각의 가중치를 통해 선형 변환(linear transform)을 수행하여 key, query, 그리고 value 값을 도출한다. 먼저 key와 query는 입력 영상에 대해 각각의 학습된 가중치가 곱하여져 계산된다. 이렇게 도출된 key와 query는 서로 곱해진 뒤 소프트맥스(softmax) 함수를 통과하며 0과 1 사이의 확률값으로 표현된다. 이후에 소프트맥스 함수의 출력값은 입력 영상에 대한 가중치 곱을 통해 얻어진 value와 행렬 요소 단위의 곱셈(element-wise multiplication)을 거치게 되며, 최종적으로 1×1의 크기를 갖는 결과값이 출력된다.

$$y_{ij} = \sum_{a,b \in N_k(i,j)} Softmax_{ab}(q_{ij}^T k_{ab}) v_{ab} \quad (3)$$

그러나 이와 같은 기본적인 key, query, value의 연산만을 수행한다면 입력 영상에서 각 픽셀의 위치 정보들이 소실되는 문제점이 발생하는데, 셀프 어텐션에서는 positional encoding을 통해 이러한 문제점을 해결할 수 있었다. 의료 영상에 대한 랜드마크 검출 문제에서는 영상에서 각 픽셀들이 갖는 상대적인 위치 정보가 매우 중요한 요소로 작용한다. 이러한 픽셀 간 상대적 위치 정보의 손실을 최소화하기 위해 셀프 어텐션에서는 각 픽셀 간의 행과 열에 대한 거리 관계를 정의한 뒤, 모든 픽셀로부터의 가능한 행 오프셋(row offset)과 열 오프셋(column offset)의 조합을 표의 형태로 표현하여 위치 정보를 보존한 채 학습을 수행한다. ResNet50을 기준으로 하였을 때 셀프 어텐션이 삽입될 수 있는 위치는 각 합성곱 블록의 위치이므로 총 네 군데가 집계된다. 따라서 본 연구의 실험에서는 구축된 데이터 셋에 대해 가장 효과적인 성능을 보인 셀프 어텐션 모델을 구하기 위해 셀프 어텐션이 추가될 수 있는 다양한 위치별 실험 결과를 통해 가장 효과적인 모델 구조를 발견하고자 하였다.

5. 실험 방법

5.1 실험 계획 및 환경

모델 학습을 위한 입력 데이터 셋은 라벨링 된 2차원의 좌표값과 620×480 픽셀 크기의 엑스레이 영상을 사용하였다. 엑스레이 영상은 한 명의 환자 당 1° 단위로 투영시킨 360장의 각도 별 2차원 전방 투영의 결과이다. 28명의 데이터를 통해 총 10,080장의 영상을 입력 데이터로 사용하였으며, 모든 영상은 96×96 픽셀 크기의 보간법을 이용해 리사이즈(resize)한 후 모델에 입력하였

다. 학습 시에는 각 특성(feature) 별 가중치를 두지 않고 16개의 독립적인 랜드마크를 학습하도록 하였다. 학습 및 검증에는 전체 28명의 데이터 셋 중 26명의 데이터를 사용하였다. 훈련 셋과 검증 셋 간 비율은 8:2로 분리하였으며 각 손실의 수렴 정도를 파악하여 과적합 여부를 확인하였다. 학습에 사용되지 않은 2명의 데이터는 최종적인 성능 평가 시에만 사용될 수 있도록 학습으로부터 분리하여 활용하였으며, Table 2는 본 연구에서 활용된 데이터 셋의 정보를 설명한다.

학습을 위한 하이퍼파라미터(hyper-parameter)는 다음과 같이 설정하였다. 배치 사이즈(batch size)는 32, 에폭(epoch)은 100으로 지정하였으며 학습률(learning rate)은 1e-3의 크기로 실험이 진행되었다. 또한 MSE(Mean Squared Error) 손실함수와 Adam 옵티마이저가 사용되었다. 전체 실험은 CUDA 11.2 버전의 GPU 환경에서 파이썬(python) 3.8 및 파이토치(pytorch) 1.9 라이브러리를 사용해 진행되었으며, 모든 실험은 랜덤시드(random seed)를 고정하여 조건에 따라 같은 결과가 나올 수 있도록 학습의 재현성을 확보하였다.

