creative commons
COMMONS DEED

저작자표시-비영리-변경금지 2.0 대한민국

이용자는 아래의 조건을 따르는 경우에 한하여 자유롭게

- 이 저작물을 복제, 배포, 전송, 전시, 공연 및 방송할 수 있습니다.

다음과 같은 조건을 따라야 합니다:

저작자표시. 귀하는 원저작자를 표시하여야 합니다.

비영리. 귀하는 이 저작물을 영리 목적으로 이용할 수 없습니다.

변경금지. 귀하는 이 저작물을 개작, 변형 또는 가공할 수 없습니다.

- 귀하는, 이 저작물의 재이용이나 배포의 경우, 이 저작물에 적용된 이용허락조건을 명확하게 나타내어야 합니다.
- 저작권자로부터 별도의 허가를 받으면 이러한 조건들은 적용되지 않습니다.

저작권법에 따른 이용자의 권리는 위의 내용에 의하여 영향을 받지 않습니다.

이것은 이용허락규약(Legal Code)을 이해하기 쉽게 요약한 것입니다.

Disclaimer

dCollection

공학석사학위논문

중심점 기반 물체 탐지 네트워크와 가우시안의 분리를 이용한 구강 CT 영상에서의 개별 치아 분할

Tooth Instance Segmentation from Dental CT Images via Point-based Detection Network and Gaussian Disentanglement

2021년 2월

서울대학교 대학원

컴퓨터공학부

이 주 상

중심점 기반 물체 탐지 네트워크와 가우시안의 분리를 이용한 구강 CT 영상에서의 개별 치아 분할

Tooth Instance Segmentation from Dental CT Images via Point-based Detection Network and Gaussian Disentanglement

지도교수 신 영 길

이 논문을 공학석사 학위논문으로 제출함

2020 년 12 월

서울대학교 대학원

컴퓨터공학부

이 주 상

이주상의 공학석사 학위논문을 인준함

2020 년 12 월

위 원 장 김 명 수 (인)

부위원장 신 영 길 (인)

위 원 이 영 기 (인)

초 록

3차원 구강 CT 영상에서 개별 치아들의 종류를 식별하고 영역을 분할하는 것은 다양한 치과치료 및 교정치료를 위해 꼭 선행되어야 하는 과정이다. 이와 같은 개별 물체 분할(Instance Segmentation) 문제들은 주로 물체 탐지(Object Detection)가 선행된 후 분할을 진행한다는 공통점이 있는데, 최근 Convolutional Neural Network(CNN)를 수반한 인공신경망 네트워크들이 물체 탐지와 분할 문제들에서 크게 두각을 나타내고 있다. 이에 힘입어 개별 치아를 분할하는 문제에도 딥러닝을 이용한 다양한 방법들이 제한되었지만, 비슷한 형태를 지닌 치아들을 구분하고 인접한 치아들의 경계를 명확히 구분하는 것은 여전히 어려운 문제로 남아있다. 본 연구에서는 가우시안 분리 로스(Gaussian Disentanglement Loss)를 도입한 치아 탐지 네트워크를 제안한다. 이때, 치아 종류를 분류하기 위한 추가적인 학습이 필요하지 않은 Point 기반 탐지 방법을 사용하였고, 이미지 상의 위치정보들을 보존할 수 있는 heatmap regression 방식을 사용하여 학습의 효과를 높였다. 또한, 새로이 고안된 가우시안 분리 로스를 적용하여 모든 인접한 치아쌍에 해당하는 heatmap들의 내적을 최소화 하는 방향으로 손실함수를 부여하여 치아 탐지 정확도의 향상을 도모하였다. 이어진 개별 치아 분할의 경우 인접한 주변 치아들로 인한 성능 저하를 줄이기 위해 픽셀 단위의 라벨링 문제를 distance map regression 문제로 치환하여 진행하였다. 그 결과 치아 탐지에서 기존 알고리즘 대비 9.1%의 평균 정밀도 향상을 보였고, 이러한 개선점에 기인하여 개별 치아 분할에서도 높은 성능을 보였다.

주요어: Object detection, Dental CT, Tooth segmentation

학번: 2019-27115

목 차

초 록 **i**

제 1 장 서론 **1**

제 2 장 관련연구 **5**

2.1 물체 탐지 (Object Detection) 5

2.2 Anchor 기반 물체 탐지 7

2.3 Point 기반 물체 탐지 8

2.4 개별 물체 분할 (Instance Segmentation) 10

2.5 개별 치아 탐지 및 분할 11

제 3 장 가우시안 분리를 이용한 치아 탐지 및 분할 **12**

3.1 개요 12

3.2 Heatmap Regression의 학습 14

3.3 치아 경계 상자 탐지 15

3.4 가우시안 분리 로스 16

3.5 거리 기반 치아 분할 18

제 4 장 실험 결과 **20**

4.1 데이터 구성 및 전처리 20

4.2 학습 방법 20

4.3 성능지표 22

4.4 결과 25

제 5 장 결론 **29**

참고문헌 **30**

Abstract **34**

감사의 글 **35**

그림 목차

그림 1.1 구강 CT 영상 1
그림 1.2 ISO 기준 치아 번호 2
그림 1.3 가우시안 분리(Gaussian Disentanglement) 3
그림 1.4 치아 Mask와 Distance Map 4

그림 2.1 물체 탐지의 응용 5
그림 2.2 Faster R-CNN 네트워크의 구조 6
그림 2.3 Anchor를 이용한 물체 탐지 과정 7
그림 2.4 Non-Maximum Suppression 8
그림 2.5 Point 기반 물체 탐지 9
그림 2.6 Mask R-CNN을 이용한 Instance Segmentation 10

그림 3.1 치아 탐지 네트워크의 전체적인 흐름 12
그림 3.2 치아 분할 네트워크 13
그림 3.3 비모수적 가우시안 분리 로스 17

그림 4.1 Heatmap Regression 학습의 정답지 20
그림 4.2 근사된 유클리드 거리 21
그림 4.3 컨퓨전 행렬 24
그림 4.4 치아 분류의 컨퓨전 행렬 결과 26
그림 4.5 가우시안 분리의 효과 27
그림 4.6 개별 치아 분할 결과 28

표 목차

표 4.1 치아 탐지 실험 결과 25
표 4.2 치아 분류 실험 결과 26

제 1 장 서론

콘빔 전산화 단층촬영 (Cone Beam Computed Tomography; CBCT) 기술은 구강 구조의 고화질 3차원 영상을 얻어낼 수 있어, 치과치료 및 교정치료의 목적으로 많이 활용된다(그림 1.1).

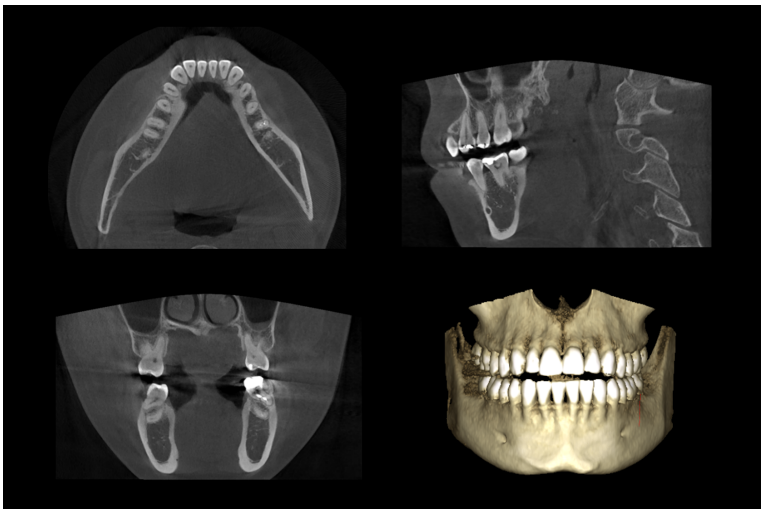

그림 1.1 구강 CT 영상

최근 의료 데이터들의 디지털화에 따라 다양한 컴퓨터 보조 진단 (Computer Aided Diagnosis; CAD) 시스템이 도입되었고, 그에 따라 임상의들이 피상적으로 반복해야 했던 업무들을 효율적으로 줄일 수 있게 되었다[1]. 특히, 교정치료 같은 경우는 성공적인 교정 및 임플란트 계획 수립을 위해 3차원 CBCT 영상에서 개별 치아들을 인식하고 그 영역을 분할해야 하는데, 이 과정을 자동화 할 수 있다면 더욱 효과적인 치료가 가능할 것이다. 그러나 개별 치아의 정밀한 탐지 및 분할은 치아들의 서로 비슷한 형태와 모양, 치아와 치조골 간의 불분명한 경계, 금속

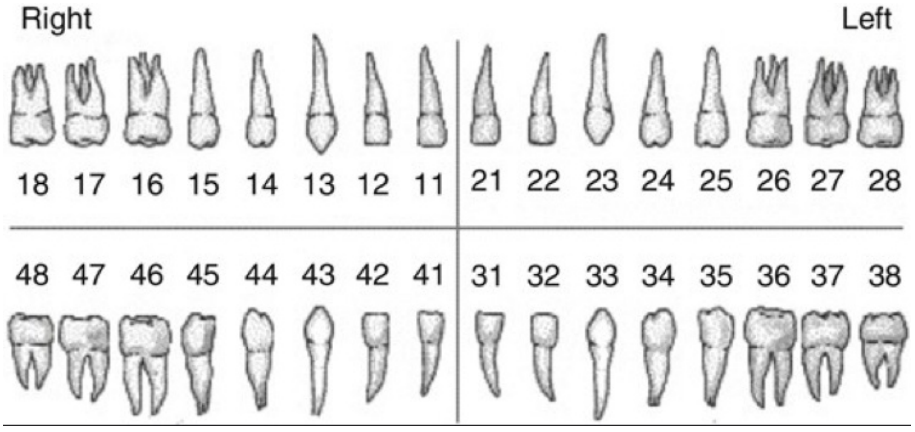

그림 1.2 ISO 기준 치아 번호

물질들로 인한 노이즈 등의 이유로 아직 어려운 문제로 남아있다.

