

다중 트레이닝 기법을 이용한 MASK R-CNN의 초음파 DDH 각도 측정 진단 시스템 연구

황석민¹, 이시욱², 이종하^{1*}

¹계명대학교 의공학과, ²계명대학교 동산병원 정형외과

A Study on a Mask R-CNN-Based Diagnostic System Measuring DDH Angles on Ultrasound Scans

Seok-Min Hwang¹, Si-Wook Lee², Jong-Ha Lee^{1*}

¹Dept. of Biomedical Engineering, Keimyung University, Daegu, South Korea

²Orthopedic at Keimyung University, Dongsan Medical Center, Daegu, South Korea

요 약 최근 영유아 성장기에 발생하는 고관절 이형성증(Developmental Dysplasia of Hip, DDH)의 숫자가 늘어나고 있다. DDH는 영유아 성장을 방해하고 다른 부작용도 많이 발생시키기 때문에 최대한 조기에 발견하여 치료해야 한다. 최근 들어 Convolutional Neural Networks (CNN) 및 개선된 Resnet50을 활용한 머신러닝 기법이 초음파 영상 분석에 많이 활용되고 있다. 연구 결과를 보면 컴퓨터 보조 이미지 분석이 의료현장에서 객관성과 생산성을 크게 향상시키고 있다. 본 연구의 결과는 정형외과에서의 난제인 초음파 영상을 통한 DDH 컴퓨터 보조 진단 알고리즘에도 충분히 활용될 수 있다는 것을 보여주고 있다. 본 논문에서는 CNN을 활용하여 DDH를 자동으로 측정하고 진단할 수 있는 컴퓨터 보조 진단 알고리즘을 제안하였다. DDH 측정을 위해 유아 고관절의 정상/비정상 판독을 위해 Acetabulum-Femoral head의 angle을 자동으로 계산하였으며 기존 영상을 딥 러닝하여 진단을 자동으로 하는 알고리즘을 설계하였다. 실험 결과 의사와 비교하여 진단의 속도와 정확도가 향상된다는 것을 확인하였다.

• 주제어 : DDH, 고관절 이형성증, 의료 영상, EOS, 인공지능, CNN

Abstract Recently, the number of hip dysplasia (DDH) that occurs during infant and child growth has been increasing. DDH should be detected and treated as early as possible because it hinders infant growth and causes many other side effects. In this study, two modelling techniques were used for multiple training techniques. Based on the results after the first transformation, the training was designed to be possible even with a small amount of data. The vertical flip, rotation, width and height shift functions were used to improve the efficiency of the model. Adam optimization was applied for parameter learning with the learning parameter initially set at 2.0×10^{-4} . Training was stopped when the validation loss was at the minimum. A novel image overlay system using 3D laser scanner and a non-rigid registration method is implemented and its accuracy is evaluated. By using the proposed system, we successfully related the preoperative images with an open organ in the operating room.

• Key Words : DDH, Developmental dysplasia of the hip, Medical image, EOS, AI, CNN

Received 21 December 2020, Revised 28 December 2020, Accepted 30 December 2020

* Corresponding Author Jong-Ha Lee, Department of Biomedical Engineering, School of Medicine, Keimyung University, 1095 Dalgubeoldae-ro, Daegu, Korea, Korea. E-mail: segeberg@gmail.com

I. 서론

1.1 연구배경

최근 영유아 성장기에 발생하는 고관절 이형성증(Developmental Dysplasia of Hip, DDH)의 숫자가 늘어나고 있다. 어린아이의 경우에는 외형적으로 증상이 나타나지 않는다. 평균적으로 유아 중 1~2% 정도 발생하고 수동적으로 관찰했을 때 15%까지 영상을 통해서 DDH를 관찰할 수 있다. 생후 1년 동안 임상적으로 조기에 발견된다면 쉽게 치료할 수 있다. 하지만 치료하지 않고 방치한다면 심각한 결과를 초래할 수 있다. 미국에서도 신생아에 대한 정기 검진이 시행되어 왔으나, 검진 방법이 획일화 되어 있지 않고 명확하지 않다. 한국은 2006년부터 영유아 검진 사업에 발달성 고관절 이형성증을 포함시켰다. 해당 증상에 대한 연구가 시작되었으나 미국에서도 미약한 만큼 한국에서는 해당 연구에 대한 기준이 명확하지 않다. DDH의 표준 검진 방법은 Ultrasonography(US)를 활용한 의사가 시각적으로 판독하는 것이다. 하지만 Ultrasonography를 이용한 표준 검진방법에 영향을 미치는 요소로는 사용자와 시스템적인 요소가 있다. 먼저 사용자의 요소를 살펴보면 의료진이 Ultrasonography를 확인하여 Acetabulum-Femoral head angle 수치 등을 추출해낸다. 그리고 의사에게 해당 수치를 토대로 진단을 실시하는데 이때 추출된 수치 값의 대한 오차에 의해서 의사가 판독하는 진단 값이 달라질 수 있다. 또한 시스템적인 요소로는 Ultrasonography 화질에 대한 여부도 큰 영향을 미친다. 표준 판독방법은 시각적 판독 방법 이기에 Ultrasonography의 해상도 및 정확도에 따라서 결과 값에 대해서도 큰 영향을 끼친다. 이 요인은 정해진 가이드라인 없이 수동으로 직접 사람이 작업하기에 오차가 많이 발생한다. 또한 Ultrasonography를 가지고 의사가 증상을 판별하기 위해서는 영상만 가지고 판단하는 게 아니고 길이나 각도 등을 부위마다 직접 재야하는데 현재 방법으로는 의료진이 수동적으로 판별하여 의사에게 전달하는 가공작업을 거치고 있다. 현 방법으로는 가공하는 데 걸리는 시간이 한 장당 10분 이상 소요되고 있는 것이 현실이다. 원활한 의료서비스를 제공하는 데에 있어서 어려움을 겪고 있다.

현재 기술의 발달로 컴퓨터의 연산속도가 매우 빠르게 증가하고 있으며 이미지나 영상을 처리하는 그래픽

픽 처리 장치(Graphics Processing Unit, GPU)의 성능이 발달하면서 Convolutional Neural Networks(CNN)와 같은 영상처리 인공지능이 산업, 과학, 심지어 의료 쪽으로도 많이 쓰이고 있다. 의료에 시각적인 판단방법은 매우 중요한 요소이며 CNN 및 개선된 Resnet50을 활용한 머신러닝 기법이 초음파 영상 분석에 많이 활용되고 있다. 그리고 관련 논문을 참고하면 영상처리 인공지능을 활용한 이미지 분석이 의료 현장에서 객관성과 생산성을 크게 향상시키고 있다. 이를 이용하면 정형외과에서의 난제인 초음파 영상을 통한 DDH 컴퓨터 보조 진단 알고리즘에도 충분히 활용될 수 있다는 것을 보여주고 있다.