Table 2: Description of the skull XCAT dataset.

Dataset	Num of patients	Num of images	Gender
Train/Val	26	9,360	12 males 14 females
Test	2	720	1 male 1 female
Total	28	10,080	13 males 15 females

5.2 평가 방법

성능 평가를 위해 MRE(Mean Radial Error)와 SD(Standard Deviation), 그리고 SDR(Successful Detection Rate)의 총 세 가지 평가 지표를 활용하였다. MRE 지표의 경우 예측된 좌표의 기하학적 방사상 오차를 나타내며 MRE 계산 시 정답 좌표값과 예측된 좌표값 간의 거리를 계산하기 위해 아래의 수식 4 을 활용하였다.

$$R = \sqrt{\Delta x^2 + \Delta y^2} \tag{4}$$

이때 Δx 와 Δy 는 2차원 좌표계에서 정답인 좌표(x1, y1)와 예측된 좌표(x2, y2)에서 x축과 y축 각각의 절대값의 차이로 계산된 두 좌표 간의 거리이자 검출 오차를 말한다. 이를 통해 MRE와 그에 대한 분산을 의미하는 SD를 계산하는 방식은 아래의 수식 5, 6과 같이 표현된다.

$$MRE = \frac{\sum_{i=1}^{N} R_i}{N} \tag{5}$$

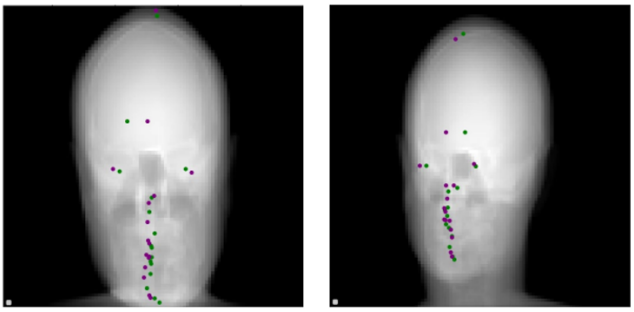
$$SD = \sqrt{\frac{\sum_{i=1}^{N}(R_i - MRE)^2}{N-1}} \tag{6}$$

직관적인 성공률을 통계적 지표로 보이기 위해 SDR을 활용하여 2mm에서 6mm의 범위 내에서의 성공적인 검출율을 계산한다. 각 랜드마크에 대해 예측된 좌표와 정답 좌표 간의 거리 차가 z 이하이면 검출에 성공한 것으로 간주하고 그렇지 않으면 오검출로 간주한다. z의 거리별 정확도에 해당하는 SDR은 백분율로 수식화 할 수 있다. 성능 평가 시 z값은 각각 2mm, 4mm, 그리고 6mm 측면에서 계산하였다. SDR은 수식 7과 같이 계산된다.

$$SDR_{(z)} = \frac{number\ of\ accurate\ detections}{number\ of\ total\ detections} \times 100\% \tag{7}$$

6. 실험 결과

6.1 모델 별 실험 결과



a) 175° a) 338°

Figure 5: a) and b) visualize the detection results derived from ResNet50 at 175 degrees and 338 degrees, respectively. (green dots: predicted values, purple dots: label values)

Figure 5는 ResNet50을 이용한 예측 결과를 시각화 한 결과이다. 한 명의 환자 데이터에서 대해 서로 다른 두 각도 별 모든 랜드마크의 예측 좌표가 표시되어 있으며, 실제 좌표와 예측된 좌표 간 차이를 시각적으로 확인할 수 있었다.