치아 영상에서 개별 치아들을 탐지 및 분할하기 위하여 다양한 연구들이 진행되었다[2, 3, 4]. 고전적인 이미지 처리 기법들을 이용한 방법론들도 많이 제안되었지만, 최근에는 다양한 분야에서 두각을 나타내고 있는 CNN을 이용한 딥러닝 기반 방법론들이 많이 도입되고 있다. 특히 개별 물체 분할(instance segmentation) 방법론은 영상에서 찾고자 하는 대상들을 탐지하고 그 영역을 분할하는 데에 적합하기에 다양한 의료영상 및 치아 영상에서도 활용되고 있다. 개별 물체 분할의 경우 보통 물체 탐지(object detection) 과정이 선행되는데, 이 과정을 통해 찾고자 하는 대상들의 종류를 분류(classification)함과 더불어 대상들의 위치 및 영역을 표현하는 경계상자(bounding box)들을 정확히 찾아낸 후 그 상자들에 대해 분할(segmentation)을 하는 방식으로 진행된다. 그렇기에 물체 탐지가 얼마나 정확하게 수행되었는지가 최종적인 개별 물체 분할의 성능과 직결될 수밖에 없다.

본 연구에서는 개별 치아 탐지 및 분할을 수행하기 위하여 CNN을 이용한 Point 기반 물체 탐지 네트워크를 제안한다. 개별 치아의 종류를 나타내는 데에는 ISO 기준의 치아 번호(그림 1.2)가 사용되었다. Point 기반 물체 탐지는 미리

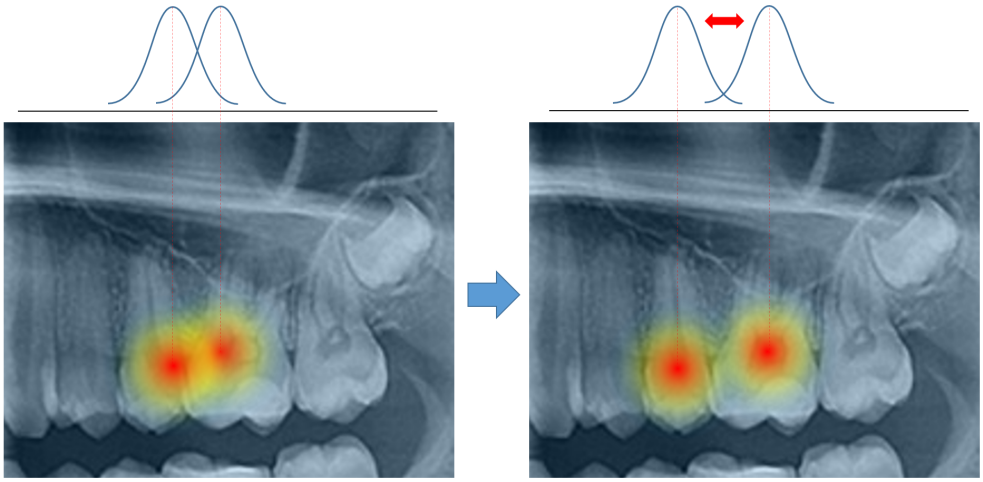

그림 1.3 가우시안 분리(Gaussian Disentanglement)

정해진 클래스들에 대해서 물체의 위치를 탐지하기 때문에, 탐지된 치아의 종류를 판별하는 추가적인 분류 과정이 필요하지 않다. Point 기반 물체 탐지는 크게 direct regression과 heatmap regression 방식으로 나눌 수 있는데[5], 이미지의 위치정보들을 보존할 수 있는 fully-convolutional network(FCN)[6] 구조를 가지고 있어서 더욱 효과적인 학습이 가능한 heatmap regression 방식을 채택하였다. 이때, 새로이 고안된 가우시안 분리 로스를 도입하여 성능을 향상시킬 수 있었다. 인접한 치아가 서로 형태가 비슷하기 때문에 같은 위치에 서로 다른 두 치아가 존재한다고 잘못 판단되는 경우 치아 탐지 성능을 크게 저하시키는 원인이 된다. 가우시안 분리 로스는 그림 1.3과 같이 heatmap 상에서 인접한 치아들의 예측된 가우시안 분포가 서로 비슷한 위치에 겹치는 것을 방지해주는 효과가 있다. 이때, 가우시안 분포의 평균과 표준편차 등의 모수(parameter)들을 사용한 통계학적인 방법론을 기반으로 손실함수를 구축하는 하는 것은 매우 어렵기 때문에 본 연구에서는 비교적 단순한 비모수적(non-parametric) 방식으로 손실함수를 설계하였다. 치아 경계 상자의 경우 탐지된 각 치아의 위치로부터 상자의 가로/세로/높이 값을

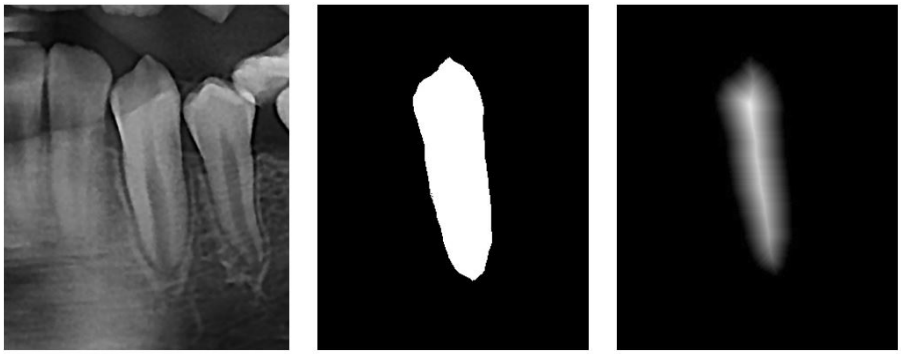

그림 1.4 치아 Mask와 Distance Map

예측하여 최종적으로 결정되게 된다. 이렇게 개별 치아들의 경계상자를 검출한 후 각 상자들에 대해 분할을 하는 단계로 넘어간다. 분할의 경우 U-Net[7] 구조를 3차원으로 확장하여 사용하였다. 또한, 치아 상자의 경계부분에 걸쳐 있는 인접한 치아들로 인해 분할의 성능이 저하되는 현상을 해소하고자 각 픽셀들이 목표하는 물체에 해당하는지 아닌지를 구별하는 분할 문제에서 각 픽셀이 물체의 표면에서 얼마나 떨어져 있는지를 예측하는 distance map regression 방식(그림 1.4)으로 치환하였다.

본 논문은 다음과 같이 구성된다. 먼저 2장에서는 관련 연구를 소개하는데, 물체 탐지 분야에서 많이 사용되는 Anchor 기반과 Point 기반의 개괄적인 내용과 함께 이를 이용한 다양한 물체 탐지 방법론을 설명한다. 3장에서는 본 연구에서 제안하는 방법론에 대하여 설명하고, 4장에서는 본 연구에서 사용된 데이터에 대한 설명과 함께 제안된 네트워크의 성능을 분석한다. 마지막으로 5장에서는 이러한 내용들을 바탕으로 결론을 맺는다.

제 2 장 관련연구

2.1 물체 탐지 (Object Detection)

물체 탐지는 주어진 이미지에서 목표하는 대상들이 어떤 위치에 존재하는지 판단하는 컴퓨터 비전의 대표적인 문제 중 하나이다.[8] 이 과정은 단순히 주어진 라벨을 이미지에 붙이는 것에서 더 나아가 이미지에 대한 더 깊은 이해를 위해 필수적으로 수행되어야 하는 과정으로 얼굴 탐지, 보행자 탐지 등의 다양한 분야에서 활용되고 있다(그림 2.1). 물체 탐지는 기본적으로 물체 영역 탐지(localization)와 물체 종류 판별(classification)로 이루어져 있는데[9], 한 이미지에 정해지지 않은 수의 물체가 동시에 존재할 수 있고, 같은 물체라 하더라도 이미지 상에서 다양한 위치와 크기로 존재할 수 있기에 기본적으로 어려운 문제에 속한다.

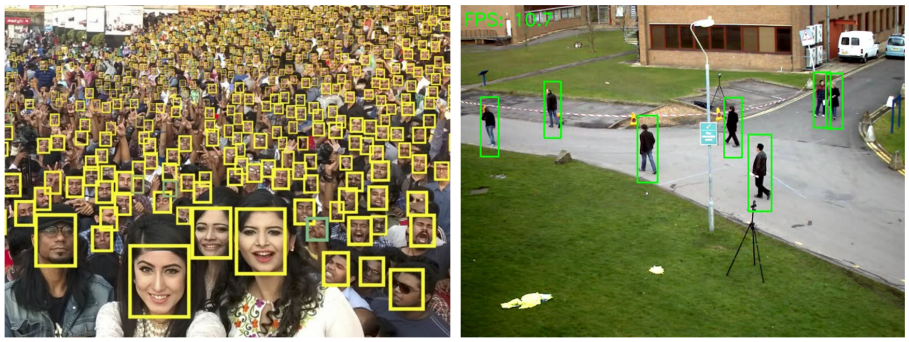

그림 2.1 물체 탐지의 응용. (좌) 얼굴 탐지 (우) 보행자 탐지

다양한 고전적인 이미지 처리 기법들을 이용한 물체 탐지 방법론들이 많이 제안되었지만[10, 11, 12], 대부분의 경우 이미지 상에서 가능한 모든 위치에 다양한 크기의 상자를 놓고 스캔하듯이 상자를 움직이며 물체를 탐지하는 방법을 사용하

고 있어 상당히 시간이 오래 걸린다는 단점이 있다. 또한 사람이 직접 설계한 특징 추출기를 이용하기 때문에 다양한 물체들을 판별하는 데 한계가 있다.

CNN을 이용한 딥러닝 기법이 컴퓨터 비전 분야에 본격적으로 도입되면서 다양한 문제에서 큰 성과를 내었는데, 이는 물체 탐지 분야에서도 마찬가지였다. 깊게 쌓인 인공신경망 구조로 인해 이미지에서 복잡한 특징을 추출하는 것이 가능해지고, 안정적인 학습이 가능해져 더 이상 사람이 손으로 특징 추출기를 설계할 필요가 없어졌다. CNN을 이용한 물체 탐지는 크게 Anchor 기반 방법과 비교적 최근에 제안된 Point 기반 방법으로 나눌 수 있다. 이 두 방식의 가장 큰 차이점은 Anchor 기반 방법과 달리 Point 기반 방법에서는 anchor box를 사용하지 않는다는 점으로, Anchor 기반 방법들의 근본적인 문제를 해결하면서도 그에 뒤지지 않는 성능과 상당히 빠른 속도를 지니고 있어 주목을 받고 있다.