현재의 인공지능 기법 중 하나인 CNN의 경우에는 영상을 가지고 인공지능이 판독하는 방식을 가지고 있는데 사람의 개입 정도에 따라서 기계 학습(Machine Learning)과 딥 러닝(Deep Learning)으로 나뉜다. 사람의 개입의 정도가 낮을수록 딥 러닝에 가까워지는데 사용자의 개입이 적어지면 그만큼 데이터 자료 양도 매우 커진다. 하지만 데이터의 양이 적은 곳에서는 기계 학습을 하기 위해 필요한 기본 데이터의 양도 부족한 것이 현실이다. 딥 러닝에 가까워질수록 데이터의 양이 커진다면 데이터 양의 부족 현상을 해결하기 위해서는 반대로 사용자의 개입을 늘리면 된다. 그래서 본 연구에서는 사용자 필터를 이용한 다중 트레이닝 기법을 제안한다.

본 연구에서는 CNN을 활용하여 적은 데이터양으로도 고관절 이형성증(Developmental Dysplasia of Hip, DDH)을 자동으로 진단하고 측정할 수 있는 다중 트레이닝 기법을 이용한 MASK R-CNN의 초음파 DDH 각도 측정 진단 시스템 연구를 제안한다. DDH 측정을 위해 유아 고관절의 정상/비정상 판독을 위해 Acetabulum-Femoral head의 angle을 자동으로 계산하였으며 기존 영상을 딥 러닝하여 진단을 자동으로 하는 알고리즘을 설계하였다.

1.2 배경 이론

1.2.1 고관절 이형성증

고관절은 골 반쪽 관절면의 비구와 하지의 대퇴골두가 연결되어 이루어지는 관절로 골반의 압력을 지지하고 서거나 달리거나 전체적인 하지의 움직임이 이루어지도록 만드는 주요 관절이다. 하지만 고관절 쪽 이

상 변형이 생기면 비구와 대퇴골두 간의 접촉 면적이 줄어들게 되고 결과적으로는 관절의 압력이 증가하게 됨으로써 관절 연골의 퇴행성 변화를 촉진시키는 원인이 된다. 이런 변형은 원활한 관절 운동을 방해하거나 관절면의 불일치를 유발하여 일상생활에 통증이나 불편감을 느끼게 하고 주변 조직에 과도한 충격을 주어서 관절 손상을 일으켜 조기에 퇴행성관절염을 유발시킨다. 고관절 이형성증은 어렸을 때 조기에 발견하면 효과적으로 치료가 가능하다. 보행기 이전에 치료하면 쉽게 치료가 가능하며 보행기 이전에는 보장구, 이중삼중 기저귀, 본 로젠(Von Rosen) 부목, 파브릭(Pavlik) 보장구 등을 이용하여 치료가 가능하다. 보장구를 이용한 치료는 6개월 이전의 신생아에게 사용하기 좋다. 보행기 이후에는 비수술적 방법으로는 뼈를 위치에 바로 잡아가지고 석고 붕대로 고정하는 방법이 있는데 이는 적용이 되는 경우가 낮고 대부분 시행되는 수술적 방법으로 피부를 절개하여 골절 부위를 맞추는 관절적 정복과 골 고정술을 통하여 치료한다. 고관절 이형성증을 보이는 신생아의 경우에 치료를 하지 않는다면 정상적으로 자연히 회복되기는 매우 힘들며 신생아 때 치료를 하지 않는다면 다리가 짧아지고 근력이 약해져 다리를 절개 되거나 가성비구가 형성되어 조기 퇴행성 변화가 발생하게 된다. 그리고 이차 증상이 발생하기도 하는데 가장 대표적으로 측만증과 요통이 발생하게 된다. 비구의 이형성이나 아탈구가 있는 경우에는 약간의 하지 단축을 보이고 다리를 약간 저는 정도의 증상만이 있거나 증상이 거의 없을 수도 있다. 하지만 추가적으로 비구 연골이나 비구순 등의 파열 및 퇴행성 변형이 일어나고 절반 정도는 고관절의 퇴행성관절염이 발생하게 됨으로 적극적인 치료가 필요하다.

1.2.2 인공지능 기반 의료영상 해석

인공지능은 학습, 문제 해결, 패턴 인식 등과 같이 주로 인간기능과 연결된 인지문제를 해결하는 데 주력하는 컴퓨터 공학 분야 중 하나이고 보통은 AI로 줄여서 표기한다. 인공지능은 현재 컴퓨터 공학의 최종 단계로 본다. 최근에 기술의 발달로 컴퓨터 연산 속도와 효율성이 증가함으로써 기계 학습 분야에서 몇 가지 분야로 발전됨으로써 이와 동시에 네트워크 컴퓨팅 또한 발전함으로써 딥 러닝이라는 하위 분야에 대한 발전 또한 증가했다. 인공지능은 사용자의 개입에 정도

에 따라 머신러닝, 딥 러닝, 강화 학습 등으로 분류한다.

머신 러닝은 패턴인식 등 학습에 사용되는 기법에 사용되는 이름이다. 기계학습은 정답 또는 기록된 데이터를 통해 학습하고 예측함으로써, 불확실한 알고리즘에서 정답을 추론해내는 과정에 대해 최적화 하고, 예측 되지 못한 변수에 대해서 숨겨진 특징 점을 추출해냄으로써, 데이터에 대해 간결하게 표현할 수 있는 알고리즘의 모음이라고 할 수 있다. 기존처럼 소프트웨어 개발자가 설계한 알고리즘에 의해 주어진 입력에 따라 프로그램 코드별로 출력하는 기존 기법과는 다르게 기계학습은 데이터를 사용해서 필터를 통한 비교 분석을 통해 통계 코드를 생성하여 사전에 생성된 모델을 바탕으로 정답을 찾아내서 포출하다. 모델의 정확성을 위해서는 데이터의 양과 질을 높이는 게 중요하다.