셀프 어텐션이 추가된 개수와 위치를 달리하여 실험한 16개의 모델 별 결과는 하단의 Table 3과 같으며, ResNet50이 갖는 각 층인 conv_2, conv_3, conv_4, conv_5에 대한 셀프 어텐션의 적용 여부는 T(True), 혹은 F(False)로 표기하였다. 셀프 어텐션 실험에서 가장 우수한 성능을 보인 모델은 총 세 개의 어텐션(T, T, T, F)이 추가된 경우로, MRE와 SDR(6mm) 측면에서 각각 2.9997mm와 94.6788%로 가장 높은 성능을 보였다. 이 경우 SD 측면에서도 3.8673mm로 두 번째로 높은 값을 보이며 안정적인 학습 성능을 나타냈다. 본 실험의 평가 지표에서는 MRE가 낮을수록 더 우수한 성능을 보이는 것으로 간주한다. 기본 ResNet50보다 더 높은 MRE값을 보인 경우는 각각 셀프 어텐션이 (F, F, F, T), (F, T, F, T), (F, F, T, T), 그리고 (F, T, T, T)에 위치한 경우였으며, 이는 모두 conv_2층에 어텐션이 적용되지 않았다는 공통점을 갖는다. 따라서 셀프 어텐션을 모델의 도입부에 적용하는 것이 랜드마크 검출

Table 3: Experimental results on 16 cases of applying Self-Attention to ResNet50.

Num of Self-Attention	Position of Self-Attention (conv_2, conv_3, conv_4, conv_5)	MRE (mm)	SD (mm)	SDR (%)		
				(2mm)	(4mm)	(6mm)
1	T, F, F, F	3.4038	6.2086	39.6441	**83.2726**	94.5052
	F, T, F, F	3.4564	4.3089	32.4653	74.3142	91.7188
	F, F, T, F	3.3054	4.8174	43.3507	77.9948	89.2014
	F, F, F, T	3.6580	5.1508	41.1719	72.4045	84.0885
2	T, T, F, F	3.0957	**3.6377**	42.0486	78.8281	92.1267
	T, F, T, F	3.3378	6.3228	**45.5816**	81.3802	93.9063
	T, F, F, T	3.1830	5.6842	44.2969	80.8594	93.5851
	F, T, T, F	3.5257	5.1822	39.9653	74.1667	89.3576
	F, T, F, T	3.9982	7.1549	39.9826	72.5955	86.6146
	F, F, T, T	3.6239	6.6797	38.4288	75.1042	91.7448
3	T, T, T, F	**2.9997**	3.8673	40.7986	81.4063	**94.6788**
	T, T, F, T	3.5314	6.3551	42.9688	77.9774	91.8142
	T, F, T, T	3.5173	6.3888	37.2917	80.0174	92.2830
	F, T, T, T	4.0847	5.7149	30.1736	68.2639	86.3715
4	T, T, T, T	3.2849	4.5663	31.7101	81.4497	94.0625
Self-Attention average	-	3.4671	5.4693	39.3252	77.3356	91.0706

∗ MRE: Mean Radial Error, SD: Standard Deviation, SDR: Successful Detection Rate

Table 4: Experimental results comparing ResNet50 with the other three attention methods.

Model	Position of Self-Attention (conv_2, conv_3, conv_4, conv_5)	MRE (mm)	SD (mm)	SDR (%)		
				(2mm)	(4mm)	(6mm)
Plain ResNet50	-	3.5367	3.7111	30.9288	73.6024	89.9653
ResNet50 + SE-block	-	**2.6965**	**3.4437**	**51.2326**	**83.2118**	93.5938
ResNet50 + CBAM	-	3.9340	6.0625	43.9410	72.0486	83.0295
ResNet50 + Self-Attention	T, T, T, F	2.9997	3.8673	40.7986	81.4063	**94.6788**

∗ MRE: Mean Radial Error, SD: Standard Deviation, SDR: Successful Detection Rate

성능 향상에 효과적인 요인으로 작용할 수 있음이 파악된다. 16개의 셀프 어텐션 모델들에서 MRE의 평균은 3.4671mm로 기본 ResNet50 모델보다 약 0.0696mm로 소폭 향상된 결과를 보였다. 또한 2mm 기준의 SDR의 경우, 셀프 어텐션이 세 번 사용된 (F, T, T, T)의 경우를 제외하고는 기본 ResNet50보다 최소 0.7813%에서 최대 14.6528%의 향상이 있었으며, 평균적으로는 8.3964% 성능이 향상되었다.