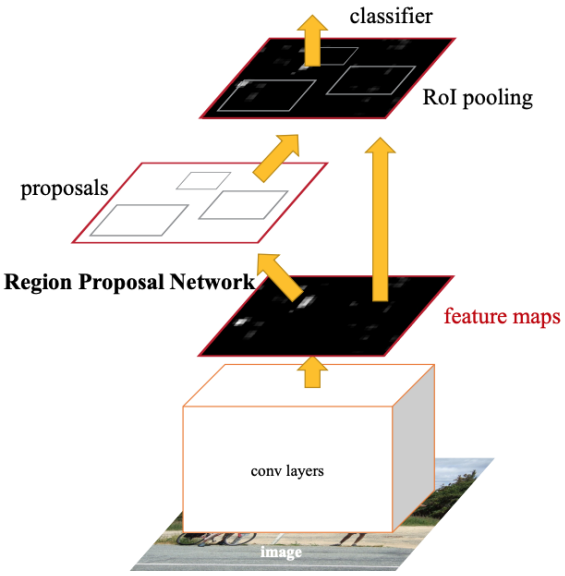

그림 2.2 Faster R-CNN 네트워크의 구조

2.2 Anchor 기반 물체 탐지

Anchor 기반 물체 탐지는 크게 2단계 검출과 1단계 검출로 나눌 수 있는데, 2단계 검출 방식의 경우 대표적으로 R-CNN(Regions with CNN)[13] 계열 네트워크가 있다. R-CNN은 딥러닝을 물체 탐지 분야에 최초로 도입한 기법으로, 이를 필두로 많은 개선이 이루어지면서 현재 여러 분야의 물체 탐지에서 가장 많이 쓰이는 Faster R-CNN 네트워크[14]가 탄생하게 되었다. Faster R-CNN은 이미지 전체를 CNN에 통과시켜 feature map을 추출한 후 이를 이용하여 물체가 존재할 가능성이 높은 영역들을 추천해주는 region proposal network(RPN)를 도입하였다는 것이 가장 큰 특징이다(그림 2.2). 이렇게 RPN을 통해 추천된 다수의 영역들에 대해 최종적으로 그 물체의 종류를 판별하고 경계상자의 정확한 범위를 예측하는 두 번째 단계를 거쳐 물체 탐지가 완료된다. 1단계 검출 방식의 경우 기존 R-CNN 계열과는 달리 각 이미지 픽셀들로부터 물체 상자를 직접적으로 regress 한다는 차이점이 있는데, 대표적으로 YOLO 계열 네트워크[15, 16, 17]들이 있다.

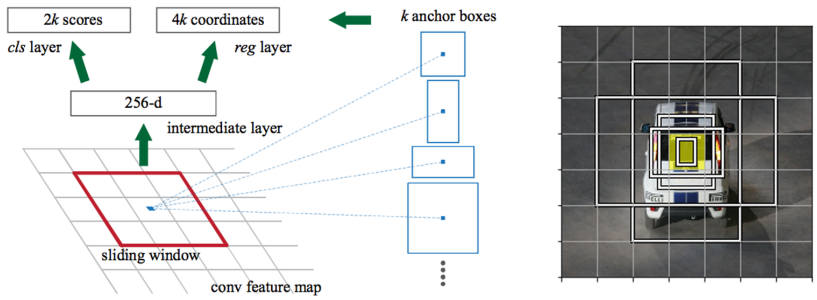

그림 2.3 Anchor를 이용한 물체 탐지 과정

2단계 검출과 1단계 검출은 이렇게 확연히 다른 특징을 보이지만, 둘 모두 anchor box를 활용한다는 공통점이 있다. Anchor box란 사용자가 미리 정의해 둔 다양한 크기와 비율의 상자들을 의미하는데, 물체가 존재할 것이라고 판단되는

위치에 anchor box를 적용하여 해당하는 영역의 이미지나 feature map을 잘라 사용할 수 있게 하는 방식을 말한다(그림 2.3). 이를 통해 사용자는 이미지 상에 물체가 다양한 크기나 비율로 존재하더라도 성공적으로 탐지할 수 있게 된다. 그렇기에 anchor box를 어떻게 정의하느냐에 따라 물체 탐지의 성능이 크게 좌우되는데[18], 높은 성능을 위해 무턱대고 많은 종류의 anchor box를 사용하게 되면 그만큼 네트워크의 처리 속도가 급격히 줄어든다는 문제가 있다. 또한 필연적으로 이미지 상의 하나의 물체에 대해 겹쳐진 다수의 경계 상자가 검출되기 때문에, 최종 단계에서 이들 중 해당 물체에 가장 적합한 경계 상자를 추출해 내는 Non-Maximum Suppression(NMS, 그림 2.4)을 필수적으로 수행해야 한다. 이 NMS 역시도 네트워크의 전체적인 처리 속도를 저하시키는 원인이 된다.

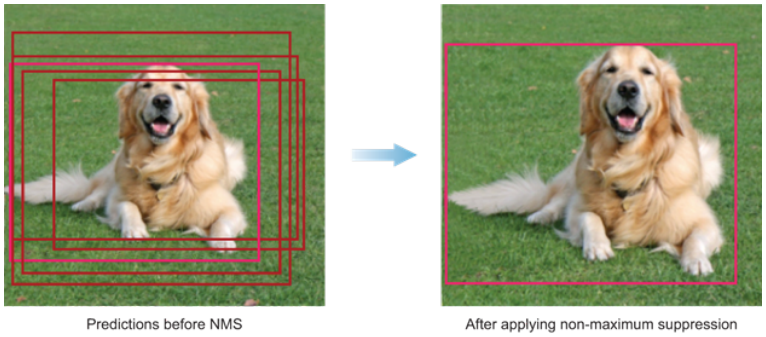

그림 2.4 Non-Maximum Suppression

2.3 Point 기반 물체 탐지

Point 기반 물체 탐지는 상술한 Anchor 기반 물체 탐지의 근본적인 문제인 anchor box 자체를 사용하지 않는다는 특징이 있다. Point 기반 물체 탐지를 처음으로 도입한 CornerNet[18]의 경우 이미지 상에서 각 물체를 감싸는 경계 상자의 좌상단 점과 우하단 점들의 위치를 예측하여 물체 탐지를 수행하는 방식을 채택하

였다. 이때, 탐지된 좌상단/우하단 점들이 서로 올바르게 짝지어져야만 제대로 된 물체 탐지 결과를 도출할 수 있는데, 이를 위해서 같은 물체는 벡터 공간 상에서 서로 가까운 위치로 사상(Embedding) 될 수 있도록 학습시키는 과정이 필요하다. 그에 비해 CenterNet[19] 같은 경우 CornerNet과는 달리 각 물체의 중심점을 기반으로 하기에 점들을 짝지어야 하는 과정이 필요하지 않다는 차이점이 있다. 이미지 상에 목표로 하는 모든 물체들의 중심점의 위치를 예측하고 난 후 각 물체의 경계상자의 크기를 예측하여 물체 탐지를 완료하게 된다(그림 2.5).

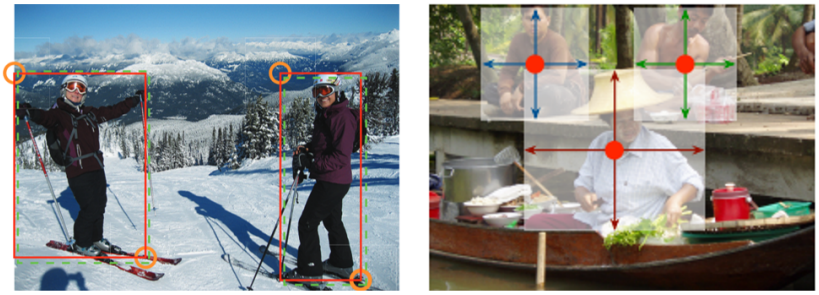

그림 2.5 Point 기반 물체 탐지. (좌) 좌상단/우하단 점 방식 (우) 중심점 방식

Stacked Hourglass 네트워크[5]의 경우 정확히는 물체 탐지 기법이 아니라 keypoint estimation 기법으로 분류되지만, 이미지 상에서 목표로 하는 위치를 찾는다는 점에서 Point 기반 물체 탐지 기법들과 유사성을 지닌다. 다만 keypoint estimation의 경우 이미지 상에서 찾아야 하는 점들의 종류 및 개수가 미리 정해져 있다는 차이점이 있다. 그렇기에 keypoint estimation의 경우 물체 탐지 보다는 이미지 상에서 사람이 어떠한 자세를 취하고 있는지 판단하는 human pose estimation 분야에 더 많이 활용된다.

이렇게 Point 기반 물체 탐지 및 keypoint estimation은 주어진 이미지 상에서 목표하는 점들의 위치를 찾아낸다는 공통점이 있다. 이때, 점들의 위치를 예측하는 방법에는 크게 직접 예측 방식[20]과 Heatmap Regression을 통한 예측 방식[21]

이 있다. 직접 예측 방식은 정해진 개수의 좌표값들을 직접적으로 regress 하는 방식으로 학습이 진행된다. 그에 비해 Heatmap Regression 방식은 이미지와 동일한 크기의 heatmap을 출력하게 되는데, 이는 각 이미지 픽셀의 위치에 찾고자 하는 물체가 존재할 확률을 나타낸다. 출력된 heatmap을 통해서 물체의 위치를 찾을 때에는 각 물체의 위치가 정규분포를 이룬다고 가정하여 heatmap 상의 피크점들의 위치를 찾는 것으로 물체의 위치를 특정할 수 있게 된다. 일반적으로 direct regression 보다는 heatmap regression 방식이 더 높은 성능을 보이는데, 이는 heatmap regression에서 fully-convolutional 네트워크 구조를 사용하여 이미지 상의 위치정보들을 보존할 수 있기 때문인 것으로 알려져 있다[6].