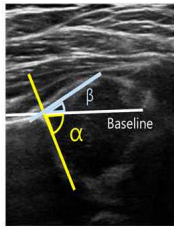
딥 러닝은 머신러닝에서 사용자의 학습 부분을 줄이거나 생략하는 기계학습의 한 분야이다. 기존의 머신러닝에서 좀 더 진화된 과정으로써 기존 머신러닝은 사용자가 모델링을 형성하는 학습과정에서 정답에 특징점을 포출하여 다른 예제에 대해서 정답을 도출하는 과정이었다면 딥 러닝은 사용자가 개입하지 않고 수많은 데이터를 통해서 분석하여 자체적으로 정답을 추론해 나가는 과정이다. 기존의 머신 러닝처럼 설명 가능한 관계 집합을 생성하는 게 아닌 비선형 알고리즘 계층을 사용하여 다양한 입력 값을 바탕으로 각 입력 값의 관련성을 찾아 요소를 바탕으로 상호 작용하는 모델링을 생성한다. 딥 러닝을 이용하면 일반 소프트웨어의 기법을 이용한 예측이나 기존 머신 러닝 방법보다 더 예측할 수 없는 복잡한 관계, 사람이 인지 못하는 부분까지 파악 할 수 있다. 딥 러닝을 하기 위해서는 머신 러닝보다 많은 양의 데이터를 필요로 하기에 대량의 데이터를 보유하고 있는 곳에서 사용하기 적합하다.

1.3 선행연구

1.3.1 표준화된 고관절 이형성증 측정방법

고관절 이형성증(DDH)를 측정하기 위해서는 현재 육안으로 검사하는 방법과 방사선을 통한 확인 방법 두 가지가 있다. 먼저 육안으로 검사하는 방법은 여러 가지가 있는데, 대퇴부 피부 주름(Unequal folds of

skin)의 비대칭 여부이다. 왼쪽과 오른쪽의 허벅지 피부의 주름이 비대칭인 경우이다. 다음 갈라아치(Galeazzi) 검사이다. 굴곡이 없는 매끄러운 바닥에 유아를 눕히고 무릎을 세웠을 때 양쪽 높이가 다른 경우이다. 다음 오토라니(Otolani) 검사이다. 탈구된 골두를 골반에 맞추었을 때 소리를 확인 하는 검사법인데 대상자를 바로 눕히고 고관절과 무릎관절을 구부린 다음 허벅지를 벌리면서 대퇴 골두 부위를 밀어 올리면 소리가 나면서 비구내로 들어가는 느낌이 감지된다. 다음 발로(Barlow)검사이다. 오토라니 검사와 다르게 탈구를 유도하는데 고관절을 벌린 상태로 허벅지를 모으면서 대퇴골두 부위를 밀어내어 탈구를 유도한다. 여기까지가 육안으로 검사하는 방법이다. 방사선을 이용한 방법에는 2가지가 있다. 일반 x-ray촬영과 초음파 촬영인데 x-ray로 확인 하려면 대퇴골두가 골반 골에 접촉된 상태가 먼저 확인되어야 하는데 신생아 시기에는 연골로 이루어져있어 x-ray로는 관측이 불가능하다. 6-7개월 이후에는 x-ray로 관측 되지만 조기 진단을 위해서는 초음파 검사가 적합하다. 다음 본 연구에 핵심인 초음파 검사인데 관측하기 위해서는 고관절의 관상 면을 촬영하는데 장골 뼈가 수평이 되어있도록 하고 삼중연골이 잡히는 부위의 초음파 사진을 촬영해야 하는데 이때 장골은 사진 수평으로부터 $\pm 5^\circ$ 안쪽에 있어야 한다. 그리고 알파각과 베타 각에 따라 이상소견을 분류한다.



Classification	Alpha angle	Beta angle	Description
1	>60	<55	Normal
1a	50-60	55-77	Immature(<3 mo)
2b	$>50-60$	55-77	≥ 3 mo
2c	43-49	>77	Acetabular deficiency
2d	43-49	>77	Everted labrum
3	<43	>77	Everted labrum
4	Unmeasurable	-	Dislocated

Fig. 1. Angles examined in US, Graf classification of DDH

1.3.2 윤곽선 검출 필터

초음파 사진을 보면 불필요한 부분이 매우 많다. 초음파 사진은 일반 CT사진과는 다르게 형태를 나타내는 특징점이 적고 형태를 나타내는 지점이 뚜렷하지 않기에 유사한 부분이 매우 많다. 유사한 부분이 바로 검출을 저해하는 요소이다. 특징점을 추출하기 위해서는 불필요한 부분을 제거해야 한다. 본 연구에서 필터

방식으로 윤곽선 검출(Edge Detection) 방식을 시도하였다. 초음파 사진자체가 흑백으로 이루어진 그레이스케일 사진이라 별도 처리 없이 진행이 가능한데 사진에서 윤곽선을 검출하는 필터 방식이다. 윤곽선이라는 것은 서로 다른 명암을 가진 경계선, 즉 픽셀의 명암이 임계값보다 급격히 변하는 지점을 윤곽선이라고 하는데 이러한 변화량은 미분 값으로 찾을 수 있다. 1차 미분에서는 기울기의 값으로 2차 미분에서는 부호로 알 수 있다. 영상에서 윤곽선 검출은 마스크로 하는데 미분연산자와 동일한 효과를 가지고 있는 필터이다. 1차 미분 방식에는 소벨, 프리윗, 로버츠 마스크가 있다. 소벨 마스크와 프리윗 마스크는 잡음에 강하고 연산 속도가 느리다. 로버츠 마스크는 연산 속도가 빠르고 엷지를 확실하게 검출할 수 있지만 잡음에 약하다. 1차 미분 방식은 가로 또는 세로에 대한 엷지를 검출 가능하다. 2차 미분 방식에는 라플라시안과 캐니 마스크가 있는데 라플라시안 마스크는 먼저 가우시안 스무딩을 통해서 잡음을 줄여준 다음에 LOG(Laplacion of Gaussian)을 사용한다. 다음은 캐니 마스크인데 먼저 가우시안 필터로 평활화를 통해 잡음 제거 후에 마스크 연산으로 엷지를 검출한 다음 Non-Maximum value 제거를 하여 잘못 검출된 엷지를 제거한다.