Table 4는 기본 ResNet50과 어텐션 모델들의 실험 결과를 나타낸다. 셀프 어텐션의 경우, 가장 우수한 성능을 보였던 (T, T, T, F)를 선택하여 기재하였다. 셀프 어텐션의 실험 결과로는 Table 3의 결과에서 가장 우수한 성능을 보였던 (T, T, T, F)의 경우를 선택하여 기재하였다. SE-block과 셀프 어텐션이 사용된 경우에는 전반적인 실험 결과에서 기본 ResNet50에서보다 기하학적 평가 지표인 MRE와 통계적 평가 지표인 SDR이 모두 향상되었다. 전체 네 개의 모델 중 가상 우수한 결과를 보인 것은 SE-block을 사용한 경우였으며 이 경우의 성능은 MRE와 SD, 2mm와 4mm 기준 SDR 측면에서 각각 2.6965mm, 3.4437mm, 51.2326% 83.2118%로 가장 높은 값을 보였다. 2mm 기준의 SDR 지표에 정답값으로 포함되는 경우는 임상적으로도 의미가 있다고 여겨진다[21]. SE-block이 적용된 경우 2mm 기준 SDR의 성능이 기본 ResNet50에서보다 20.3038%나 오르며 추가적인 개선을 통한 성능 향상의 가능성을 보였다. 해당 결과는 엑스레이 영상에서 채널 어텐션을 적용하여 랜드마크를 검출할 시 채널 별 가중치에 대한 고려가 우수한 성능 도출에 큰 영향을 미친다는 것을 의미한다. 두 번째로 우수한 성능을 보인 어텐션 모델은 셀프 어텐션을 사용한 경우였으며, 이 경우 6mm 기준 SDR이 94.6788%로 전체 실험 중 최고 성능을 보였다.

이에 반해, CBAM을 사용한 실험 결과에서는 MRE와 SD 등의 기하학적 지표 측면에서 하락된 성능이 확인되었다. 특히 SD값은 6.0625mm으로 크게 하락되었는데, 이는 Gnathion과 Nasion 등 성공적으로 예측하기 어려운 몇몇 랜드마크의 특성이 분산을 단번에 높였기 때문인 것으로 분석되었다.

