

구강 스캐너 영상 세그멘테이션을 위한 실시간 딥 모델 개발

유 호 진*, 조 항 재*, 김 재 곤**, 전 경 구°

Real-Time Deep Model for Intraoral Scanner Image Segmentation

Hojin Yoo*, Hangjae Cho*, Jaegon Kim**, Kyungkoo Jun°

요 약

최근 치과에서 구강 스캐너를 사용한 3차원 치아 모형 제작 시스템 사용이 증가하고 있다. 스캐너는 스캔 된 여러 장의 2차원 영상을 이용하여 3차원 모델을 재구성한다. 이 과정에서 세그멘테이션을 이용하여 재구성에 불필요한 손가락, 혀 등의 물체를 제거한다. 모형 구성의 정확도를 위해서 시각화가 실시간으로 진행되어야 하기에, 세그멘테이션 과정에 주어지는 시간은 제한적이다. 실시간 세그멘테이션 관련 기존 연구들은 브랜치 구조를 통해 적은 수의 컨볼루션으로 정확도의 감소를 줄이면서 실시간 요구사항을 만족시켜왔다. 본 연구에서는 브랜치 구조의 존성을 줄여 추론 시간을 감소시키는 모델, IntraOralScanner Network (IOSNet)을 제안한다. 브랜치를 줄이되, 남은 브랜치의 레이어를 보강하여 속도 향상과 정확도를 유지하였다. 이를 통해, 기존 실시간 모델인 PIDNet-small보다 추론 속도는 62.2% 향상시켰고(NVIDIA GTX 1060 기준), 정확도는 mIoU 기준 0.929로 기존보다 2.7% 감소에 불과했다. 상용 스캐너 제품 출시에 필요한 세그멘테이션 수행 시간 목표인 10ms 내(동일 GPU 기준), 30 FPS를 달성하였다.

Key Words : Intraoral Scanner, Real-time Segmentation, Branch, Deep Model, 3D Modeling

ABSTRACT

Recently, the use of intraoral scanners for creating three-dimensional models of teeth has been increasing in dental practices. These scanners reconstruct 3D models using multiple 2D images obtained through scanning. In this process, segmentation is used to remove unnecessary objects such as fingers and tongues from the reconstruction. Real-time visualization is crucial for achieving the accuracy of the model construction; thus, the time given to the segmentation process is limited. Previous studies on real-time segmentation have achieved real-time requirements by reducing the number of convolutions using a branch structure while minimizing the decrease in accuracy. In this study, we propose the IntraOralScanner Network (IOSNet), a model that reduces branch structure dependency to decrease inference time. We improved the remaining branches' layers to enhance speed and maintain accuracy, achieving a 62.2% increase in inference speed (compared to NVIDIA GTX 1060) and only a 2.7% decrease in accuracy, as measured by mIoU, compared to the existing real-time model PIDNet-small. We achieved the segmentation time goal required for commercial scanner product release, i.e., 10ms (on the same GPU) and 30 FPS.

* 이 논문은 인천대학교 2021년도 자체연구비 지원(2021-0093)에 의하여 연구되었습니다.

• First Author : Dept. of Embedded Systems Engineering Incheon National University, yoohj0416@inu.ac.kr, 정회원

° Corresponding Author : Dept. of Embedded Systems Engineering Incheon National University & Energy Excellence & Smart City Lab., kjun@inu.ac.kr, 정회원

* Dept. of Embedded Systems Engineering Incheon National University, hangjae99@inu.ac.kr, 학생회원

** Dept. of Industrial Engineering Incheon National University, jaegkim@inu.ac.kr, 종신회원

논문번호 : 202305-091-C-RN, Received April 28, 2023; Revised July 13, 2023; Accepted July 20, 2023

1. 서론

구강 스캐너(Intraoral Scanner)는 치아의 표면을 촬영하고 얻은 거리정보를 바탕으로 3D 포인트 클라우드(point cloud) 형태의 치아 모형을 생성하는 기기를 의미한다^[1]. 구강 스캐너는 환자의 치아와 직접적으로 접촉하지 않고 정보를 획득할 수 있기에 환자 불편함 감소, 시간 절약, 치료 과정 간소화와 같은 장점이 존재한다. 또한, 스캔한 치아 모형을 실시간으로 확인하여 치아의 크기, 구조 등을 쉽게 파악할 수 있다^[2].

구강 스캐너의 동작 원리는 다음과 같다. 먼저, 패턴 모양의 레이저(laser)를 치아 표면에 투영시키고, 센서(sensor)를 이용하여 반사된 레이저가 돌아오는 데 걸린 시간을 측정한다. 측정된 시간에 삼각측량법을 적용해 치아 표면과의 거리를 구한다. 위의 동작을 빠른 속도로 반복하여 얻은 여러 거리정보를 재구성하여 3D 포인트 클라우드를 완성한다.

치아 포인트 클라우드를 구성하기 위해서는 촬영된 이미지에 세그멘테이션(segmentation) 적용을 적용해야 한다. 이는 불확실한 물체가 촬영되면 포인트 클라우드 구성을 방해할 가능성이 있기 때문이다. 예를 들어, 구강 내에서 치아와 잇몸의 위치는 고정되어 있어서, 스캐너가 측정하는 거리정보에 큰 변화가 발생하지 않아 구성에 문제가 발생하지 않는다. 하지만, 입 벽, 혀, 손가락과 같이 측정 중에 위치가 변할 수 있는 물체는, 포인트 클라우드를 재구성하는 과정에서 오류를 일으킬 가능성이 존재한다. 또한, 이러한 물체들의 정보는 치아 치료에서 활용되지 않는 정보이기에 제외하는 것이 적절하다. 그림 1과 같이 불확실한 물체를 제거하면 사용자가 원하는 정보만을 획득할 수 있다.

하지만, 구강 스캐너에 세그멘테이션을 적용하는 데

는 한계가 존재한다. 첫 번째, 실시간으로 이루어져야 한다는 점이다. 사용자가 치아 모형이 구성되는 과정을 실시간으로 확인하고, 구성이 잘 이루어지는지 알 수 있어야 하기 때문이다. 두 번째, 컴퓨팅 자원이 제한적이다. 병원에서는 주로 스캐너를 노트북에 연결하여 사용하기 때문에 컴퓨팅 자원이 한정적이다. 이러한 제약 조건에서 알고리즘으로 동작하는 세그멘테이션을 적용하면 정확도가 크게 떨어질 수 있다는 문제가 있다. 따라서, 실시간으로 세그멘테이션이 이뤄질 수 있는 딥러닝 모델을 적용해야 한다. 거리정보를 3D 포인트 클라우드로 재구성하는 과정에서도 컴퓨팅 자원을 사용하기 때문에, 이 역시 고려되어야 한다.