2.4 개별 물체 분할 (Instance Segmentation)

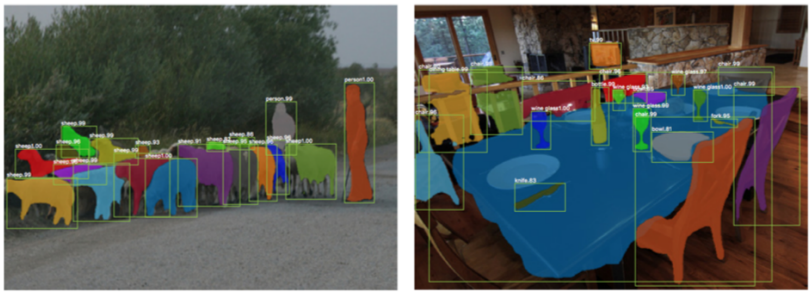

그림 2.6 Mask R-CNN을 이용한 Instance Segmentation

이미지의 각 픽셀들이 어떤 물체에 해당하는지를 예측하는 분할(Segmentation)은 크게 시맨틱 분할(Semantic Segmentation)과 개별 물체 분할(Instance Segmentation)으로 나눌 수 있다. 단순히 이미지의 모든 픽셀들이 어떤 물체에 해당하는지 예측하는 시맨틱 분할과는 달리, 개별 물체 분할의 경우 이미지에서 각 물체들이 존재하는 영역을 구분한 후, 각 영역에 대해 개별적으로 분할을 진행한다는 점에서 차이가 있다. 그렇기에 개별 물체 분할은 필연적으로 물체 탐지

문제와 그 결을 같이 하게 된다. 대표적인 개별 물체 분할 네트워크인 Mask R-CNN[22](그림 2.6)의 경우도 상술했던 Anchor 기반 물체 탐지 방법 중 하나인 Faster R-CNN을 개량한 것으로, 검출된 물체들의 경계상자들을 대상으로 분할을 수행하는 방식을 채택하고 있다. 이 외에도 다양한 개별 물체 분할 방식들이 있지만 대부분 물체 탐지가 선행된다는 공통점을 지닌다. 물체 경계 상자들만 성공적으로 검출된다면, 일반적으로 그 영역 내에서 분할을 학습하는 것은 비교적 단순한 문제이므로 개별 물체 분할에 있어서 그 성능을 결정하는 데에는 물체 탐지가 결정적인 영향을 끼친다.

2.5 개별 치아 탐지 및 분할

구강 CT 영상 및 구강 파노라마 영상에서 개별 치아를 분할하는 다양한 방법이 제안되었다. ToothNet[4]의 경우 상술한 Mask R-CNN 구조를 3차원으로 확장하여 채택하였고, 서로 다른 종류의 치아들을 구분하기 위해 Similarity Matrix를 도입하였다. [2]의 경우도 동일하게 3차원 Mask R-CNN 구조를 채택하였으나, 보다 나은 분할 성능을 위해 원본 영상을 상악과 하악 영역으로 분리한 후 각 영역이 지면과 평행을 이루도록 정렬하여 사용하였다. 또한 각 치아 영역에 대한 Segmentation을 수행할 때 단순히 각 픽셀들이 치아 영역에 해당하는지 여부를 판단하는 문제에서 각 픽셀이 치아의 표면에서 얼마나 떨어져 있는지를 예측하는 distance map regression 방식으로 치환하였다. [3]에서는 구강 파노라마 영상에서 치아들의 위치를 찾아내기 위해 Anchor 기반의 방법이 아닌 Point 기반의 방식을 사용해서 각 치아의 중심점들의 좌표값들을 직접적으로 regress 하는 방식을 채택하였다. 이때 인접한 치아들의 좌표값들이 서로 겹치는 것을 방지하기 위해 Spatial Distance Regularization Loss 를 도입하였다. 치아의 중심점을 기반으로 하기에 CenterNet과 유사하지만 32개로 정해진 개수의 치아에 대해 좌표값들을 예측하므로 keypoint estimation과도 유사성을 띈다.

제 3 장 가우시안 분리를 이용한 치아 탐지 및 분할

3.1 개요

본 연구에서는 구강 CT 영상에서 개별 치아 영역들을 분할하는 프레임워크를 개발하였다.

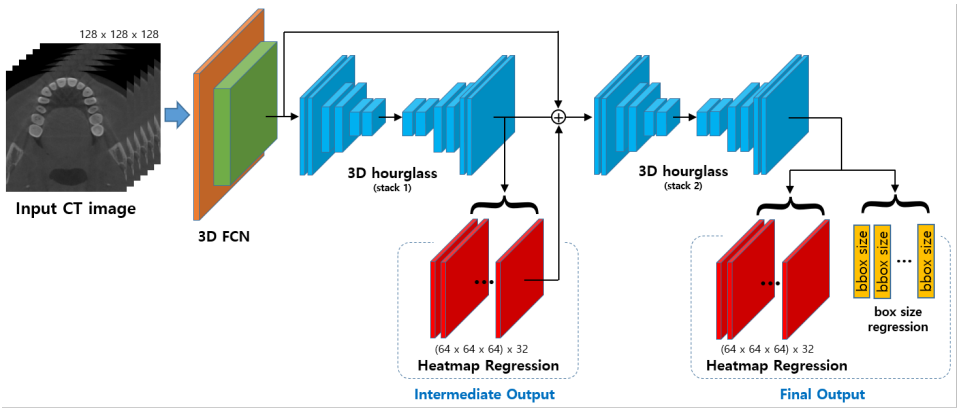

그림 3.1 치아 탐지 네트워크의 전체적인 흐름

먼저, 치아 탐지의 전체적인 흐름은 그림 3.1과 같다. 학습에 사용되는 구강 CT 영상은 크기가 128 x 128 x 128로 조절되어 네트워크에 입력으로 들어간다. 이후 짧은 FCN 레이어를 통과한 다음, 이어지는 encoding-decoding 구조를 통과하여 heatmap regression에 사용될 feature map이 추출된다. 이때 사용된 특징 추출기의 backbone으로는 hourglass[5]를 기반으로 하였고, 이를 3차원으로 확장해서 사용하였다. 전체적인 네트워크는 2-stack 구조를 따르는데, 이는 네트워크의 중간단계의 산출물도 손실함수에 포함시켜 학습에 도움을 주기 위함이다. 첫 번째 hourglass에서 추출된 feature map을 이용하여 산출된 heatmap 결과들은 이전의

feature map들과 합쳐져 두 번째 hourglass에 통과된다. 마지막 단계에서는 최종적인 heatmap 결과와 더불어 각 치아에 해당하는 경계상자의 크기도 regression을 통해 예측한다.

치아 탐지가 완료되면, 그 결과를 바탕으로 독립적인 분할 네트워크(그림 3.2)를 통해 치아 분할을 수행하게 된다. 탐지된 치아 경계상자들을 기준으로 원본 이미지를 자른 후 크기를 128 x 64 x 64로 조절하여 네트워크에 입력으로 넣는다. 분할 네트워크는 U-Net[7]을 기반으로 하여 3차원으로 확장해서 사용하였다. 입력과 동일한 크기의 이미지가 출력되면, 이를 통해 탐지된 치아 경계상자의 어떤 부분이 치아에 해당하는지를 판별하여 최종적으로 개별 치아 분할이 완료된다.

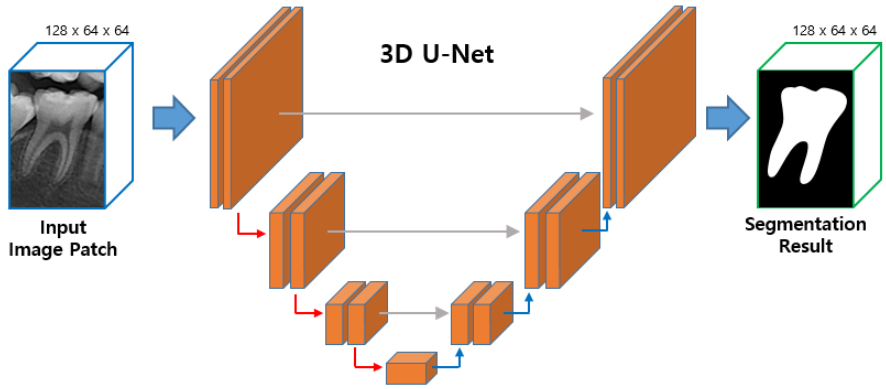

그림 3.2 치아 분할 네트워크

이어지는 3.2절부터는 3.5절까지 에서는 상술한 네트워크의 구조를 더 자세히 다루고, 각 부분에서 어떤 방식으로 손실함수들을 적용하여 학습을 진행하였는지 서술한다.

3.2 Heatmap Regression의 학습

위의 3.1절에 나타나있듯이, 제안된 치아 탐지 네트워크는 구강 CT 이미지를 입력으로 받아서 heatmap을 결과로 출력한다. FCN과 hourglass 등의 backbone 레이어들을 통과하여 추출된 feature map은 이어서 heatmap regression 레이어를 통과하게 된다. 이때, 한 이미지에서 탐지하고자 하는 치아의 종류가 총 32개이므로 총 32개의 heatmap regression 레이어가 병렬 구조를 이루고 있으며, 이를 통해 인해 총 32개의 heatmap을 하나의 묶음으로 출력하게 된다. 이러한 heatmap regression 과정을 학습하기 위한 손실함수로는 Focal Loss[18]가 사용되었다. 이는 다른 Point 기반 탐지 기법들에서도 많이 사용되는 손실함수로, 정답지의 heatmap에서 0인 부분이 0이 아닌 부분보다 월등히 많다는 특징 때문에 단순한 Mean Squared Error Loss(MSE Loss) 등의 손실 함수로는 효과적으로 학습되기 어렵기 때문이다. Focal Loss를 이용한 손실함수는 다음과 같은 식으로 표현된다.

$$
\begin{aligned}
L_{heat} &= Focal(\mathbf{x}_{heat}, \mathbf{y}_{heat}) \\
&= \frac{-1}{N} \sum_{p} \begin{cases} (1-\mathbf{x}(p))^{\alpha} \log\left(\mathbf{x}(p)\right) & \text{if } \mathbf{y}(p) = 1 \\ (1-\mathbf{y}(p))^{\beta}\, \mathbf{x}(p)^{\alpha} \log\left(1-\mathbf{x}(p)\right) & \text{otherwise} \end{cases}
\end{aligned} \tag{3.1}
$$

여기서 $\mathbf{x}_{heat}$는 네트워크에서 출력된 heatmap 결과, $\mathbf{y}_{heat}$는 정답지를 의미하고, 픽셀 p의 위치에서의 값은 각각 $\mathbf{x}(p)$와 $\mathbf{y}(p)$로 나타낸다. 수식의 조건에서 알 수 있듯이, 정답지에서 물체가 존재하는 피크점에서와 피크점이 아닌 곳에서의 계산식이 다르다. 그렇게 각 픽셀의 위치에서 계산된 손실값을 모두 더한 뒤, 정답지의 피크점 개수로 나누면 최종적으로 Focal Loss의 값을 얻어낼 수 있다. α와 β의 값은 $\mathbf{x}_{heat}$와 $\mathbf{y}_{heat}$가 전체 손실함수에 얼마나 영향을 미치는지를 조절하는데, 상대적으로 개수가 적은 피크점 쪽이 전체 손실함수에 더 큰 영향을 미칠 수 있도록 $\alpha = 2$, $\beta = 4$로 설정하였다.