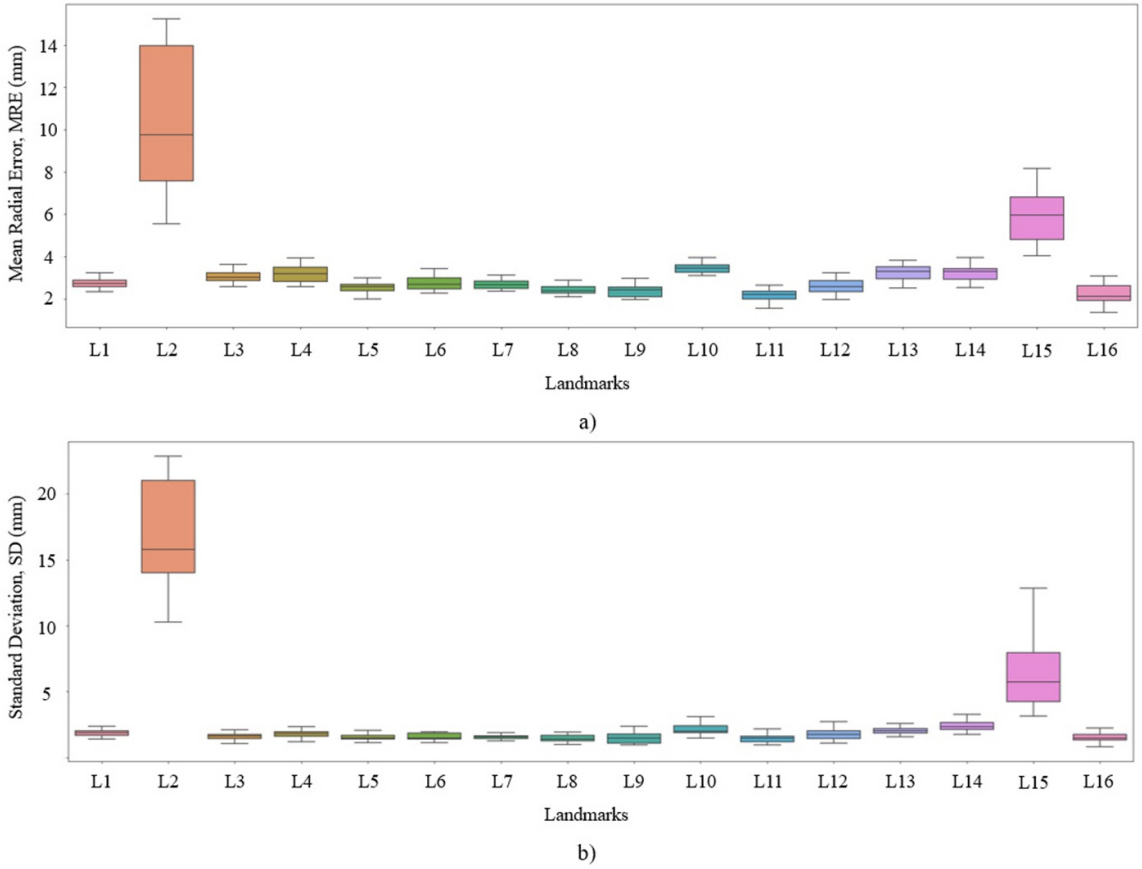


Figure 6: Boxplot visualization of the a) distribution of MRE and the b) distribution of SD in terms of landmarks through the average detection results of 16 model using self-attention.

6.2 랜드마크 별 예측 결과 분석

Figure 6는 Table 3에서 보인 실험 결과의 랜드마크 별 a) MRE와 b) SD의 성능 분포도를 보인다. Gnathion(L2)과 Nasion(L15)의 경우, 평균 MRE는 각각 10.5264mm, 6.1600mm이었으며 박스플롯 시각화를 통해 예측 범주가 넓음을 알 수 있었다. 이외 랜드마크들의 MRE는 3.4714mm 이하의 값을 보였기 때문에 L2, L15와 같은 일부 랜드마크의 특성이 전체의 성능을 하락시키고 있음을 확인할 수 있었다. 하관에 위치한 Gnathion의 검출 성공률이 낮은 이유로는 턱 주변의 연조직이 뚜렷한 돌출 없이 경사져 있으며, 엑스레이 촬영 시 안면 근육의 충분히 이완되지 않아 연조직 변형 및 비정형 연조직의 형태를 초래했기 때문인 것으로 분석된다. 또한 Gnathion은 두개골의 가장자리에 위치하지 않아 모델이 특성을 추출하여 좌표를 검출하는 것에 어려움이 있는 것으로 예상된다. Nasion은 코 뼈와 이마 뼈 사이에 위치하는데, 해당 위치는 해부학적으로 뼈가 중첩되어 있기 때문에 식별에 어려움이 있는 것으로 파악되었다. 또한 Nasion의 위치는 이마와 코의 모양에 따라 사람마다 다르며 정면 곡선이 명확하지 않은 경우에는 전문가가 수동으로 라벨링 할 경우에도 어려움을 겪는 지점이다[22]. 글로벌 로컬 인코더(global local encoder)와 패치 기반 어텐션 등을 적용한 최신의 연구들에서도 Nasion의 랜드마크 지점은 여전히 다른 랜드마크들에 비해 낮은 성능을 보이며 검출에 어려움을 겪었다[23].