본 논문의 기여 사항은 다음과 같다.

- 기존에 연구되었던 실시간 세그멘테이션 모델의 구조가 구강 스캐너에 적절하지 못한 이유를 모델 조정을 통해 알아내었다.
- 구강 스캐너에서 적용 가능한 실시간 세그멘테이션 모델 IOSNet을 제안하였다.
- 실험을 통해 제안한 모델이 구강 스캐너에 적용 가능함을 확인하였다.

구강 스캐너 실시간 세그멘테이션 모델을 제안하는 과정을 다음과 같다. 먼저 2장 구강 스캐너 세그멘테이션 모델과 관련된 논문들을 소개하고, 3장에서는 구강 스캐너에 실시간 세그멘테이션 모델을 적용하기 위해 문제점을 파악하고 해결 방법을 제안한다. 4장에서는 제안한 모델이 구강 스캐너에서 사용하기에 적절한지 확인하기 위한 실험 과정과 결과를 보여준다. 마지막으로 5장 결론에서 마무리 짓는다.

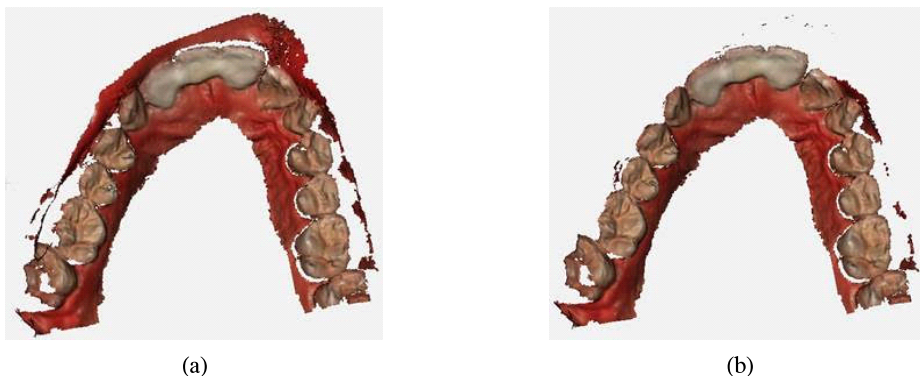


그림 1. 구강 스캐너로 생성된 치아 모형 (a) 세그멘테이션 적용 전 (b) 세그멘테이션 적용 후
 Fig. 1. Dental models generated by an intraoral scanner. (a) Before segmentation. (b) After segmentation.

II. 본 론

2.1 관련 연구

1절에서는 구강 스캐너 실시간 세그멘테이션 모델과 관련된 연구들에 대해 소개하고자 한다. 먼저, 1.1항에서는 치의학 분야에 활용된 인공지능 기술들을 소개한다. 1.2항에서는 의미론적 세그멘테이션(semantic segmentation) 모델들이 연구되었던 과정들을 보여주고자 한다. 1.3항에서는 의미론적 세그멘테이션을 실시간으로 처리하기 위해 노력했던 방법들을 소개하고, 본문에서 제안하는 모델의 기반이었던 PIDNet^[3]에 대해 간단하게 설명하고자 한다.

2.1.1 치의학 인공지능 활용

최근 치의학의 다양한 분야에서 인공지능 기술이 주목받아 활용되고 있다. 기존에는 의사가 치아를 직접 확인하고 진단하였지만, 인공지능을 활용하여 진단, 분석, 그리고 시술을 보조하는 방법들이 연구되고 있다. 예를 들어, 방사선 사진을 이용해 정중과잉치(Mesiodens)를 검출하는 기술^[4]이 제안되었다. 정중과잉치는 비정상적으로 위치하고 있는 치아를 의미한다. 또한, 방사선 사진에서 치근단(Periodontally Compromised Teeth, PCT) 및 치주골의 손상 여부를 판단하는 방법^[5]이 연구되었다. 해당 연구에서의 인공지능은 짧은 시간 내에 방사선 전문의의 진단 내용과 비슷하게 판단한 것으로 확인되었다.

치아 보철물 제작 분야에서도 인공지능 기술이 활발히 적용되고 있다. 손실된 치아의 템스맵(depth map)을 만들어 생성 모델의 입력으로 사용해 보철물의 디자인을 추천하는 방법^[6-9]이 활발하게 연구되고 있다. 보철물 제작 과정에서 발생할 수 있는 문제를 예측하고 보완하는 기술^[10]도 연구되고 있다. 이는 사람의 개입으로 나타날 수 있는 실수와 문제를 미리 탐지함으로써 보철물 제작 속도 및 정확도 향상에 기여하고 있다.

2.1.2 의미론적 세그멘테이션 모델

의미론적 세그멘테이션은 입력 이미지의 모든 픽셀을 해당하는 클래스 레이블에 매핑(mapping)하는 작업이다. 세그멘테이션 모델은 기본적으로 인코더-디코더(Encoder-Decoder) 구조를 사용해 각 픽셀에 대한 클래스를 구한다. 지금까지 연구된 세그멘테이션 모델은 크게 두 가지로 나눌 수 있다. 단계마다 하나의 컨볼루션 레이어(convolution layer)를 사용하는 단일 컨볼루션 방법과, 한 단계에서 여러 개의 레이어를 사용하고 정보를 합친 후 추론하는 특징 피라미드 네트워크(Feature

Pyramid Network, FPN)^[11]를 적용한 방법이다.

FCN^[12]은 이미지 분류에 사용되던 컨볼루션 신경망(Convolutional Neural Network, CNN)을 의미론적 세그멘테이션에 사용하기 위해 최초로 고안된 모델이다. 이미지 분류 모델의 완결 연결 계층(Fully Connected Layer, FCN) 대신 디코더를 사용하여 모델의 출력 크기를 입력 이미지와 일치시켰다. U-Net^[13]은 같은 크기의 특징 맵끼리 연결하는 스킵 연결(skip connection)을 사용하여 이전 레이어의 지역적 정보를 활용해 학습하고 추론하는 방법을 제안하였다. SegNet^[14]은 다운 샘플링(down-sampling) 단계에서의 정보를 업 샘플링(up-sampling) 레이어에서 사용해 공간적 정보의 손실을 최소화하는 방법을 연구하였다.