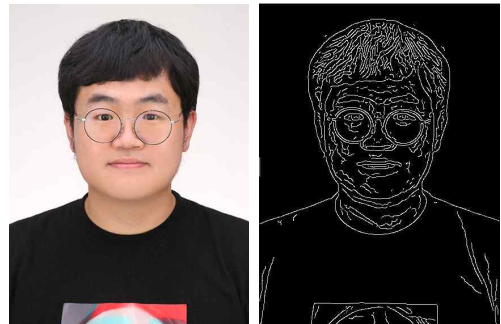


Fig. 2. Original image (Left), An image with the edge filter applied (Right)

1.4 연구 목적

선행연구에 기재한 바와 같이 유아의 고관절 이행성증(DDH)를 진단하기 위해서는 조기에 진단하여야 하고 진단할 수 있는 방법으로 가장 확실한 것은 초음파를 이용하는 방법이다. 하지만 대상자의 연령이 낮은 점과 초음파 사진이라는 특성 때문에 진단에 있어서

시간과 노력에 비해서 적은 양의 데이터를 추출한다. 추출한 데이터 중에서도 사용 가능한 데이터는 얼마 되지 않는다. 초음파 사진이 육안으로 식별이 가능할 만큼 깨끗하게 나와야 하고 고관절의 장골과 베이스라인이 5도 이하로 나와야 사용 가능한 사진이 되기 때문이다. 현재 방법으로는 의뢰인이 직접 사용 가능한 사진을 판독해야 한다. 방법으로 직접 사진을 가지고 컴퓨터로 길이와 각도를 재서 수치를 표시한 다음 유무를 판단한다. 이 과정에 있어서 매우 시간이 오래 소모된다. 최근 AI에 대한 연구가 활발하게 진행되고 있는 만큼 이에 따라 기존 프로그램으로 판독하기 어려운 초음파 사진도 AI를 통해서 판독이 가능할 것이다. 초음파 사진은 비슷한 부분이 많아서 사람의 육안으로는 특징점을 검출하기 매우 힘들다. 하지만 CNN을 통하여 특징점을 추출한다면 초음파 영상으로도 특징점을 추출 가능하다. 머신 러닝이나 딥 러닝을 사용하기 위해서는 데이터의 양이 매우 많아야 하는데 적은 데이터를 보유하고 있다면 AI기술을 사용하기 매우 어렵다. 기술을 사용하는데 있어서 대부분 데이터가 적거나 있지만 가공되지 않은 데이터를 보유하고 경우가 대부분이다. 많은 가공 데이터를 확보하기 위해 적은 양으로 머신 러닝이나 딥 러닝을 돌리기 위한 초기 모델링을 만들기 위한 방법의 필요성이 요구된다. 기존 머신 러닝보다 사람의 개입이 많아지고 데이터의 양은 줄일 수 있는 방법에 대해 고안하였고 본 다중 트레이닝을 이용한 Mask-RCNN 초음파 DDH 검출 방법을 제안하였다. 해당 방법을 통해서 연구실이나 기업 병원 등에서 초기에 적은 양의 데이터로 많은 량의 데이터를 가공할 수 있는 시드 모델링을 생산하여 초기에 AI 환경을 구축하는 데 큰 도움을 줄 수 있다.

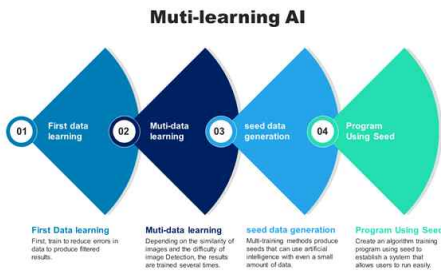


Fig. 3. The goal of the system

II. 연구 내용

2.1 연구 방법

본 연구는 적은 양의 데이터를 가지고 다중 트레이닝 방법을 이용하여 AI를 이용하여 이후 들어오는 가공되지 않은 데이터를 쉽게 처리하여 사용자의 접근성을 향상시키는 시스템의 개발을 목적으로 한다.

먼저 기존 선행되었던 DDH의 진단 방법의 경우에는 장골(Ilium)과 구개(Acetabular roof)를 먼저 영역 검출한 다음 구개의 각도와 장골의 각도를 얻어와서 이를 다시 알고리즘으로 계산하여 값을 도출해 내었다. 하지만 실제 의사의 판독에서는 장골과 구개를 이용한 영역 검출법으로 각도를 계산하는 것이 아닌 장골과 구개의 해부학적인 이해를 기반으로 α 각도와 β 각도에 해당하는 부분을 색출해 내어 각도를 계산해낸다.

따라서 의사의 검출 방법을 Mask R-CNN 방법을 사용하여서 장골과 구개가 이루는 모양을 먼저 Segmentation한다. 사진의 유효성을 판별하는 장골의 2개의 포인트와 α 각도와 β 각도를 계산 할 수 있는 3개의 포인트 총 4개의 포인트(장골 포인트와 각도 포인트 1개 중복)를 이용한다. 이를 연결하여 의사의 판단에 필요한 DDH의 핵심적인 값을 도출해 내고자 한다.

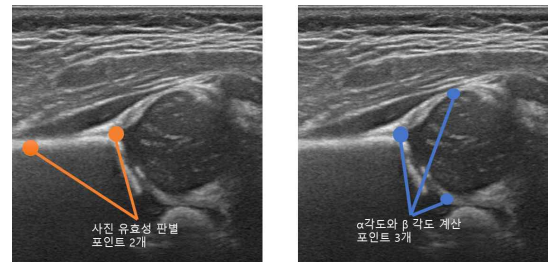


Fig. 4. Four points on ultrasound scans

기존 AI의 사람을 찾아내기 위한 연구에서도 segmentation과 keypoint를 동시에 활용하면 AI가 모델링을 형성하는데 같은 양의 데이터에 대해 많은 정보가 주어지기에 더 높은 성능을 도출한 사례를 바탕으로 장골과 구개의 4개의 포인트에 segmentation을 활용하여 딥 러닝 모델링을 생성하고자 한다.

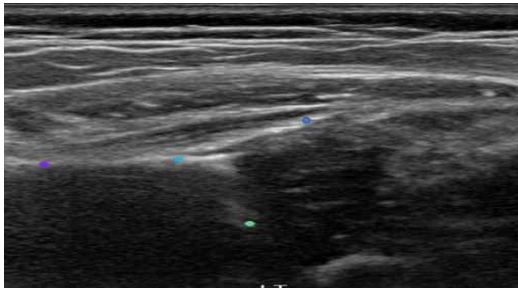


Fig. 8. Point marks for the second training

2.2 최종 결과

2차 결과까지 트레이닝이 끝나면 1차 모델링과 2차 모델링을 이용하여 다중으로 결과를 ai를 구동한다. 다중으로 구동하게 되면 장골과 구개가 이루는 길이와 각을 얻어낼 수가 있다. 해당 값을 토대로 의사가 판단을 보조할 수 있는 ai 시스템이 완성된다.