7. 결론 및 향후 연구 방향성

본 논문에서는 랜드마크 라벨링의 효율을 높이고 의료 영상의 제한된 데이터 문제를 극복할 수 있는 기하학적 데이터 증강 기법을 활용하여 두개골 XCAT 랜드마크 데이터 셋을 구축하였다. 해당 데이터 셋에 대해 ResNet50과 세 가지의 어텐션 기법들을 적용하여 엑스레이 영상을 학습한 다양한 모델들 간의 랜드마크 검출 성능 차이를 비교 분석해보았다. 실험 결과, SE-block의 채널 어텐션 방식이 적용된 모델이 SDR(6mm) 측면에서 94.6788%의 성능으로 가장 우수함을 보였다. 또한 랜드마크 별 실험 결과를 바탕으로 SD 측면의 지표가 상대적으로 낮은 이유를 파악하였다. 이는 Gnathion과 Nasion 등 얼굴의 하관과 중앙에 위치한 랜드마크의 해부학적 구조의 특성으로 인해 성공적인 검출이 어렵기 때문이다. 그러나 Gnathion과 Nasion을 제외한 기타 14개의 랜드마크와 같은 경우 합성곱 신경망과 어텐션 모듈을 활용하여 안정적인 랜드마크 예측 성능에 도달할 수 있었다. 향후 연구 방향으로는 의료 영상에 대한 딥러닝 모델의 임상적인 응용 가능성 및 실용성을 높이기 위해 의료 영상의 회색조(gray scale) 특성을 고려한 데이터 전처리 방법을 개발할 계획이다. 또한 검출 성공률

이 상대적으로 낮은 특정 랜드마크들로 인한 높은 분산을 낮추기 위한 새로운 데이터 증강 기법 등을 모색할 예정이다. 이를 통해 MRE의 분산이 높은 특정 랜드마크에 집중한 학습을 통해 모델의 전체 성능을 개선해 나갈 수 있을 것으로 보인다.

감사의 글

This work was partly supported by the Technology development Program of MSS [S3146559] and by the Korea Medical Device Development Fund grant funded by the Korea government (the Ministry of Science and ICT, the Ministry of Trade, Industry and Energy, the Ministry of Health & Welfare, the Ministry of Food and Drug Safety) (Project Number: KMDF_PR_20200901_0016, 9991006689)

References

[1] A. Kaur and C. Singh, "Automatic cephalometric landmark detection using zernike moments and template matching," *Signal, Image and Video Processing*, pp. 117–132, 2015.

[2] I. El-Fegh, M. Galhood, M. Sid-Ahmed, and M. Ahmadi, "Automated 2-d cephalometric analysis of x-ray by image registration approach based on least square approximator," in *2008 30th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*, 2008, pp. 3949–3952.

[3] R. D. J.H. Hwang, M.G. Kim, "Evaluation of automated cephalometric analysis based on the latest deep learning method," *The Angle Orthodontist*, pp. 329–335, 2021.

[4] Daseong and Han, "A supervised learning framework for physics-based controllers using stochastic model predictive control," *Journal of the Korea Computer Graphics Society*, pp. 9–17, 2021.

[5] Jewon, T. Ahn, T. Gu, and Kwon., "Motion generation of a single rigid body character using deep reinforcement learning," *Journal of the Korea Computer Graphics Society*, pp. 13–23, 2021.

[6] Z. Feng, J. Kittler, M. Awais, P. Huber, and X. Wu, "Wing loss for robust facial landmark localisation with convolutional neural networks," *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 2235–2245, 2018.

[7] W. Li, Y. Lu, K. Zheng, H. Liao, C. Lin, J. Luo, C.-T. Cheng, J. Xiao, L. Lu, C.-F. Kuo, *et al.*, "Structured landmark detection via topology-adapting deep graph learning," in *European Conference on Computer Vision*, 2020, pp. 266–283.

[8] Q. Liu, J. Deng, J. Yang, G. Liu, and D. Tao, "Adaptive cascade regression model for robust face alignment," *IEEE Transactions on Image Processing*, pp. 797–807, 2016.

[9] S.-E. Wei, V. Ramakrishna, T. Kanade, and Y. Sheikh, "Convolutional pose machines," in *Proceedings of the IEEE conference on Computer Vision and Pattern Recognition*, 2016, pp. 4724–4732.