최근에는 인코더에 FPN을 적용하여 여러 크기의 물체에 대응하는 모델을 연구하고자 하였다. PSPNet^[15]은 기존 단일 컨볼루션 레이어가 아닌 FPN을 사용해 세그멘테이션을 진행하여 정확도를 증가시킨 모델이다. DeepLab V2^[16]은 팽창(dilated) 컨볼루션 레이어를 사용해 여러 크기의 이미지를 한 단계에서 처리하는 효과를 내는 Artous Spatial Pyramid Pooling (ASPP)를 통해 정확도를 향상시키는 방법을 제안하였다. DeepLab V3+^[17]는 다양한 컨볼루션 레이어를 적용한 Xception 모델^[18]에 ASPP를 적용하고 물체의 작은 부분을 감지하는 분기를 만들어 세밀하게 세그멘테이션하는 모델은 연구하였다.

2.1.3 실시간 세그멘테이션 모델 PIDNet

PIDNet^[3]은 출력값을 측정하여 오차를 계산한 후 필요한 제어값을 구하는 Propotional-Integral-Differential (PID) 제어기의 개념을 Convolutional Neural Networks (CNN)에 적용한 실시간 세그멘테이션(segmentation) 모델이다. PIDNet은 기존의 실시간 세그멘테이션 모델들이 채택한 인코더-디코더(encoder-decoder) 구조^[19-21]를 사용하지 않고 브랜치(branch) 구조^[22-25]를 사용한다. 하지만, 지금까지 브랜치 구조를 채택한 모델들은 2개의 브랜치를 사용하여 세그멘테이션을 수행하였다. 이 모델들은 작은 크기 물체들의 경계선 부분 픽셀이 붕괴되는, 이른바 오버슈트(overshoot) 현상이 발생하는 경향이 있다. PIDNet은 3개의 브랜치를 사용해 이 현상을 보완하고자 하였다.

PIDNet에서 사용한 3개의 브랜치는 다음과 같다. Propotional (P) 브랜치, Integral (I) 브랜치, Derivative (D) 브랜치. P 브랜치에서는 고해상도의 특징맵에서 세부적인 정보를 분석하고 유지하는 역할을 담당한다. I 브랜치는 지역적 또는 전체적 문맥(context) 정보를 유

지한다. 마지막으로, D 브랜치는 경계선을 예측하기 위한 특징을 추출하기 위한 구간이다. D 브랜치는 물체의 경계선을 예측하도록 학습되어야 하기에, ground truth (GT) 에서 경계선을 검출하여 D 브랜치 학습에 사용한다. 또한, 각 브랜치에 포함된 정보를 융합하기 위한 Boundary - attention-guided fusion (Bag) 모듈을 제안하여 3개의 브랜치를 가진 실시간 세그멘테이션 모델을 완성하였다.

2.2 실시간 세그멘테이션을 위한 모델 경량화

2절에서는 실시간 세그멘테이션 모델을 구강 스캐너에 적용하기 위해 기존 모델의 문제점을 파악하고 해결 방안을 적용한 모델을 제안하고자 한다. 모델을 제안하기 전에, 2.1항에서 본 연구에서 구강 스캐너 이미지에 세그멘테이션 모델을 사용하는 이유와 과정을 설명한다. 2.2항에서는 이전 모델인 PIDNet이 실시간 세그멘테이션 모델로 동작하는데 발생하는 한계점을 소개하고, 2.3장에서 이를 해결하기 위한 방법을 제안한다.

2.2.1 구강 스캐너 영상 세그멘테이션

세그멘테이션을 구강 스캐너 이미지에 적용하는 이유는 치아 모형을 구성할 때 방해가 되거나 불필요한 물체를 제외하기 위함이다. 구강을 스캔할 때 치아와 잇몸은 위치가 크게 변하지 않기 때문에 스캐너의 거리정보를 획득하는 데 어려움이 없다. 하지만, 혀, 입 벽, 손가락 등과 같이 동일한 부분이 스캔 될 때마다 위치가 변하는 물체는 거리정보 또한 변하기 때문에, 치아 모형 구성에 방해가 된다. 또한, 사용자가 구강 스캐너를 통해 얻고자 하는 정보는 치아와 잇몸의 3D 모델이기에 해당 물체를 이미지에서 제거하는 과정이 필요하다.

구강 스캐너 세그멘테이션 적용 과정은 그림 2와 같다. 첫 번째, 구강 스캐너에서 생성된 베이어(bayer) 이미지를 세그멘테이션 모델의 입력으로 사용한다. 세그멘테이션 모델은 이미지의 각 픽셀을 이진 분류하는 모델이다. 이진 분류는 제거 대상인 타겟(target)과 제거

대상이 아닌 것에 해당한다. 두 번째, 세그멘테이션 결과를 이용해 타겟 마스크(mask)를 생성한다. 마지막으로, 타겟 마스크를 스캐너에 수집된 거리정보에 반영하여 불필요한 물체를 제거하며 치아 모형을 구성한다.

2.2.2 PIDNet 경량화의 한계점

PIDNet은 컨볼루션 레이어로 이루어진 3개의 브랜치 구조를 하고 있다. 이 브랜치 구조가 모델의 추론 시간이 증가하는 결과를 초래하고 있다. 그 이유는 각각의 브랜치가 순차적으로 계산을 하기 때문이다. 예를 들어 I 브랜치 첫 번째 단계를 계산하고 나면, P 브랜치의 첫 번째 단계, 그리고 D 브랜치 레이어의 첫 번째 단계 순으로 순차적 계산이 이뤄져야 한다. P와 D 브랜치는 I 브랜치의 정보를 활용하기 때문에, 병렬 처리가 불가능하고 순차적인 추론이 필수적이다.

순차적으로 계산하는 데 있어 발생하는 다른 문제는 그래픽 카드의 캐시 메모리(cache memory)를 적극적으로 활용하지 못한다는 점도 있다. 각각의 브랜치 계산에 사용되는 정보가 달라 매번 메모리에서 데이터를 불러오는 데 많은 시간을 소모하기 때문이다. 만약 하나의 브랜치만 계산한다고 했을 때는, 계산 결과들이 그대로 캐시 메모리에 저장되어 재활용될 수 있어서 추론 시간을 줄이는데 이점이 있다. 이러한 원인으로 인하여 PIDNet을 추론하는데 부가적인 지연시간이 발생하고, 짧은 시간 내에 처리가 이뤄져야 하는 구강 스캐너에 적용하기에 부적절하다.