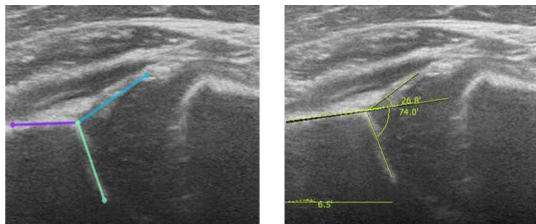


Fig. 9. Final results of the multi-AI system

III. 연구 내용

3.1 연구 결과

아래와 같은 데이터를 활용하였다.

총 데이터: 2,692장

유효데이터 : 1,241 / 2,692장

모델링 생성을 위한 데이터 : 321 / 1,241장

모델링 트레이닝 데이터 : 288 / 321장

모델링 테스트 데이터 : 33 / 321장

모델링을 통한 테스트 데이터 920 / 1,241장

2016년부터 2019년까지 확보한 총 2,692장의 데이터를 동산의료원으로부터 확보하였다. 이 중 실험에 활용한 유효한 데이터는 1,242장이었다. 이 중 321장을 랜덤하게 활용하여 모델을 생성하였다. 사용자 임의로

모델링을 위한 트레이닝 데이터와 학습데이터로 구분하였는데, 본 논문에서는 288장은 트레이닝 데이터, 33장은 테스트 데이터로 활용하였다.

학습 모델 생성을 위해 본 논문에서 제안한 다중트레이닝 알고리즘을 사용하였으며 학습 변수는 $2.0e-3$ 로 시작하여 validation loss가 가장 낮은 지점에서 중지하였다. 생성된 모델을 바탕으로 920장으로 테스트를 진행하였다.

동산의료원으로부터 확보한 322개의 모델링 데이터 중 288개의 트레이닝 데이터와 33개의 검증 데이터를 임의로 분리하여 학습시켜 모델링을 생성하였다. 시스템의 정확도는 920개의 데이터 중 유효데이터(동산의료원 정형의사 판독결과 판독 가능한 US 데이터) 389개 중 AI는 369개를 검출하여 94.86%의 검출 성능을 확인하였다.

3.1.1 유효 데이터

환자의 움직임, 기기의 유동성 등으로 인해 의료진이 촬영하여 초음파 사진의 데이터를 수집할 때 가공되지 않은 데이터 중 유효한 데이터는 20% 이하다. 대상이 영유아라면 더욱더 유효 데이터를 수집하기 어려워진다. 본 연구에서는 2,692개의 가공되지 않은 데이터 중 1,242개의 데이터를 실험 논문에 사용되는 데이터로 분류(1차 분류)하였다. 그 후 AI를 통한 테스트 데이터 920개에서 AI의 성능을 판별하기 위해서 테스트 데이터의 정답을 가려내기 위해 판독이 가능한 유효한 389개의 데이터를 한 번 더 분류(2차 분류)하였다.

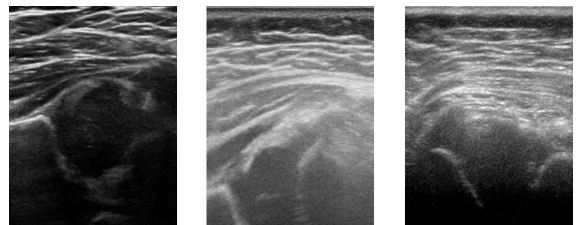


Fig. 10. Comparison of ultrasound scans

3.1.2 US 초음파 사진 검출 방법

유아 고관절 이형성증 초음파 사진을 검출하기 위해 2세 미만의 영유아를 대상으로 시행하였으며, 초음파 기기는 필립스사의 HD7- XE를 사용하였고 유아별 30장 이상 촬영을 실시하였다.

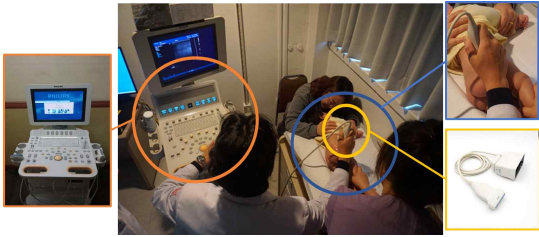
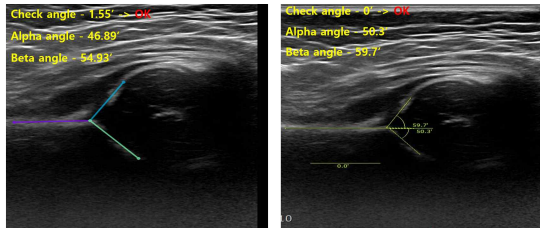


Fig. 11. Schematic diagram for US

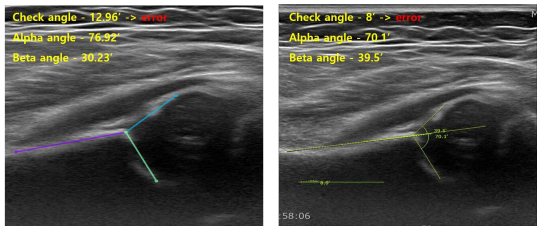
3.1.3 시스템 성능 측정을 위한 데이터 분류

성능 측정을 위하여 Appropriate detection, Fail detection, Different result, False image, Inappropriate detection, Insufficient image 총 6가지로 분류하였다. 여기에서 OK는 정상인 데이터에 정상적으로 값을 도출했을 경우를 의미하며 Error는 값을 도출했으나 base angle 값이 정상 범주가 아니어서 비정상 이미지를 의미하며 NA는 non-available로 값을 도출해 낼 수 없는 경우를 의미한다.