또한 네트워크의 전체적인 구조가 2-Stack 구조를 이루고 있는데, 이러한 구조 역시 다양한 문제에서 사용되고 있다. 이는 네트워크의 중간단계 산출물도 손실함수에 포함시키기 위함인데, 이러한 학습 방식을 Intermediate Supervision[5]이라고 한다. 이 방식을 적용하면 단순히 CNN 레이어를 깊게 쌓는 것과 비교해서 같은 깊이로도 더 높은 학습효과를 볼 수 있다고 알려져 있다. 각 stack에서의 출력값 $\mathbf{x}_{heat}(s)$에 대해 intermediate supervision을 적용한 손실함수는 다음과 같다. $w(s)$는 각 stack이 전체 손실함수에 얼마나 영향을 끼치는지 결정하는 가중치 값으로, 여기서는 전부 1로 설정하였다. 본 연구에서는 가장 기본적인 2-Stack 구조를 사용하였으므로, s=2번의 Focal Loss를 반영하여 L_{heat}를 계산하게 된다.

$$L_{heat} = \sum_{s} w(s) Focal(\mathbf{x}_{heat}(s), \mathbf{y}_{heat}) \tag{3.2}$$

3.3 치아 경계 상자 탐지

개별 치아들을 분할하기 위해서는 앞서 설명한 치아 탐지 네트워크를 통해 얻어낸 heatmap을 기반으로 각 치아들의 경계상자를 얻어내야 한다. 먼저 각 치아들의 중심점의 위치를 특정해야 하는데, 이는 32개의 heatmap 결과들 각각에서 최댓값을 가지는 픽셀의 좌표를 통해 예측하게 된다. 치아 번호 t=1번부터 t=32번까지에 해당하는 heatmap인 $\mathbf{x}_t(p)$에서 각 치아의 중심점들의 위치들을 찾는 과정은 다음과 같이 표현된다.

$$P_{centers} = \{ \underset{p}{argmax} \left(\mathbf{x}_t(p) \right) \mid \text{for } t \in [1, 32] \} \tag{3.3}$$

치아들의 중심점을 찾은 후, 해당하는 치아 경계상자의 크기를 예측하여 치아 탐지를 완료하게 된다. 각 치아 경계상자의 가로, 세로, 높이 또한 네트워크로 예측하게 되는데, 이를 위한 regression의 손실함수는 MSE Loss를 사용하였다.

$$L_{bbox} = MSE(\mathbf{x}_{bbox}, \mathbf{y}_{bbox}) \tag{3.4}$$

탐지된 치아의 중심점의 좌표 $P_{center} = (x_{center}, y_{center}, z_{center})$와 예측된 경계상자의 크기 $(W_{bbox}, H_{bbox}, D_{bbox})$를 통해 최종적으로 치아의 경계상자를 다음과 같이 표현한다.

$$\begin{aligned} Bbox = (&x_{center} - W_{bbox}/2,\ x_{center} + W_{bbox}/2, \\ &y_{center} - H_{bbox}/2,\ y_{center} + H_{bbox}/2, \\ &z_{center} - D_{bbox}/2,\ z_{center} + D_{bbox}/2) \end{aligned} \tag{3.5}$$

3.4 가우시안 분리 로스

가우시안 분리 로스는 본 연구에서 새로이 고안된 손실함수로, 앞서 설명한 heatmap regression의 학습의 효과를 높이기 위해 도입되었다. 치아 CT 영상의 특성상 비슷한 형태를 가진 치아들이 서로 가까이 인접해 있기 때문에, 목표하는 치아의 위치를 정확히 탐지하는 데에 어려움이 따른다. 특히, 한 치아에 대한 heatmap 상에 여러 위치에서 false positive가 많이 발생하거나 인접한 두 치아에 대한 heatmap들이 같은 위치에 가우시안 분포를 형성하는 경우가 많아 치아 탐지의 성능이 저하된다.

가우시안 분리 로스는 이렇게 인접한 치아에 대해 heatmap regression이 겪는 어려움을 완화하는 역할을 한다. 이 손실함수를 적용하면 인접한 두 치아의 heatmap 상에 예측된 가우시안 분포가 서로 가까울수록 큰 손실을 발생시키게 된다. 두 가우시안 분포가 얼마나 가까운지를 판단하기 위해서는 각 분포의 평균과 표준편차 등의 모수들을 알아야 하는데, 이러한 모수적(paremetric) 방법을

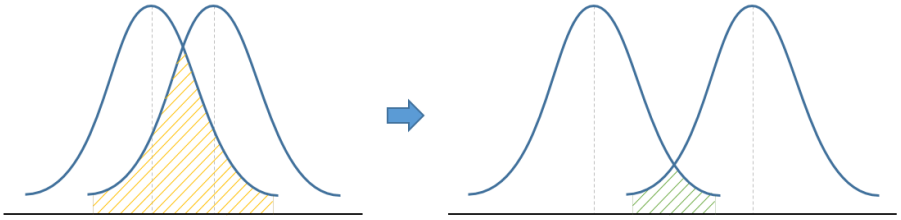

그림 3.3 비모수적 가우시안 분리 로스의 원리

적용하는 데에는 복잡한 통계학 이론들과 수식들이 필요하고, 이를 손실함수로 유도해내는 것 또한 쉽지 않다. 그렇기에 본 연구에서는 가우시안을 분리하기 위하여 비교적 간단한 비모수적 방식으로 손실함수를 설계하였다.

두 치아에 대한 가우시안 분포가 서로 가까울 경우 그림 3.3에서처럼 두 분포 모두에서 확률밀도함수의 값이 높은 지점이 많아지게 된다. 그에 비해 만약 두 분포가 충분히 멀리 있다면 대부분의 위치에서 두 확률밀도함수의 값이 모두 높게 나타나지는 않는다. 그러므로 만약 모든 위치에서 두 확률밀도함수를 곱한다면, 분포가 겹치는 부분에서만 높은 값이 나오고 다른 위치들에서는 0에 가까운 값들이 나올 것이다. 이러한 현상을 이용해서 비모수적 가우시안 분리 로스(GD Loss)를 다음과 같은 식으로 나타내었다.

$$L_{GD} = \sum_{(i,j)\in N} \sum_{p} \mathbf{x}_i(p) \odot \mathbf{x}_j(p) \tag{3.6}$$

여기서 $\mathbf{x}_i(p)$와 $\mathbf{x}_j(p)$는 각각 i번째, j번째 치아에 해당하는 heatmap을 뜻하고, 수식의 기호 $\odot$는 두 heatmap을 각 픽셀별로 곱셈하는 연산을 뜻한다. 그렇게 얻어진 결과물에서 각 픽셀들의 합을 구하는데, 이러한 과정을 모든 (i, j) 쌍들 중에서 인접한 치아 쌍들에 대해 적용한 후 전체를 합산하면 최종적으로 GD Loss가 구해지게 된다. 그림 1.2에 나타나 있듯이 ISO 기준 치아 번호는 11번부터 48번까지 있는데, 이들 중 인접한 치아는 치아번호가 서로 1이 차이나는 치아들이 있고,

추가적으로 11번과 21번, 그리고 31번과 41번 치아도 인접한 치아로 간주하였다.

상술한 heatmap regression과 intermediate supervision, 경계상자의 크기 예측, 그리고 가우시안 분리 로스까지 합하여 최종적으로 치아 탐지 네트워크를 학습하기 위한 손실함수는 다음과 같다.

$$L = \lambda_1 L_{heat} + \lambda_2 L_{bbox} + \lambda_3 L_{GD} \tag{3.7}$$

λ_1, λ_2, λ_3은 각각의 손실함수가 전체 손실함수에 얼마나 반영되는지를 나타내는 가중치이다. 본 연구에서는 $\lambda_1 = 0.1$, $\lambda_2 = 0.1$, $\lambda_3 = 1$이 사용되었다.

3.5 거리 기반 치아 분할

치아 탐지가 성공적으로 진행되었다면 탐지된 치아 경계상자들을 기준으로 원본 이미지를 잘라서 얻어진 작은 이미지 조각에는 오직 치아 한 개만 포함되어 있을 것이라고 기대할 수 있다. 그렇기에 이미지 조각들의 각 픽셀들이 치아에 해당하는지 아닌지를 판별할 수 있다면 치아 영역을 성공적으로 분할해 낼 수 있을 것이다. 분할을 학습하는 데에는 많은 분야의 분할 문제에서 높은 성능을 보이고 있는 U-Net 구조의 네트워크를 를 3차원으로 확장해서 사용하였다.

그러나 구강 CT 영상의 특성상 인접한 치아들이 너무 가깝게 붙어 있어서 어떻게 이미지를 자르더라도 중심이 되는 치아를 제외한 다른 치아의 일부분이 이미지 조각에 포함될 수밖에 없다. 그러므로 단순히 이미지 조각의 각 픽셀이 치아인지 아닌지 만을 판별하도록 학습을 하게 된다면 중심이 되는 치아 영역뿐만 아니라 주변 치아 영역에 해당하는 픽셀들까지 치아라고 판별될 가능성이 높다. 그렇게 되면 주변 영역의 픽셀들에서 많은 false positive가 발생할 것이고, 이는 곧 개별 치아 분할의 성능을 저하시키는 원인이 될 것이다. 이를 해결하기 위해 본

연구에서는 [2]에서 제안된 거리 기반 분할(distance-based segmentation)을 도입하였는데, 이는 기존의 픽셀 단위의 분류 문제를 픽셀 단위의 regression 문제로 치환한 것이다. 그에 따라 학습을 위한 손실 함수도 분류를 위한 Binary Cross Entropy Loss(BCE Loss)를 대신하여 MSE Loss가 사용되었다. 거리 기반 분할 학습에 사용된 손실함수는 다음과 같다.

$$L_{dist} = MSE(\mathbf{x}_{dist}, \mathbf{y}_{dist}) \tag{3.8}$$

여기서 $\mathbf{x}_{dist}$는 거리 기반 분할 네트워크의 출력값을 의미하고, $\mathbf{y}_{dist}$는 정답지로 사용된 distance map을 뜻한다. 학습이 완료된 후 실제로 분할을 수행하고자 할 때에는 네트워크에서 출력된 distance map을 정해진 값으로 thresholding 하여 최종적인 개별 치아 분할 결과를 얻어내게 된다.