2.2.3 Branch 제거

불필요한 지연시간을 없애고 추론 시간을 줄이기 위해 PIDNet에서 브랜치를 그림 3과 같이 제거한 모델인 Intra Oral Scanner Network (IOSNet) 를 제안한다. IOSNet은 PIDNet에서도 small에 해당하는 모델을 기반으로, P 브랜치와 D 브랜치를 제거하고 I 브랜치를 사용해 구성하였다. Parallel Aggregation Pyramid Pooling Module (PAPPM)과 세그먼트 헤드(segment head)는 동일하게 사용하였다.

브랜치를 제거로 인해 기존 PIDNet보다 정확도가 떨어질 가능성이 존재한다. 하지만, 구강 스캐너에 필요한 세그멘테이션 태스크는 타 세그멘테이션 태스크에 비해 간단하기에 정확도 하락을 상쇄할 수 있다. 예를 들어, 도로의 물체들을 구별하는 CamVid 데이터셋^[26]은 가로등, 사람 등의 물체의 크기가 작아 식별하기 어려운 경우가 존재한다. 그에 반해 구강 스캐너 사진에는 등장하는 물체들은 상대적으로 작지 않다. 또한, 두 가지로 세그멘테이션 하는 구강 스캐너에 비해, 식별해야

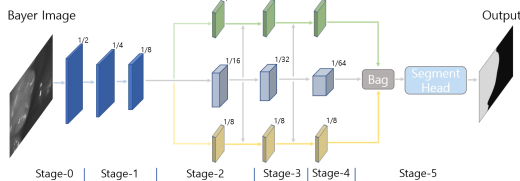


그림 2. 구강 스캐너에 실시간 세그멘테이션 모델 PIDNet을 적용하는 과정
Fig. 2. Applying real-time segmentation model PIDNet to the intraoral scanner.

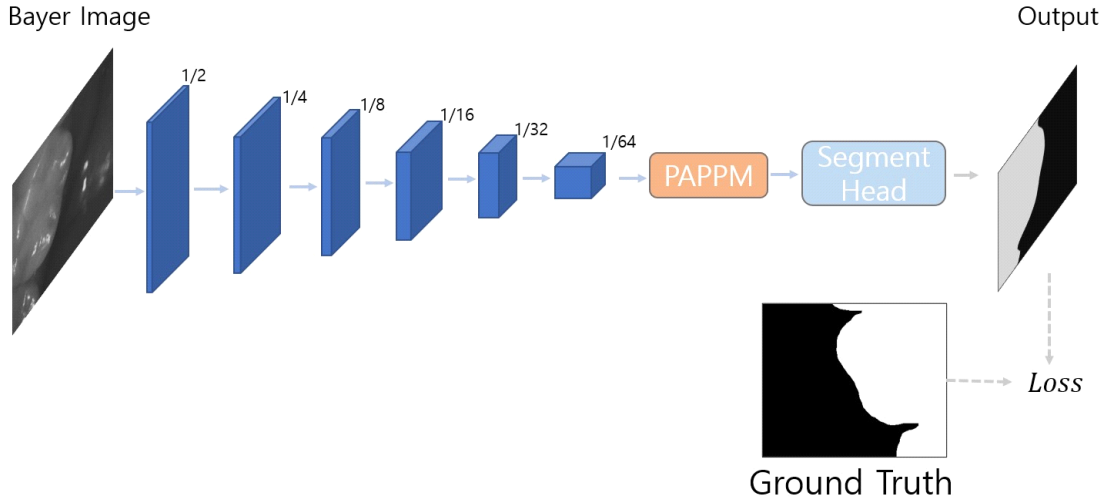


그림 3. 제안된 모델 IntraOralScanner Network (IOSNet) 의 기본 구조 개요
 Fig. 3. An overview of the basic architecture of proposed IntraOralScanner Network (IOSNet).

하는 카테고리의 개수가 32개로 태스크의 난이도가 존재한다.

PIDNet 보다 구조가 간단한 모델을 사용하지 않고 PIDNet을 경량화하여 사용한 이유는 PIDNet이 실시간 세그멘테이션에 적합한 구조를 하고 있기 때문이다. 기존의 세그멘테이션 모델들은 인코더-디코더 구조를 사용한다. 하지만, 이 구조는 인코더를 통해 생성해낸 특징 맵을 디코더를 이용해 입력 크기와 동일하게 만들어야 하기 때문에 수행 시간이 늘어날 수밖에 없다. 그에 반해 PIDNet은 모델의 출력에 단순히 보간법을 적용하여 세그멘테이션을 수행하는 방법을 채택해 수행 시간을 획기적으로 단축하였다. 단순히 컨볼루션 레이어를 이어 붙이는 방법이 아닌, 문맥 (context) 정보를 효율적으로 전달할 수 있는 PAPPM과 세그먼트 헤드 구조를 PIDNet에서 가져왔다. 이와 더불어, 추론 시간에 악영향을 미치는 브랜치만을 제거함으로써 구강스캐너에 적용가능한 실시간 세그멘테이션 모델인 IOSNet을 구성하였다.

두 개의 브랜치를 제거함으로써 2.2항에서 언급한 단점들을 보완하면서 속도를 향상시켰다. 하지만, 브랜치의 제거는 정확도 하락으로 이어지기에, 남은 브랜치의 컨볼루션 레이어를 보완하여 정확도 감소를 최소화 하였다. 이를 통해, 구강 스캐너 세그멘테이션 태스크에 만족할 수 있는 정도의 정확도를 이뤄낼 수 있는 IOSNet을 완성하였다. IOSNet은 한 장의 사진을 10ms 내에 처리해야 하는 구강 스캐너의 세그멘테이션 모델로 적합하다고 판단하였다. 모델의 학습은 식 (1)과 같

이 cross entropy loss를 사용해 진행하였다.

$$Loss = -\frac{1}{n} \sum_{i=1}^n \frac{1}{l} \sum_{j=1}^l \sum_{k=1}^c t_k \log p_{i,j,k} \quad (1)$$

식에서 l 은 이미지의 픽셀 개수를 의미하고, c 는 분류해야 할 클래스의 개수를 의미한다. t_k 는 k 번째 클래스에서의 정답 값을 의미하고 0 또는 1 값을 가진다. p 는 예측한 값을 의미한다.