먼저 Appropriate detection는 AI와 doctor의 결과 값이 같은 경우로 둘 다 OK 또는 Error를 체크했을 때이다. [ex : AI(Check angle "Error"), Doctor(Check angle "Error")]



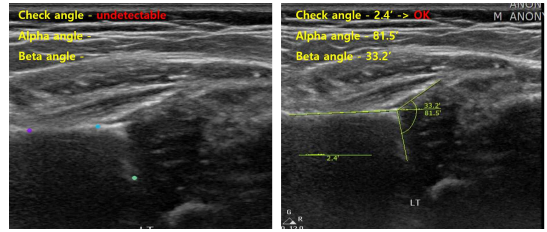
(a) AI(Check angle "CK") (b) Doctor(Check angle "CK")



(c) AI(Check angle "Error") (d) Doctor(Check angle "Error")

Fig. 12. Appropriate detection

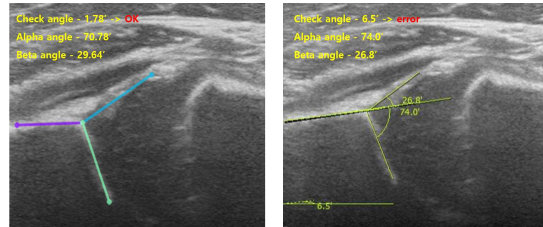
Fail detection은 doctor가 정상적인 사진으로 판단했는데도 AI가 검출을 하지 못한 경우이다. [ex : AI(Check angle "NA"), Doctor(Check angle "OK")]



(a) AI(Check angle "NA") (b) Doctor(Check angle "CK")

Fig. 13. Fail detection

Different result는 AI와 doctor의 진단 값이 서로 상이한 경우로 각각 OK, Error 값을 상이하게 표현한 경우이다. [ex : AI(Check angle "OK"), Doctor(Check angle "Error")]

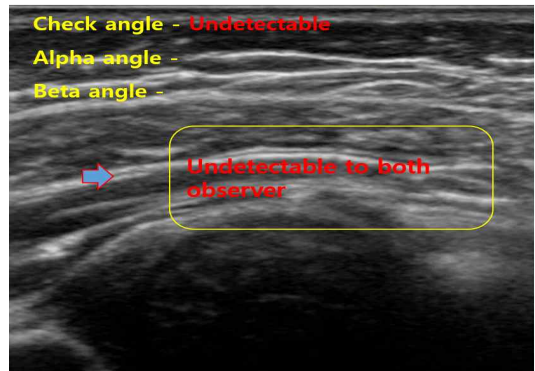


(a) AI(Check angle "CK") (b) Doctor(Check angle "Error")

Fig. 14. Different result

False image는 AI도 doctor도 측정 불가능한 image이다. [ex : AI(Check angle "NA"), Doctor(Check angle "NA")]

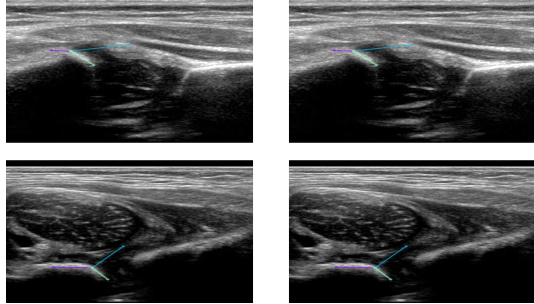
AI and Doctor (Check angle 'error')



(a) AI and Doctor(Check angle "NA")

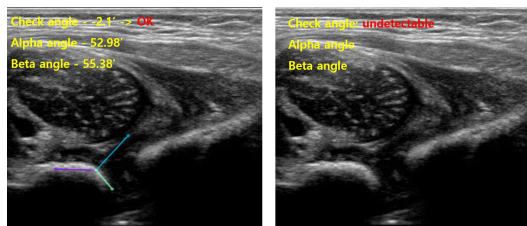
Fig. 15. False image

Inappropriate detection은 doctor가 측정 불가능한 이 이미지에도 불구하고 AI가 측정한 경우이다. [ex : AI(Check angle “NA”), Doctor(Check angle “NA”)]



(a) AI(Check angel “CK”) (b) Doctor(Check angle “NA”)
Fig. 16. Inappropriate detection

Insufficient image는 Ai가 측정했지만 해상도 등으로 시각적으로 불충분한 이유로 Doctor가 측정하지 못한 경우이다.



(a) AI(Check angel “CK or Error”) (b) Doctor(Check angle “NA”)
Fig. 17. Insufficient image

총 920개 데이터 중 Appropriate detection은 413개, Fail detection은 98개, Different result은 30개, False image은 336개, Inappropriate detection은 9개, Insufficient image는 34개로 분류되었다. 아래 그림을 보면 왼쪽부터 각각 'Appropriate Detection', 'False Image', 'Different Result', 'Insufficient Image', 'Fail Detection', 'Inappropriate Detection' 이다.

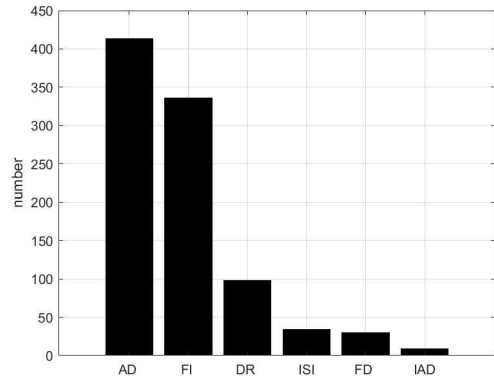


Fig. 18. Detection results

3.1.4 다중 트레이닝 AI 모식도

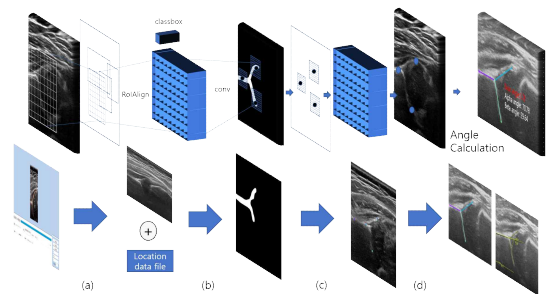


Fig. 19. Schematic diagram for multi-training model

전체적인 다중 트레이닝에 대한 그림이다. (a)에서 CVAT 프로그램을 이용하여 트레이닝 위치를 지정해준다. 1차 트레이닝 결과에 해당하는 이미지를 변환한다(b). 변환된 이미지에 2차 트레이닝을 실시한다(c). 이후 최종 결과를 도출해낸다(d).