제 4 장 실험 결과

4.1 데이터 구성 및 전처리

본 연구에서는 총 120개의 구강 CT 데이터가 사용되었고 그 중 학습에 80장, 검증에 20장, 실험에 20장을 사용하였다. 각 구강 CT 데이터는 512 x 512 x 512의 크기를 가지고 있으며 각 픽셀의 값은 Hounsfield Scale을 따르기 때문에 가장 밀도가 낮은 부분인 -1024부터 가장 밀도가 높은 부분인 3072으로 이루어져 있다. 전반적으로 밝게 나오거나 어둡게 나오는 등 데이터마다 픽셀 값들의 분포가 다르기에, 학습에 사용되기 전 각 데이터 별로 정규화 과정을 거쳐서 사용되었다.

4.2 학습 방법

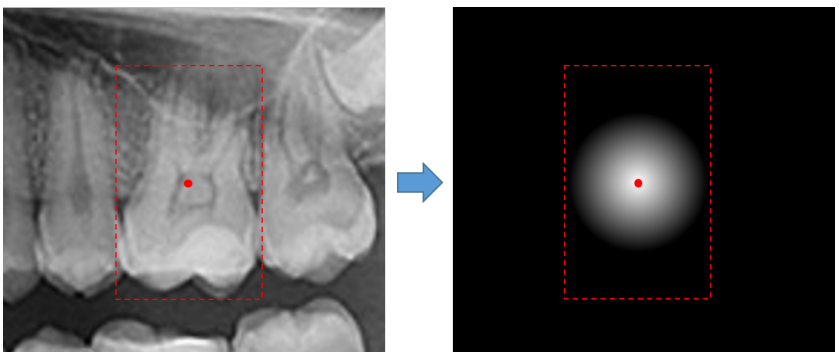

그림 4.1 (좌) 치아 이미지 (우) Heatmap Regression 학습의 정답지

Heatmap Regression을 학습하는 데에 사용하기 위한 정답지로는 Gaussian Distribution Heatmap을 생성하였는데, 입력 이미지와 동일한 크기의 0으로 채워진 이미지에서 찾고자 하는 치아의 중심점을 기준으로 하는 가우시안 분포를

따라 값을 채우는 방식으로 만들었다(그림 4.1). 이때, 가우시안 분포의 표준편차 값은 해당하는 치아의 크기에 비례하도록 설정하였다. 그렇게 각 학습 데이터 마다 모든 치아들에 해당하는 총 32개의 heatmap 정답지를 생성하였고, heatmap regression 학습 단계의 손실함수에 사용되었다. 또한 치아들의 중심점을 기준으로 경계상자의 너비, 높이, 깊이에 해당하는 값들도 메타데이터로 구조화 하여 학습에 사용하였다.

개별 치아 분할을 위한 학습은 각 치아 영역에 해당하는 픽셀들의 위치를 나타낸 Mask 데이터를 기반으로 distance map을 생성하여 정답지로 활용하였다. 생성한 distance map의 각 픽셀들의 값은 치아의 표면으로부터의 거리를 나타내는 것으로, 치아 외부는 모두 0으로 채워져 있지만 치아 표면부터 시작해서 치아 내부로 들어갈수록 거리가 증가하는 형식으로 되어있다. Distance map을 생성하는 알고리즘으로는 [23]에서 제안된 간단한 Two-Pass 알고리즘을 이용하였다. 또한 픽셀들 간의 거리는 유클리드 거리(Euclidean Distance)로 정의하였는데, 이때 대각선 방향의 거리가 유리수가 아니기 때문에 근사된 유클리드 거리(그림 4.2)를 사용하였다.

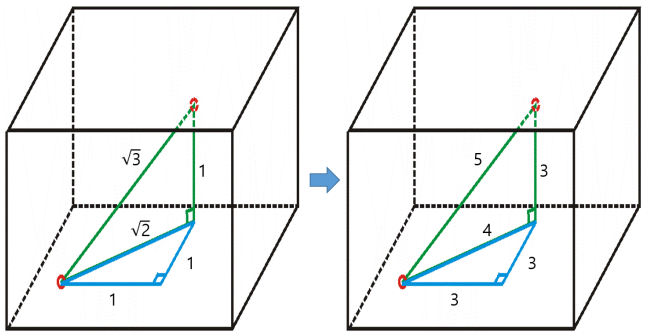

그림 4.2 (좌) 3차원 공간상의 유클리드 거리 (우) 근사된 유클리드 거리

4.3 성능지표

1) Intersection over Union(IOU): 물체 탐지의 성능을 결정짓는 데에는 탐지된 물체의 경계상자가 정답지의 경계상자와 얼마나 많이 겹치는가를 기준으로 한다. 이때, 두 상자가 서로 겹쳐지는 정도를 나타내는 수치로 IOU가 가장 많이 쓰인다. 두 상자 A, B 영역의 IOU를 계산하는 수식은 다음과 같다.

$$IOU = \frac{|A \cap B|}{|A \cup B|} \tag{4.1}$$

2) 정밀도(Precision)와 재현율(Recall): 정밀도와 재현율은 보통 분류 기법의 성능을 나타내는 지표로 많이 사용되는데, 각각의 수식은 다음과 같다.

$$precision = \frac{true\,positives}{true\,positives + false\,positives} \tag{4.2}$$

$$recall = \frac{true\,positives}{true\,positives + false\,negatives} \tag{4.3}$$

이러한 정밀도와 재현율은 물체 탐지 분야에서도 활용되는데, 이때 탐지된 경계상자들이 true positive로 분류될 수 있는지의 여부를 결정짓는 방법이 바로 IOU를 고려하는 것이다. 탐지된 경계상자와 그에 해당하는 정답지 경계상자와의 IOU 값이 특정 threshold 보다 크다면 그 물체는 성공적으로 탐지되어 true positive로 포함시킬 수 있지만, 만약 그렇지 않을 경우는 false positive로 분류된다. 탐지된 경계상자들 중 성공적인 탐지의 비율이 높을수록 정밀도가 증가하고, 전체 정답지의 물체들 중 성공적으로 탐지된 물체가 많을수록 재현율이 증가하게 된다.

3) 평균 정밀도(Average Precision; AP): 탐지된 경계상자들을 신뢰도(Confidence) 순으로 정렬하여 각각의 신뢰도 수준에 대한 정밀도와 재현율의 값으로 정밀도-재현율 곡선(PR-Curve)을 그릴 수 있는데, 이 곡선 아래의 면적이

바로 평균 정밀도에 해당한다. 면적이 넓을수록, 즉 재현율이 증가함에 따른 정밀도의 감소폭이 낮을수록 평균 정밀도의 값이 커지게 되어 더욱 높은 성능의 물체 탐지라고 볼 수 있다. 이때, 상술한 정밀도와 재현율을 판단하기 위해 사용되었던 IOU threshold 값에 따라 평균 정밀도의 값이 달라지는데, 물체 탐지 분야에서는 보통 IOU threshold 값으로 50을 사용한다. 그렇게 계산된 평균 정밀도는 AP50으로 표기한다.

4) Object Include Ratio(OIR): 본 연구의 최종적인 목표는 물체 탐지가 아닌 개별 물체 분할인 만큼, 분할의 성능에 크게 영향을 미칠 수 있는 요소도 평가에 포함해야 할 것이다. 상술했듯 개별 물체 분할은 먼저 물체 탐지 과정을 거쳐 추출된 경계상자들에서 분할을 진행하게 되는데, 아무리 분할의 성능이 좋아도 만약 경계상자가 물체를 온전히 포함하고 있지 못하면 성공적으로 물체를 분할하지 못 할 것이다. 그러므로 단순히 경계상자들의 IOU만을 고려하기 보다는 상자가 실제로 물체를 얼마나 잘 포함하고 있는지의 여부도 평가해야 할 것이다. 그렇기에 본 연구에서는 모든 물체 탐지 결과에 대해 OIR 값을 구한 후 평균을 취하여 성능 지표에 포함시켰다. 물체 영역 M에 대해 탐지된 경계상자의 영역 D에 대한 OIR을 계산하는 수식은 다음과 같다[2].

$$OIR = \frac{|M \cap D|}{|M|} \tag{4.4}$$

5) 컨퓨전 행렬(Confusion Matrix): 개별 치아 탐지의 경우, 치아 영역의 경계상자를 잘 검출하는 것뿐만 아니라 치아 종류까지 정확히 분류하는 것도 매우 중요하다. 이렇게 여러 개의 클래스로 이루어진 분류 문제에서 성능을 평가하기 위해 자주 사용되는 것으로 컨퓨전 행렬(예시: 그림 4.3)이 있다. 컨퓨전 행렬은 가로축은 예측된 클래스, 세로축은 실제 클래스로 놓고 네트워크가 수행한 모든

분류에 대해 해당하는 칸에 더하는 방식으로 형성된다. 그 후 모든 칸의 값들을 각 칸이 해당하는 클래스에 속한 물체들의 총 개수로 나누어 완성한다. 성공적인 분류 네트워크라면 컨퓨전 행렬의 대각 성분들에만 값이 몰려 있을 것이다. 행렬의 크기가 클 경우 각 칸의 값에 따라 색을 입혀서 한눈에 알아볼 수 있도록 하는 경우가 많다.