III. 실험

3.1 구강 스캐너 동작 방법 및 구동 환경

본 연구의 실험에서는 유무선 겸용 방식의 구강 스캐너를 사용하였다. 해당 스캐너는 구조화된 조명 패턴 (structured light pattern)을 이용해 촬영된 물체의 거리를 측정한다. 스캐너에 탑재된 카메라는 384 * 320 픽셀(pixels) 크기의 이미지를 600 프레임 레이트(frame per second, fps)로 촬영이 가능하다. 촬영 과정의 90% 정도를 거리정보 수집을 위해 사용하고, 나머지는 치아의 컬러 매핑(color mapping)을 위해 비트맵(bitmap) 형태의 베이어(bayer) 이미지를 촬영한다. 스캐너는 Windows 환경에서 동작하고, 컴퓨팅 자원은 Intel Core i5-6500와 NVIDIA GTX 1060 6GB를 사용한다.

3.2 데이터셋 및 하이퍼파라미터

실험을 위해 구강 스캐너를 사용해 한국 성인 35명

을 대상으로 구강 내부를 직접 촬영하였다. 약 4만 장의 비트맵 사진을 수집하였고, 연속된 사진을 10장씩 샘플링(sampling)하여 4,618장의 사진을 데이터로 사용하였다. 이 중 80%인 3,694장을 임의로 선택하여 학습 데이터로 사용하고 나머지 924장을 테스트 데이터로 사용하였다.

수집된 이미지는 OpenCV에서 공개한 이미지 레이블링 툴인 Computer Vision Annotation Tool (CVAT)를 사용하여 세그멘테이션 GT를 생성하였다. CVAT를 로컬 환경에 설치하고 도커(Docker)를 이용해 구동 후 작업을 진행하였다. 레이블은 구강 스캐너 이미지에서 제거되어야 하는 부분인 혀, 입벽, 손가락 등을 타겟으로 분류하였다. 그리고 치아 3D 모델 구성에 필수 요소인 이빨과 잇몸을 논-타겟(non-target)으로 분류하였다. 학습 데이터 및 테스트 데이터 예시 사진은 그림 4와 같고, 클래스별 데이터 개수는 표 1과 같다.

데이터는 입력 데이터로 사용하기 전에 384 * 384 픽셀 크기로 제로 패딩(zero padding)을 진행하고 데이터 증강을 위해 랜덤 플립(random flip)을 적용한다. 학습 시 배치(batch) 크기는 6, 최적화 알고리즘은 확률적 경사 하강법(Stochastic Gradient Descent, SGD)를 사용하고, 학습률(learning rate)은 0.001로 설정하였다. ImageNet^[27]으로 사전 학습된 파라미터를 가진 모델을

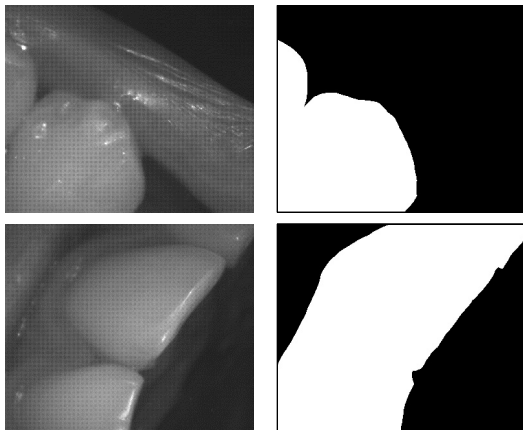


그림 4. 구강 스캐너 베이퍼 이미지 예시와 Ground Truth (왼쪽: 원본 이미지, 오른쪽: Ground Truth)
Fig. 4. Example of intraoral scanner bayer image and Ground Truth (Left: original image, Right: Ground Truth)

표 1. 클래스별 데이터 개수
Table 1. Number of data per class.

| Class Name | Train | Test |
|------------|-------|-------|
| Non-target | 7,183 | 1,776 |
| Target | 3,191 | 785 |

총 50 에폭(epoch)을 수행하였다.

3.3 학습 결과 비교 방법

IOSNet이 구강 스캐너의 실시간 세그멘테이션 모델로 적절하다는 것을 알아보기 위해 다른 모델과 두 가지 항목에 대해 평가를 진행한다. 첫 번째 평가 항목은 추론 속도이다. 구강 스캐너는 치아 모형 재구성 알고리즘이 동작하는 데에 컴퓨팅 자원을 사용하기 있기에, 세그멘테이션 모델 적용에 적절한 자원 분배가 필요하다. 따라서, 목표하는 추론 속도 이하의 모델을 사용하는 것이 필수이다. 두 번째 평가 항목은 세그멘테이션 정확도이다. 모델의 추론 속도가 목표 이하여도 정확도가 만족하지 못할 수준이라면 구강 스캐너에 적용할 수 없다. 정확도 평가 지표로는 [28]에서와 같이 mean Intersection over Union (mIoU)를 사용한다.

타 모델과의 비교는 SFNet-ResNet18, PP-LiteSeg-T1, PIDNet-small, PIDNet-tiny, PIDNet-nano와 진행하였다. SFNet은 특징 매칭(feature matching)을 통해 모델이 추출한 특징의 의미를 학습할 수 있도록 하였다. PP-LiteSeg는 디코더의 가용되는 컨볼루션 레이어의 채널 수를 줄여 실시간 세그멘테이션이 가능하도록 만든 모델이다. PIDNet-small은 PIDNet에서 컨볼루션 레이어를 가장 적게 사용한 모델이다. PIDNet-tiny와 PIDNet-nano는 추론 시간 단축을 위해 PIDNet-small의 컨볼루션 레이어 개수를 조정된 모델이다. PIDNet-small에서 컨볼루션 레이어 기준 채널 수, 스테이지 3과 4에서의 블록 개수, PAPPM의 채널 수를 조절하였다. 각 부분에서의 조절 세부 사항은 표 2와 같다.