[10] J. Wang, K. Sun, T. Cheng, B. Jiang, C. Deng, Y. Zhao, D. Liu, Y. Mu, M. Tan, X. Wang, *et al.*, "Deep high-resolution representation learning for visual recognition," *IEEE transactions on pattern analysis and machine intelligence*, pp. 3349–3364, 2020.

[11] H. Lee, M. Park, and J. Kim, "Cephalometric landmark detection in dental x-ray images using convolutional neural networks," in *Medical imaging 2017: Computer-aided diagnosis*, 2017, p. 101341W.

[12] B. Bier, M. Unberath, J.-N. Zaech, J. Fotouhi, M. Armand, G. Osgood, N. Navab, and A. Maier, "X-ray-transform invariant anatomical landmark detection for pelvic trauma surgery," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*, 2018, pp. 55–63.

[13] Y. Song, X. Qiao, Y. Iwamoto, and Y.-w. Chen, "Automatic cephalometric landmark detection on x-ray images using a deep-learning method," *Applied Sciences*, p. 2547, 2020.

[14] W. P. Segars, G. Sturgeon, S. Mendonca, J. Grimes, and B. M. Tsui, "4d xcat phantom for multimodality imaging research," *Medical physics*, pp. 4902–4915, 2010.

[15] D. S. W.R. Proffit, H.W. Fields, "Contemporary orthodontics," *Elsevier Health Sciences*, p. 8, 2006.

[16] A. Maier, H. G. Hofmann, M. Berger, P. Fischer, C. Schwemmer, H. Wu, K. Müller, J. Hornegger, J.-H. Choi, C. Riess, *et al.*, "Conrad—a software framework for cone-beam imaging in radiology," *Medical physics*, p. 111914, 2013.

[17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2016, pp. 770–778.

[18] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2018, pp. 7132–7141.

[19] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[20] P. Ramachandran, N. Parmar, A. Vaswani, I. Bello, A. Levskaya, and J. Shlens, "Stand-alone self-attention in vision models," *Advances in Neural Information Processing Systems*, vol. 32, 2019.

[21] C.-W. Wang, C.-T. Huang, M.-C. Hsieh, *et al.*, "Evaluation and comparison of anatomical landmark detection methods for cephalometric x-ray images: a grand challenge," *IEEE transactions on medical imaging*, pp. 1890–1900, 2015.

[22] Jie, W. Yao, T. Zeng, S. He, Y. Zhou, J. Zhang, W. Guo, and Tang, "Automatic localization of cephalometric landmarks based on convolutional neural network," *American Journal of Orthodontics and Dentofacial Orthopedics*, pp. e250–e259,, 2022.

[23] M. Lee, M. Chung, and Y.-G. Shin, "Cephalometric landmark detection via global and local encoders and patch-wise attentions," *Neurocomputing*, pp. 182–189, 2022.

〈저자소개〉

이 효 정

- 2017-2020 서울여자대학교 생명환경공학전공, 바이오인포매틱스전공 (학사)
- 2021-현재 이화여자대학교 컴퓨터의학과 (석사)
- 관심분야: Deep Learning, Landmark Detection, Object Detection
- https://orcid.org/0000-0002-2046-3091

마 세 리

- 2017-2021 서울여자대학교 소프트웨어융합학과 (학사)
- 2022-현재 이화여자대학교 휴먼기계바이오공학부 (석사)
- 관심분야: Deep Learning, Landmark Detection, Object Detection
- https://orcid.org/0000-0003-2893-1969

최 장 환

- 2010-2014 Stanford University, Mechanical Engineering (박사)
- 2015 Stanford University, School of Medicine (Postdoctoral Fellow)
- 2016 한국전자통신연구원, 의료정보연구실 (선임연구원)
- 2017-현재 이화여자대학교, 휴먼기계바이오공학부 (조교수, 부교수)
- 관심분야: Computer Vision, Data-informatics, Medical Vision
- https://orcid.org/0000-0001-9273-034X