3.1.5 시스템 성능 측정

시스템 측정을 위하여 모델 평가 분류 지표에 따라 구별해야 한다. 정답을 True와 false로 나누어야 함에 따라 이때 분류 모델도 OK(True)와 Error로 구분한다. 그렇게 하면 2x2 매트릭스를 만들 수가 있는데 NA를 제외한 데이터 총 510개 중 True Positive(TP)는 320개, False Positive(FP)는 48개, False Negative(FN)는 49개, True Negative(TN)는 93개로 검출된다.

Table 1. 2x2 matrix based on true/false classification

	AI OK	AI Error
Doctor OK	320	48
Doctor Error	49	93

3.1.6 시스템 분류 성능평가 지표

위에 분류에 따라서 성능지표를 계산 할 수 있는데 아래의 계산식에 의해 Precision(정밀도), Recall(재현율), Accuracy(정확도), F1 score, Fall-out를 검출할 수 있다.

$$\begin{aligned}
 (Precision) &= \frac{TP}{TP + FP} \\
 (Recall) &= \frac{TP}{TP + FN} \\
 (F1-score) &= 2 \times \frac{1}{\frac{1}{Precision} + \frac{1}{Recall}} = 2 \times \frac{Precision \times Recall}{Precision + Recall} \\
 (Accuracy) &= \frac{TP + TN}{TP + FN + FP + TN} \\
 Fall-out(FPR) &= \frac{FP}{TN + FP}
 \end{aligned}$$

Fig. 20. Equations for performance parameters

성능평가 지표에 의해 산출된 식에 의한 결과 값은 Precision(정밀도) 0.87, Recall(재현율) 0.87, Accuracy(정확도) 0.81, F1 score 0.87, Fall-out 0.34이다.

Table 2. The values of the performance parameters

	value
Precision(정밀도)	8.796e ⁻³
Recall(재현율)	8.672e ⁻³
Accuracy(정확도)	8.098e ⁻³
F1 score	8.684e ⁻³
Fall-out	3.404e ⁻³

머신 러닝의 효율을 높이고자 vertical flip, rotation, width&height shift 등의 기법을 실시하였으며 모델의 파라미터 학습을 위해서 애덤 최적화 기법을 사용하여 학습변수는 2.0 x 10e-4에서 시작하여 validation loss가 가장 낮은 안정화 지점에서 중지하였다.

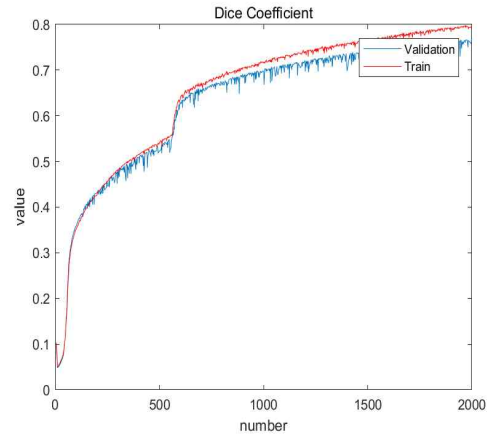


Fig. 21. Dice coefficients for the validation and training datasets

Dice coefficient는 다이스 계수라고 불리는데 IOU(Intersection over union)의 개념이랑 비슷하다. 두 영역의 교차되는 부분의 넓이를 모든 부분을 계산했을 때의 값으로 나눈 값이다. 예측된 경계의 정확도를 평가하고자 할 때 많이 쓰이고 예측된 경계 상자와 실제 참값에 대한 IOU를 잡고자하는 해당 상자의 정확도라 생각하면 된다. 1에 가까울수록 성능이 좋고 0에 가까울수록 성능이 나쁘다.

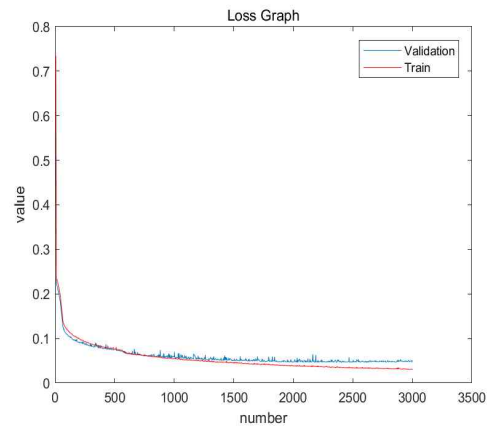


Fig. 22. Loss values for the validation and training dataset

loss value graph는 손실 값을 나타내는 것으로 ai가 정답과 확인할 때 출력 값이 정답에 가까워질수록 0에 가까워진다.

3.1.7 시스템 진단 수치 비교

시스템에 의해서 정상적으로 판단된 유효 데이터로 AI와 Doctor의 각도 차이를 비교해 보면 큰 차이를 보이지 않는 것을 확인했다.

base angle은 max는 6.93도 min은 0도 평균은 1.81도 차이를 확인하였다.

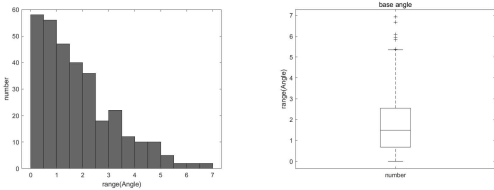


Fig. 23. Base angle range

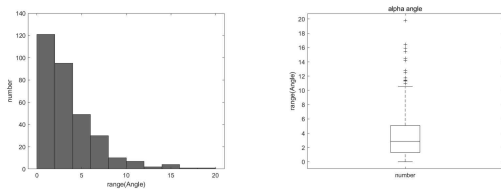


Fig. 24. Alpha angle range

beta angle은 max는 25.92도 min은 0.01도 평균은 4.51도 차이를 확인하였다.

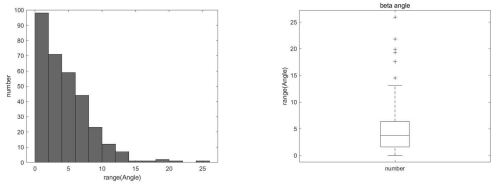


Fig. 25. Beta angle range

IV. 고찰

본 논문에서 다중트레이닝 기법을 이용하여 기존 트레이닝 방법보다 적은 양의 데이터로 트레이닝 실시 하였으며 다른 이미지 영상보다 복잡한 초음파 사진으로 본 논문에서 제시한 다중 트레이닝 기법을 사용하였을 때 기존 방법보다 성능이 뛰어난 것을 확인하였다. 본 기법을 이용하게 되면 적은 양의 데이터로 ai를 구축해야 하는 소규모 연구실이나 개발팀에서도 사용될 것으로 보이고 본 기법을 베이스로 하여 많은 양의

데이터에 대해 러닝을 수월하게 하는 시드로 활용 가능할 것으로 보인다.