표 2. PIDNet 컨볼루션 레이어 및 블록 개수 세부 사항
Table 2. Details of the number of convolution layers and blocks in PIDNet.

| Model | #Reference Ch. | #Blocks (Stage 3, 4) | #PAPPM Ch. |
|--------------|----------------|----------------------|------------|
| PIDNet-small | 32 | 3 | 96 |
| PIDNet-tiny | 24 | 2 | 64 |
| PIDNet-nano | 16 | 2 | 32 |

3.4 추론 시간 비교

IOSNet과 다른 모델들의 구강 스캐너의 구동 환경에서의 추론 속도를 측정하여, IOSNet이 구강 스캐너의 실시간 세그멘테이션 모델로 적절함을 알아보았다. 모델별 추론 속도, 모델 크기, 그리고 모델의 연산량을 나타내는 GFLOPs는 표 3과 같다. SFNet, PP-LiteSeg

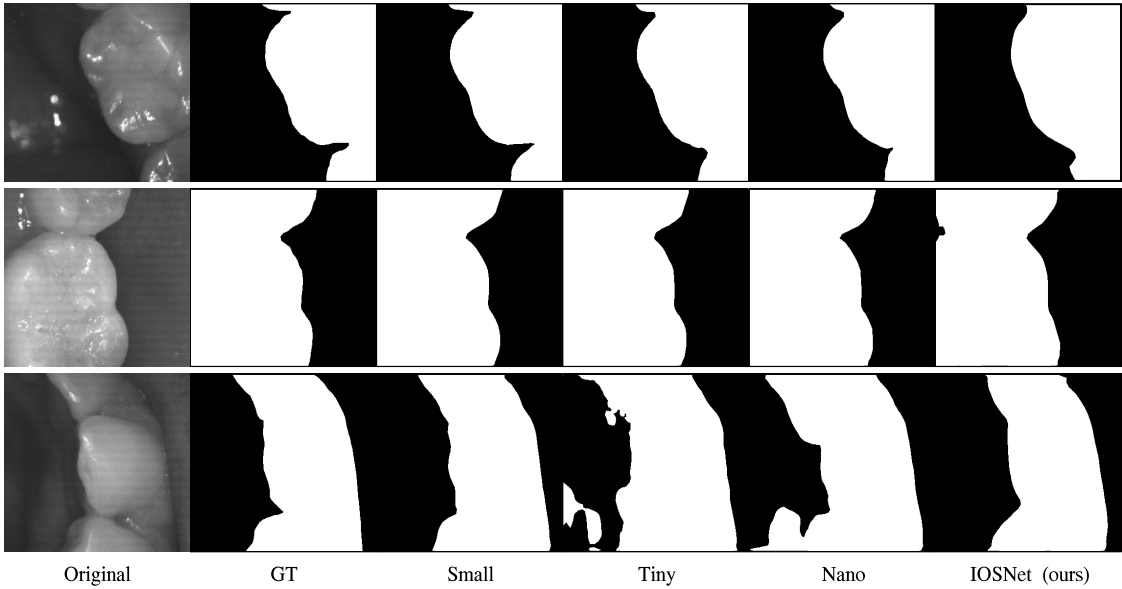


그림 5. 모델별 구강 스캐너 이미지 세그멘테이션 결과 비교
 Fig. 5. Comparison of intraoral scanner image segmentation results by model.

는 GFLOPs와 파라미터 개수에 비해 성능이 기준치에 못 미쳐 추론 시간 측정 비교는 제외하였다.

추론 시간 측정 결과 PIDNet-small부터 PIDNet-nano까지 순서대로 시간이 감소하였다. 이는 모델에 사용된 컨볼루션 레이어의 개수와 그 채널 수가 줄어들었기 때문이다. 제안된 모델인 IOSNet은 FLOPs와 모델 크기 모두 기존 PIDNet-small 보다 큰 것을 알 수 있다. 이는 컨볼루션 레이어를 보완하였기 때문이다. 그럼에도 불구하고 추론 시간이 네 개의 모델 중에서 가장 낮은 것을 확인할 수 있다. 이는 브랜치의 존재가 모델 추론 시간에 큰 영향을 끼친다는 것을 알 수 있다.

표 3. 실시간 세그멘테이션 모델 추론 시간, GFLOPs 및 모델 크기 비교 GFLOPs는 384 * 384 해상도 입력 기준. (굵은 글씨: 가장 빠른 추론 시간)
 Table 3. Comparison of inference time, GFLOPs, and model size for real-time segmentation models. GFLOPs are based on 384*384 resolution input. (bold: the fastest inference time)

| Model | Time (ms) | #GFLOPs | #Params | Size (MB) |
|----------------|------------|---------|---------|-----------|
| SFNet-resnet18 | - | 32.04 | 13M | 52.6 |
| PP-LiteSeg-T1 | - | 26.89 | 8M | 31.1 |
| PIDNet-small | 15.9 | 3.56 | 7.72M | 31.3 |
| PIDNet-tiny | 14.6 | 1.91 | 3.5M | 14.3 |
| PIDNet-nano | 13.7 | 0.40 | 1.33M | 5.4 |
| IOSNet (ours) | 9.8 | 2.15 | 9.76M | 39.6 |

IOSNet의 추론 시간은 구강 스캐너가 실시간 처리를 하기 위해 요구되는 시간인 10ms 내에 도달하였다.

3.5 세그멘테이션 정확도 비교

모델별 세그멘테이션 정확도 비교는 표 4와 같다. 컨볼루션 레이어 개수를 조절하지 않고 브랜치도 제거하지 않은 기존 모델인 PIDNet-small의 mIoU가 가장 높게 나타났고, 크기를 조절한 모델인 PIDNet-tiny, small 순으로 정확도가 떨어졌다. PIDNet-tiny와 small은 이미지의 특징 정보를 저장할 수 있는 컨볼루션 레이어의 개수가 기존보다 감소하였기 때문에 정확도가 하락한 것으로 추측된다. IOSNet의 세그멘테이션 정확도가 PIDNet의 모델들과 비교했을 때 가장 낮게 나타났다. IOSNet은 브랜치 구조를 사용해 여러 정보를 동시에 학습할 수 있었던 이전 모델과 달리 학습할 수 있는 정보가 제한적이기 때문에 낮은 정확도 수치를 보인다. 하지만, 타 실시간 세그멘테이션 모델인 SFNet과 PP-LiteSeg와 비교했을 때 높은 정확도를 보인다. 이는 IOSNet이 브랜치 제거로 인해 성능이 하락하였지만, 실시간 세그멘테이션 모델로 준수한 성능을 보이고 있음을 알 수 있다.