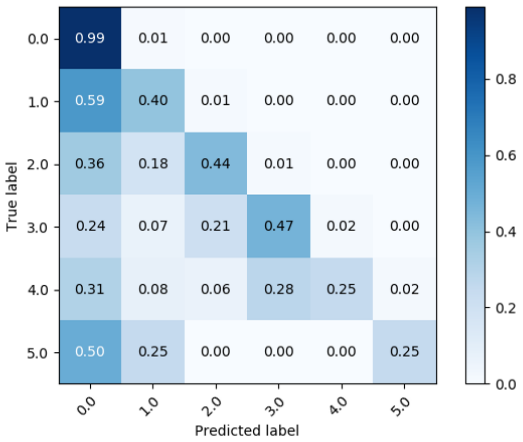

그림 4.3 분류의 성능평가를 위한 컨퓨전 행렬의 예시

4.4 결과

Methods	AP50	OIR	mIOU
Faster R-CNN 3D	69.08	0.781	0.545
Stacked Hourglass 3D	72.73	0.822	0.587
CenterNet 3D	81.82	0.847	0.695
Ours	**90.91**	**0.966**	**0.704**
Ours w/ margin	N/A	**0.99**	N/A

표 4.1 치아 탐지 실험 결과

실험은 본 연구에서 제안된 네트워크를 기존의 물체 탐지 분야 및 keypoint estimation 분야에서 가장 많이 사용되고 있는 딥러닝 모델들인 Faster R-CNN, Stacked Hourglass Network, CenterNet과 비교하며 진행하였다. 세 모델 모두 기본적으로는 2D 이미지 상에서만 동작하도록 설계되어 있어서 3차원으로 확장하여 실험하였다. Ablation study의 측면에서는 실험에 사용된 CenterNet과 제안된 모델은 구조적으로 동일한 네트워크를 기반으로 하고 있기 때문에, 이 둘의 차이를 통해 GD Loss의 효과를 입증할 수 있다. 성능지표로는 앞서 설명한 AP50, OIR, 그리고 평균 IOU(mIOU)를 사용하였다.

먼저, 치아 탐지 실험을 진행하였는데 결과는 표 4.1에 나타나 있다. 제안된 모델이 세 성능지표 모두에서 다른 네트워크들보다 높은 성능을 보이는 것을 확인할 수 있었다. 또한 탐지된 경계상자에 5에서 10 픽셀 정도의 적당한 margin을 추가하였을 경우 탐지된 치아를 완벽하게 포함하여 OIR의 값이 1로 나타나는 것이 확인되었다.

이어서 치아 분류에 있어서도 성능을 평가하였다. 탐지 및 분류가 완료된 치아 경계상자들이 실제로 그 위치에 치아를 잘 포함하고 있다면 true positive로,

Methods	Precision	Recall
Faster R-CNN 3D	0.806	0.731
Stacked Hourglass 3D	0.853	0.826
CenterNet 3D	0.891	0.883
Ours	**0.932**	**0.919**

표 4.2 치아 분류 실험 결과

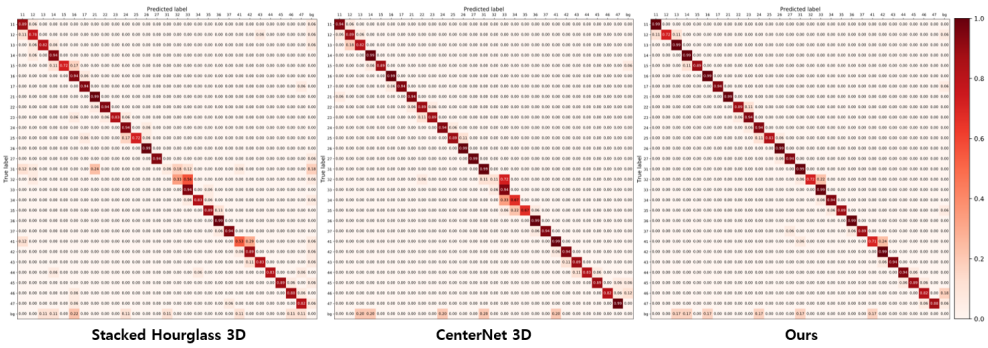

그림 4.4 치아 분류의 컨퓨전 행렬 결과

아니라면 false positive로 하여 정밀도 및 재현율을 계산하였다. 표 4.2에 나타나 있듯이 치아 분류에 있어서도 다른 네트워크들보다 좋은 성능을 보였다. 또한 치아 분류의 성능을 한눈에 확인하기 위해 컨퓨전 행렬로도 나타내었다(그림 4.4). 인접한 치아들 중에서도 특히 그 크기가 작은 31, 32, 41, 42번 치아들에서 어려움을 겪는 다른 네트워크들과는 달리, 제안된 네트워크에서는 대부분의 치아에서 컨퓨전 행렬의 대각성분 쪽에 값이 몰려있는 것을 확인할 수 있다.

실험 결과에 나타나 있듯이, 인접한 치아들로 인해 치아 탐지의 성능이 저하된다는 것을 알 수 있다. 이를 해결하기 위해 본 연구에서는 상술한 가우시안 분리 로스를 도입하였는데, 이는 heatmap regression 단계에서 인접한 치아들의

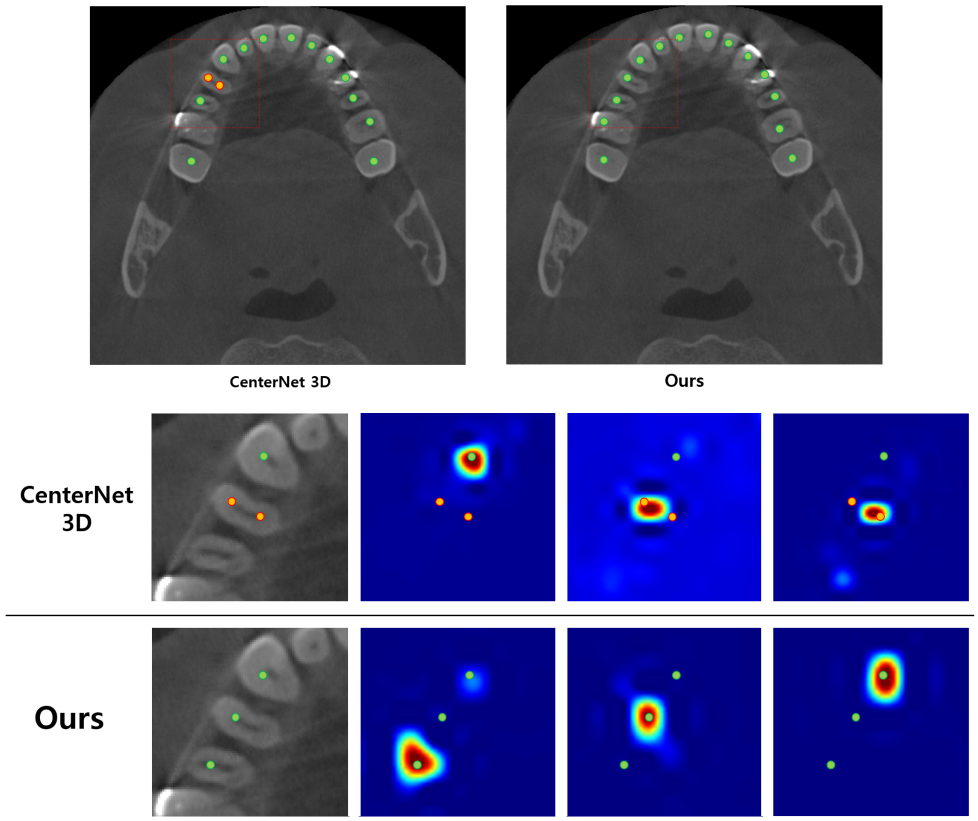

그림 4.5 가우시안 분리의 효과

가우시안 분포가 겹치지 않게 학습되는 효과가 있었다. 그림 4.5에 나타나 있듯이, CenterNet과 같은 기존의 Point 기반 물체 탐지 네트워크에서 한 치아에 두 개의 피크점이 몰려있는 경우가 있었다. 해당 부분의 heatmap을 시각화하여 살펴보면 가우시안 분포가 서로 가깝게 위치하고 있다. 하지만 가우시안 분리 로스를 적용하였을 때에는 heatmap 상의 가우시안이 성공적으로 분리되어 있는 모습이 확인되었다.

위의 치아 탐지 부분에서 언급했듯이, 경계상자의 크기에 소량의 margin을 추가하는 경우 치아를 완벽하게 포함하게 된다. 개별 물체 분할에서는 탐지된 경

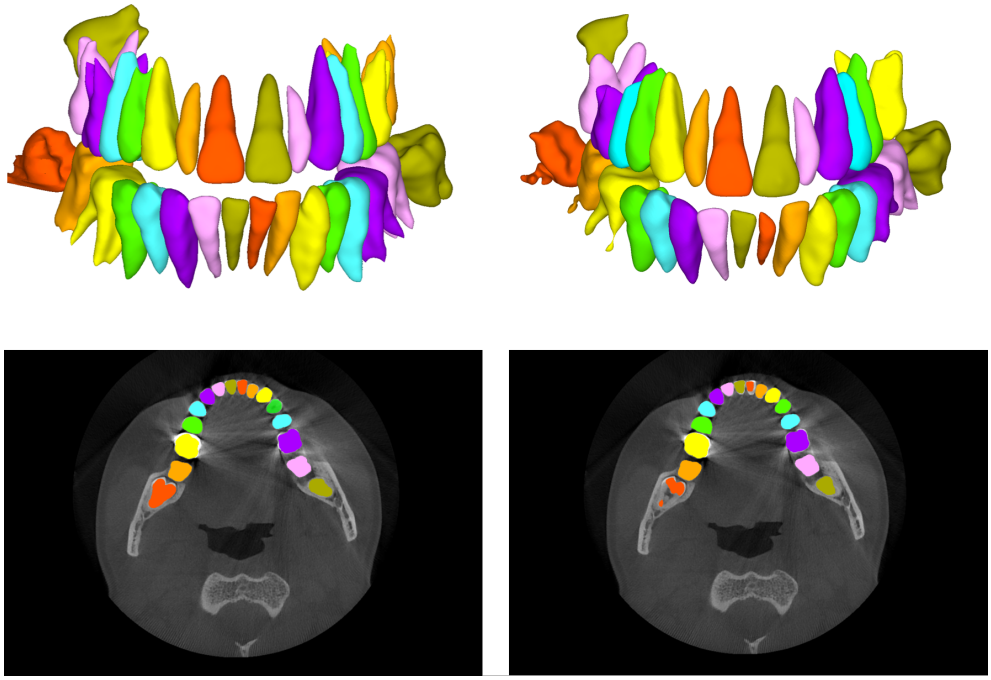

그림 4.6 (좌) 치아 영역 정답지 (우) 개별 치아 분할 결과

계상자들이 물체를 얼마나 잘 포함하고 있는가가 최종적인 분할의 성능에 큰 영향을 미치는데, 본 연구에서 제안된 네트워크는 치아 탐지의 성능에서 OIR이 1로 나타나기에 개별 치아 분할에서도 높은 성능을 보일 것으로 기대할 수 있다. 실제로 탐지된 치아 경계상자들에 대해 거리 기반 치아 분할도 학습하여 그 결과를 확인하였는데 그림 4.6에 나타나 있듯이 대부분의 치아에 대해 성공적으로 분할이 되는 것을 확인할 수 있었다.