논문에 사용된 기법은 인공지능에서 사용자의 개입에 따른 정도에 따라 머신 러닝, 딥 러닝 등으로 구분된다. 원본 영상에서 판단에 사용되는 부분을 먼저 트레이닝한다. 검출된 부분 이외에 영역을 모두 제거하여 노이즈에 대한 필터링을 강화시킨다. 각도에 해당하는 지점을 재 트레이닝한다. 기존 트레이닝 방법보다 적은 양의 데이터로도 필요 부분을 검출시킬 수 있도록 만들었다.

실험에 사용된 데이터는 2,692장이지만 초음파 특성상 흔들림이 심하거나 노이즈가 심하거나 사람의 시각으로 확인할 수 없는 데이터는 전부 배제하였다. 그 결과 트레이닝에 사용된 데이터는 총 1,242장이었다. 이 중에서 모델링 구축을 위한 데이터로 321장을 가져왔다. 생성된 모델링을 통해서 테스트할 데이터로 920장으로 분류하였다. 먼저 모델링 구축을 위해서는 모델링 형성을 위한 트레이닝 데이터와 학습 안정화를 위한 테스트 데이터가 필요하다. 각각 288개의 트레이닝 데이터와 33개의 테스트 데이터로 학습했다. 2차 트레이닝은 1차 트레이닝에 사용한 데이터를 변환한 파일로 학습하였다. 유효 데이터를 수집 시에 가장 어려웠던 점은 대상자가 영유아라는 점과 초음파가 CT나 MRI 같은 다른 의학 영상보다 아티팩트(Artifact)가 크다는 점이다. 경력 5년 이상의 의료진이 촬영했지만 정상 데이터 획득률은 50%에도 미치지 못했다. 의학적인 요소가 추가되었기 때문에 의학적인 측정방식으로 해당 사진을 6가지로 분류하였다. 6가지의 판별 요소는 실제로는 복잡한 기준을 필요로 한다. 정량적 측정을 위하여 기준을 의사와 AI의 사진판별 답안지를 OK, Error, NA로 선택지를 고정하였다. OK는 정상인 데이터에 정상 범위의 값을 나타낼 경우를 의미한다. Error는 정상 데이터지만 base angle값이 정상 범위가 아니라서 비정상 이미지를 의미한다. NA는 non-available로 값을 도출해 낼 수 없는 경우를 의미한다. 성능 측정 방식은 일반적인 모델 평가 분류 지표에 의해서 구별하였다. 의사가 감지할 수 없는 NA 선택지를 제외한 데이터 510개를 검출하였다. 의사와 인공지능의 선택지의 OK와 Error를 2가지 정답을 기준으로 총 4가지로 계산하였다. AI의 성능 평가로 성능평가지표와 Dice coefficient와, loss graph를 사용하였으며 결과 값은 아주 양호하게 검출되었다. 의학적인 성능 평가로는 알

과 베타각의 차이를 비교하였다. 평균으로 비교해 보면 4도 이하의 차이가 발생했다. 의시간에도 동일한 사진을 기준 측정 시 평균적으로 발생하는 차이의 범 위라서 좋은 결과다.

ACKNOWLEDGMENTS

This research was supported by Basic Science Research Program through the National Research Foundation of Korea(NRF) funded by the Ministry of Education (NRF-2017R1D1A1B04031182).

REFERENCES

[1] J. K. Author, "지속적인 심장질환을 위한 인체 삽입형 생체 센서의 무선전력전송 시스템," 융합신호처리학회 논문지, vol. 13, no. 1, pp. 57-64, 2019.

[2] D. P. Kingma, J. Ba, "Adam: A method for stochastic optimization," 2014.

[3] K. He, G. Gkioxari, P. Dollár, R. Girshick, "Mask R-CNN," Proceedings of the IEEE International Conference on Computer Vision (ICCV), pp. 2961-2969, 2017.

[4] A. Krizhevsky, I. Sutskever, G. Hinton, "Imagenet classification with deep convolutional neural networks," In NIPS, 2012.

[5] K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition," In CVPR, 2016.

[6] E. Meijering, "Cell segmentation: 50 years down the road life sciences," IEEE Signal Process. Mag., vol. 29, no. 5, pp. 140-145, 2012.

[7] R. W. Paton, S. Hossain, K. Eccles, "Eight-year prospective targeted ultrasound screening program for instability and at-risk hip joints in developmental dysplasia of the hip," J Pediatr Orthop, vol. 22, pp. 338-341, 2002.

[8] R. W. Paton, S. Hossain, "Histopathological image analysis: A review," IEEE Rev. Biomed. Eng., vol. 2, pp. 147-171, 2009.

[9] E. A. Roovers, M. M. Boere-Boonekamp, R. M. Castelein, G. A. Zielhuis, T. H. Kerkhoff, "Effectiveness

of ultrasound screening for developmental dysplasia of the hip," Arch Dis Child Fetal Neonatal Ed, vol. 90, pp. F25-F30, 2005.

[10] K. J. Holen, A. Tegnander, T. Bredland, O. J. Johansen, O. D. Saether, S. H. Eik-Nes, T. Terjesen, "Universal or selective screening of the neonatal hip using ultrasound? A prospective, andomised trial of 15,529 newborn infants," J Bone Joint Surg Br, vol. 84, pp. 886-890, 2002.

저자 소개

황 석 민 (Seok-Min Hwang)



2016년 3월 : 계명대학교
의용공학과(공학사)
2020년 12월 : 계명대학교
의학과(공학석사)
2013년 1월~현재 : 인공지능
진단연구실 연구원
관심 분야 : AI, CNN, medical
image processing, 신호처리

이 시 옥 (Uk-Si Lee)



2000년 3월 : 계명대학교
의학과(의학사)
2014년 2월 : 계명대학교
의학과(의학석사)
2016년 2월~현재 : 계명대학교
의학과(의학박사)
관심 분야 : AI, CNN, medical
image processing, 신호처리

이 종 하 (Jong-Ha Lee)



2016년 8월 : New York
University, Electrical Engineering
(공학석사)
2012년 8월 : Temple University,
Electrical Engineering (공학박사)
2012년 9월~현재 : 계명대학교
의과대학 의용공학과 부교수
관심 분야 : 컴퓨터 진단, 인공지능, 지능형 생체센서