모델별 세그멘테이션 결과를 사진으로 표현한 결과는 그림 5와 같다. PIDNet-small은 결과는 GT와 비교했을 때 타깃과 논 타깃을 잘 구별하고, 세밀한 부분까지 세그멘테이션하는 것을 볼 수 있다. 하지만, PIDNet-tiny와 nano는 small보다 정밀도가 떨어지는

표 4. 실시간 세그멘테이션 모델 정확도 비교
Table 4. Comparison of accuracy for real-time segmentation models.

| Model | mIoU |
|----------------|---------------|
| SFNet-resnet18 | 0.8799 |
| PP-LiteSeg-T1 | 0.8759 |
| PIDNet-small | 0.9554 |
| PIDNet-tiny | 0.9454 |
| PIDNet-nano | 0.9359 |
| IOSNet (ours) | 0.9296 |

모습을 보인다. 또한, 세 번째 결과에서 볼 수 있듯이 오동작하는 경우가 있는데, 이는 다른 모델들에 비해 문맥 정보를 수용할 수 있는 컨볼루션 레이어의 개수가 떨어지기 때문이라 추측할 수 있다.

IOSNet은 다른 모델들보다 눈에 띄게 작은 부분에 대해 세그멘테이션하는 능력이 떨어지는 것을 알 수 있다. 이는 작은 물체를 학습할 수 있는 브랜치가 없음으로 인해 발생하는 현상이다. 하지만, 구강 스캐너는 초당 30장 이상의 이미지를 모형으로 재구성하기 때문에 이전 결과로 보정이 가능하다. 때문에, 이미지에 등장한 불확실한 물체를 대략적으로만 구분해내어도 치아 모형 구성이 가능하다. 따라서, 목표로 한 추론 시간 내에 수행이 가능한 IOSNet을 채택하는 것이 적절하다.

IV. 결 론

본 논문에서는 구강 스캐너에 탑재할 수 있는 의미론적 세그멘테이션 모델을 제안하였다. 구강 스캐너에 적용하기 위해서는 최우선으로 고려되어야 하는 것은 처리 속도였다. 실시간으로 치아 모형이 구성되는 모습을 사용자에게 시각화해주어야 하기 때문이다. 이를 위해, 우리는 기존의 실시간 세그멘테이션 모델인 PIDNet에서 추론 시간에 영향을 줄 수 있는 브랜치를 제거하여 목적을 달성하고자 하였다.

브랜치를 제거한 모델인 IOSNet은 목표한 대로 Windows와 C++ 환경에서 이미지 한 장을 10ms 이내로 처리하는 데 성공하였다. 비록 세그멘테이션 정확도는 기존 모델보다 하락하였지만, 이전 사진을 통한 보정이 가능하기에 IOSNet의 정확도는 스캐너에 사용하기에 무리가 없다. 또한, 작은 컴퓨팅 자원을 사용함으로써 구강 스캐너의 치아 모형 재구성 알고리즘의 동작에 지장을 주지 않는 세그멘테이션 모델을 제안하였다고 볼 수 있다.

추후 연구 주제로는 세밀한 부분을 세그멘테이션할

수 있도록 IOSNet에 기존의 브랜치를 추가할 수 있는 방법을 찾아보고자 한다. 구강 스캐너에서 사용할 수 있는 컴퓨팅 자원이 한정적이기 때문에, 구강 스캐너의 특화되지 않은 기존 방법을 그대로 적용하는 것은 한계가 있을 것으로 예상된다. 하지만, 병렬 처리를 사용하거나 브랜치에서 계산되는 과정을 최소화하여 소요 시간을 단축하는 방법이 있을 것으로 추측된다.

References

- [1] N. D. Kravitz, C. Groth, P. E. Jones, J. W. Graham, and W. R. Redmond, "Intraoral digital scanners," *J. Clin. Orthod.*, vol. 48, no. 6, pp. 337-347, 2014.
- [2] I. Christopoulou, E. G. Kakkamanos, M. A. Makrygiannakis, I. Bitsanis, P. Perlea, and A. I. Tsolakis, "Intraoral scanners in orthodontics: A critical review," *Int. J. Environ. Res. and Public Health*, vol. 19, no. 3, p. 1407, 2022. (<https://doi.org/10.3390/ijerph19031407>)
- [3] J. Xu, Z. Xiong, and S. P. Bhattacharyya, "Pidnet: A real-time semantic segmentation network inspired from pid controller," *arXiv preprint arXiv:2206.02066*, 2022. (<https://doi.org/10.48550/arXiv.2206.02066>)
- [4] E. Ha, K. J. Jeon, Y. H. Kim, J. Kim, and S. Han, "Automatic detection of mesiodens on panoramic radiographs using artificial intelligence," *Scientific reports*, vol. 11, no. 1, p. 23061, 2021. (<https://doi.org/10.1038/s41598-021-02571-x>)
- [5] Y. Ariji, M. Fukuda, Y. Kise, et al., "Contrast-enhanced computed tomography image assessment of cervical lymph node metastasis in patients with oral cancer by using a deep learning system of artificial intelligence," *Oral surgery, oral medicine, oral pathology and oral radiology*, vol. 127, no. 5, pp. 458-463, 2019. (<https://doi.org/10.1016/j.oooo.2018.10.002>)
- [6] J. Hwang, S. Azernikov, A. A. Efros, and S. X. Yu, "Learning beyond human expertise with generative models for dental restorations," *arXiv preprint arXiv:1804.00064*, 2018.