제 5 장 결론

구강 CT 이미지에서는 일반적인 물체 탐지 및 분할 방법들만으로는 좋은 결과를 얻어내기가 쉽지 않다. 그래서 본 연구에서는 구강 CT 도메인에서 더욱 효과적으로 치아를 탐지 및 분할하기 위한 프레임워크를 제안하였다. 비슷한 형태의 치아들을 구분해야 하는 어려움을 극복하기 위해 치아 분류에 대한 추가적인 학습이 필요하지 않은 Point 기반 물체 탐지 방법을 채택하였고, 이미지의 위치 정보들을 보존하여 효율적인 학습이 가능한 heatmap regression 방식을 사용하였다. 특히 인접한 치아들의 위치를 정확히 구분하여야 하는 문제를 해결하고자 새로이 고안된 가우시안 분리 로스를 도입하여 정확도의 향상에 기여하였다. 이렇게 물체 탐지 부분에서 높은 성능을 보인 것에 힘입어, 이어진 치아 분할에서도 성공적인 결과를 얻었다.

개선사항으로는 Point 기반 물체 탐지 방법의 특성상 이미지에 존재하지 않는 치아, 즉 발치된 차아나 결손 치아에 대해서는 제대로 판단을 내리지 못한다는 점이 있다. 이를 해결하기 위해서는 탐지된 모든 치아 상자에 대해, 해당하는 상자에 실제로 치아가 존재하는지 아닌지에 대한 추가적인 분류를 수행하는 모듈을 추가하는 방법이 있다. 또한 탐지된 각 치아들이 3차원 상에 다양한 각도로 존재하는데, 이 치아들의 각도를 예측하여 모든 치아들을 지면에 수직한 방향으로 정렬을 한 후 분할을 진행한다면, 데이터의 일관성이 강화되어 더욱 높은 성능의 분할 결과를 기대할 수 있다.

참고문헌

[1] B. Van Ginneken, B. T. H. Romeny, and M. A. Viergever, "Computer-aided diagnosis in chest radiography: a survey," *IEEE Transactions on Medical Imaging*, vol. 20, no. 12, pp. 1228–1241, 2001.

[2] M. Chung, M. Lee, J. Hong, S. Park, J. Lee, J. Lee, I.-H. Yang, J. Lee, and Y.-G. Shin, "Pose-aware instance segmentation framework from cone beam ct images for tooth segmentation," *Computers in Biology and Medicine*, p. 103720, 2020.

[3] M. Chung, J. Lee, S. Park, M. Lee, C. E. Lee, J. Lee, and Y.-G. Shin, "Individual tooth detection and identification from dental panoramic x-ray images via point-wise localization and distance regularization," *arXiv preprint arXiv:2004.05543*, 2020.

[4] Z. Cui, C. Li, and W. Wang, "Toothnet: Automatic tooth instance segmentation and identification from cone beam ct images," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2019, pp. 6368–6377.

[5] A. Newell, K. Yang, and J. Deng, "Stacked hourglass networks for human pose estimation," in *European Conference on Computer Vision*. Springer, 2016, pp. 483–499.

[6] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2015, pp. 3431–3440.

[7] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," in *International Conference on Medical Image Computing and Computer-Assisted Intervention*. Springer, 2015, pp. 234–241.

[8] Z.-Q. Zhao, P. Zheng, S.-t. Xu, and X. Wu, "Object detection with deep learning: A review," *IEEE Transactions on Neural Networks and Learning Systems*, vol. 30, no. 11, pp. 3212–3232, 2019.

[9] P. F. Felzenszwalb, R. B. Girshick, D. McAllester, and D. Ramanan, "Object detection with discriminatively trained part-based models," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 9, pp. 1627–1645, 2009.

[10] D. G. Lowe, "Distinctive image features from scale-invariant keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, 2004.

[11] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," in *2005 IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR'05)*, vol. 1. IEEE, 2005, pp. 886–893.

[12] R. Lienhart and J. Maydt, "An extended set of haar-like features for rapid object detection," in *Proceedings. International Conference on Image Processing*, vol. 1. IEEE, 2002, pp. I–I.

[13] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 580–587.

[14] S. Ren, K. He, R. Girshick, and J. Sun, "Faster r-cnn: Towards real-time object detection with region proposal networks," in *Advances in Neural Information Processing Systems*, 2015, pp. 91–99.

[15] J. Redmon and A. Farhadi, "Yolo9000: better, faster, stronger," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2017, pp. 7263–7271.

[16] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You only look once: Unified, real-time object detection," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2016, pp. 779–788.

[17] J. Redmon and A. Farhadi, "Yolov3: An incremental improvement," *arXiv preprint arXiv:1804.02767*, 2018.

[18] H. Law and J. Deng, "Cornernet: Detecting objects as paired keypoints," in *Proceedings of the European Conference on Computer Vision (ECCV)*, 2018, pp. 734–750.

[19] X. Zhou, D. Wang, and P. Krähenbühl, "Objects as points," *arXiv preprint arXiv:1904.07850*, 2019.

[20] A. Toshev and C. Szegedy, "Deeppose: Human pose estimation via deep neural networks," in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, 2014, pp. 1653–1660.

[21] J. J. Tompson, A. Jain, Y. LeCun, and C. Bregler, "Joint training of a convolutional network and a graphical model for human pose estimation," in *Advances in Neural Information Processing Systems*, 2014, pp. 1799–1807.

[22] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask r-cnn," in *Proceedings of the IEEE International Conference on Computer Vision*, 2017, pp. 2961–2969.

[23] A. Rosenfeld and J. L. Pfaltz, "Distance functions on digital pictures," *Pattern Recognition*, vol. 1, no. 1, pp. 33–61, 1968.

Abstract

Individual tooth segmentation and identification from cone beam computed tomography images are preoperative prerequisites for orthodontic treatments. Recent studies that involve convolutional neural networks for instance segmentation, typically preceded by object detection and classification, show groundbreaking performances on different medical image domains. While point-based detection frameworks show significant outcomes on dental images, distinguishing adjacent teeth remains challenged due to their similar topology. In this study, we propose a point-based tooth localization network by introducing Gaussian Disentanglement loss. The proposed network first performs heatmap regression accompanied by box size regression for all anatomical teeth. A novel Gaussian Disentanglement penalty is employed by minimizing the sum of pixel-wise multiplication of heatmaps for all adjacent tooth pairs. Subsequently, individual tooth segmentation is applied by converting pixel-wise labeling task to distance map regression task in order to minimize false positives on adjacent teeth. The experimental results demonstrate that the proposed algorithm outperforms state-of-the-art approaches by increasing the average precision of detection by 9.1%, consequently leading to high performance of individual tooth segmentation.

Keywords: Object detection, Dental CT, Tooth segmentation

Student Number: 2019-27115

감사의 글

연구실에서 2년 동안 생활하며 많은 것을 배웠습니다. 수업을 듣기만 했던 학부 때와는 다르게 직접 문제에 대해 고민하고 해결해 나가는 과정이 결코 쉽지만은 않았지만, 감사한 분들 덕분에 무사히 마칠 수 있었다고 생각합니다.

먼저 신영길 교수님께 감사의 말씀을 전합니다. 학부생 시절부터 교수님을 만나게 되어 석사과정을 마칠 때까지 언제나 아낌없는 조언과 지도를 해주신 덕분에 많은 것을 배울 수 있었습니다.

또한 연구실에서 동고동락 하며 많은 추억을 쌓은 선후배님들께 감사의 인사를 드립니다. 언제나 연구에 최선을 다하시고 이번에 같이 졸업하게 되어 영광인 지강 형, 지선 누나께 감사드립니다. 자신만의 길을 멋지게 개척해 나가시는 민규 형, 코로나로 힘든 상황 속에서 랩장으로서 연구실을 잘 이끌어주신 지완 형, 항상 후배들에게 모범이 되고 아낌없는 조언을 해주시는 민경 누나께 특별히 감사의 말씀을 전하고 싶습니다. 언제나 연구실의 큰 기둥이 되어주는 민창이, 비록 다른 연구실에서 생활하게 되었지만 멀리서도 응원을 아끼지 않으신 진규 형께 감사드립니다. 항상 묵묵히 자기만의 연구를 멋지게 하고 계신 승환 형, 경휘 누나께 감사와 응원의 말을 전하고 싶습니다. 매사에 열심히 임하고 항상 연구실 분위기를 밝게 해주는 든든한 후배 채은이와 재용이에게도 감사의 말을 전합니다.

이번에 같이 졸업하게 된 강용 형과 상욱이에게도 감사의 말과 응원의 말을 전하고 싶습니다. 입학부터 졸업까지 매일을 함께하며 바로 옆자리에서 서로 힘이 되어준 상욱이 덕분에 2년 동안의 연구실 생활을 무사히 마칠 수 있었다고 생각합니다. 기쁠 때나 슬플 때나 항상 곁에서 아낌없는 격려와 응원을 해주신 강용 형께도 특별히 감사드립니다.

연구실 생활을 하는 동안 제 인생에 가장 큰 영향을 주신 민영 형께 깊은 감사의 말을 전합니다. 후배들이 자신만의 연구를 할 수 있게 진심을 다해 지도해주시고 어려운 문제에 봉착했을 때에도 자신감을 잃지 않게 응원해주셨기에 언제나 마음이 든든했습니다. 앞으로 교수가 되셔서도 학생들에게 좋은 가르침을 주시는 훌륭한 교수가 되실 거라 믿어 의심치 않습니다.

마지막으로 항상 변함없는 믿음과 사랑을 주시는, 세상에서 가장 존경하는 부모님께 감사와 사랑의 마음을 전합니다.