- (<https://doi.org/10.48550/arXiv.1804.00064>)
- [7] S. Tian, M. Wang, F. Yuan, et al., "Efficient computer-aided design of dental inlay restoration: A deep adversarial framework," *IEEE Trans. Medical Imaging*, vol. 40, no. 9, pp. 2415-2427, 2021.
(<https://doi.org/10.1109/TMI.2021.3077334>)
- [8] S. Tian, M. Wang, N. Dai, et al., "DCPR-GAN: Dental crown prosthesis restoration using two-stage generative adversarial networks," *IEEE J. Biomed. and Health Informatics*, vol. 26, no. 1, pp. 151-160, 2021.
(<https://doi.org/10.1109/JBHI.2021.3119394>)
- [9] S. Tian, R. Huang, Z. Li, et al., "A dual discriminator adversarial learning approach for dental occlusal surface reconstruction," *J. Healthcare Eng.*, 2022.
(<https://doi.org/10.1155/2022/1933617>)
- [10] K. Kim, "Evaluation of clinical adaptation of dental prostheses printed by 3-dimensional printing technology," *J. Technologic Dentistry*, vol. 41, no. 1, pp. 1-7, 2019.
(<https://doi.org/10.14347/kadt.2019.41.1.1>)
- [11] T. Lin, P. Dollár, R. Girshick, K. He, B. Hariharan, and S. Belongie, "Feature pyramid networks for object detection," in *Proc. IEEE Conf. CVPR*, pp. 2117-2125, 2017.
(<https://doi.org/10.48550/arXiv.1612.03144>)
- [12] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. CVPR*, pp. 3431-3440, 2015.
(<https://doi.org/10.48550/arXiv.1411.4038>)
- [13] O. Ronneberger, P. Fischer, and T. Brox, "U-net: Convolutional networks for biomedical image segmentation," *MICCAI 2015: 18th Int. Conf. Proceedings, Part III 18*, pp. 234-241, Munich, Germany, Oct. 2015.
(<https://doi.org/10.48550/arXiv.1505.04597>)
- [14] V. Badrinarayanan, A. Kendall, and R. Cipolla, "Segnet: A deep convolutional encoder-decoder architecture for image segmentation," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 39, no. 12, pp. 2481-2495, 2017.
(<https://doi.org/10.48550/arXiv.1511.00561>)
- [15] H. Zhao, J. Shi, X. Qi, X. Wang, and J. Jia, "Pyramid scene parsing network," in *Proc. IEEE Conf. CVPR*, pp. 2881-2890, 2017.
(<https://doi.org/10.48550/arXiv.1612.01105>)
- [16] L. Chen, G. Papandreou, I. Kokkinos, K. Murphy, and A. L. Yuille, "DeepLab: Semantic image segmentation with deep convolutional nets, atrous convolution, and fully connected crfs," *IEEE Trans. Pattern Anal. and Mach. Intell.*, vol. 40, no. 4, pp. 834-848, 2017.
(<https://doi.org/10.48550/arXiv.1606.00915>)
- [17] L. Chen, Y. Zhu, G. Papandreou, F. Schroff, and H. Adam, "Encoder-decoder with atrous separable convolution for semantic image segmentation," in *Proc. ECCV*, pp. 801-818, 2018.
(<https://doi.org/10.48550/arXiv.1802.02611>)
- [18] F. Chollet, "Xception: Deep learning with depthwise separable convolutions," in *Proc. IEEE Conf. CVPR*, pp. 1251-1258, 2017.
(<https://doi.org/10.48550/arXiv.1610.02357>)
- [19] M. Gamal, M. Siam, and M. Abdel-Razek, "ShuffleSeg: Real-time semantic segmentation network," *arXiv preprint arXiv:1803.03816*, 2018.
(<https://doi.org/10.48550/arXiv.1803.03816>)
- [20] H. Li, P. Xiong, H. Fan, and J. Sun, "Dfanet: Deep feature aggregation for real-time semantic segmentation," in *Proc. IEEE/CVF Conf. CVPR*, pp. 9522-9531, 2019.
(<https://doi.org/10.48550/arXiv.1904.02216>)
- [21] M. Orsic, I. Kreso, P. Bevandic, and S. Segvic, "In defense of pre-trained imagenet architectures for real-time semantic segmentation of road-driving images," in *Proc. IEEE/CVF Conf. CVPR*, pp. 12607-12616, 2019.
(<https://doi.org/10.48550/arXiv.1903.08469>)
- [22] C. Yu, J. Wang, C. Peng, C. Gao, G. Yu, and N. Sang, "Bisenet: Bilateral segmentation network for real-time semantic segmentation," in *Proc. ECCV*, pp. 325-341, 2018.
(<https://doi.org/10.48550/arXiv.1808.00897>)
- [23] R. P. Poudel, U. Bonde, S. Liwicki, and C.

Zach, "Contextnet: Exploring context and detail for semantic segmentation in real-time," *arXiv preprint arXiv:1805.04554*, 2018.

(<https://doi.org/10.48550/arXiv.1805.04554>)

- [24] R. P. Poudel, S. Liwicki, and R. Cipolla, "Fast-scnn: Fast semantic segmentation network," *arXiv preprint arXiv:1902.04502*, 2019.

(<https://doi.org/10.48550/arXiv.1902.04502>)

- [25] C. Yu, C. Gao, J. Wang, G. Yu, C. Shen, and N. Sang, "Bisenet v2: Bilateral network with guided aggregation for real-time semantic segmentation," *Int. J. Computer Vision*, vol. 129, pp. 3051-3068, 2021.

(<https://doi.org/10.48550/arXiv.2004.02147>)

- [26] G. J. Brostow, J. Fauqueur, and R. Cipolla, "Semantic object classes in video: A high-definition ground truth database," *Pattern Recognition Lett.*, vol. 30, no. 2, pp. 88-97, 2009.

(<https://doi.org/10.1016/j.patrec.2008.04.005>)

- [27] J. Deng, W. Dong, R. Socher, L.-J. Li, K. Li, and L. Fei-Fei, "ImageNet: A large-scale hierarchical image database," *2009 IEEE Conf. CVPR*, pp. 248-255, 2009.

(<https://doi.org/10.1109/CVPR.2009.5206848>)

- [28] H. Ibrahim, A. Salem, B. Yagoub, H. S. Kang, and J.-W. Suh, "Semantic object segmentation using conditional generative adversarial network with residual connections," *J. Korea Inst. Inf. and Commun. Eng.*, vol. 26, no. 12, pp. 1919-1925, 2022.

(<https://doi.org/10.6109/jkiice.2022.26.12.1919>)

유 호 진 (Hojin Yoo)



2021년 2월: 인천대학교 임베디드시스템공학과 학사 졸업
2023년 2월: 인천대학교 임베디드시스템공학과 석사 졸업
<관심분야> 딥러닝, 컴퓨터 비전

조 항 재 (Hangjae Cho)



2018년 3월~현재: 인천대학교 임베디드시스템공학과 재학
<관심분야> 인공지능, 컴퓨터 비전

김 재 곤 (Jaegon Kim)



2001년 2월: 한국과학기술원 공학박사 졸업
2003년 8월~현재: 인천대학교 산업경영공학과 교수
<관심분야> 산업인공지능, 알고리즘

전 경 구 (Kyungkoo Jun)



2001년 5월: Purdue Univ. 공학박사 졸업
2004년 3월~현재: 인천대학교 임베디드시스템공학과 교수
<관심분야> 인공지능, 시스템 소프트웨어, 알고리즘