

스마트 헬스 시티 구현을 위한 얼굴 영상기반 심박수 추출 딥러닝 모델 최적화

김 광 호*, Yuli Sun Hariyani*, 한 승 우**, 이 현 빈***, 손 량 희****, 박 철 수^o

Optimization of Deep Neural Networks for Heartrate Estimation from Face Video Stream to Implement Smart Health-City

Gwangho Kim*, Yuli Sun Hariyani*, Seungwoo Han**, HyeonBin Lee***,
Ryanghee Sohn****, Cheolsoo Park^o

요 약

컨택트 센서 대신에 비접촉 비디오 이미지로부터 생체 정보를 얻고자 하는 연구가 활발히 진행되고 있는데, 얼굴 영상으로부터 심박수 추정을 하는 연구가 한 예이다. 과거의 경우 얼굴의 작은 움직임을 이용하여 신호처리 방식으로 심박수를 추정하였는데, 최근에는 얼굴의 색 변화를 딥러닝 알고리즘으로 분석하여 심박수를 추정하는 연구가 진행 중이다. 본 논문에서는 딥러닝 분석 방법을 이용하여 얼굴 영상으로부터 심박수를 추정하고자 하며, 특히 정확도 향상을 위한 딥러닝 구조 최적화 연구를 수행하였다. 얼굴 영상 추출 알고리즘을 사용하여 움직임 정보를 손실시키고, 얼굴 시맨틱 세그멘테이션 방법을 통해 피부 정보만 추출하여 이를 딥러닝 알고리즘의 입력으로 사용한다. 또한 딥러닝에 존재하는 하이퍼 매개변수는 지금까지 대다수의 연구에서 개발자의 경험에 의해 결정되었지만, 본연구에서는 이를 베이지안 최적화와 하이퍼 밴드 기법을 통해 자동 최적화된 하이퍼 매개변수를 가지는 모델을 제안한다.

Key Words : Heartrate, semantic segmentation, video stabilize, BOHB, CNN

ABSTRACT

Lots of studies are conducted to extract information about user's condition from video streams instead of contact-based sensors, such as the estimation of heart rate from facial video streams. Heart rate provides various information about health conditions. In the past, it was extracted relying on traditional signal processing method detecting small movements of his/her face. However, recently researchers has conducted the heart rate estimation process using deep learning approach analyzing color changes of a face. In this paper, the heart rate is estimated from facial video streams using a deep learning model and a facial semantic segmentation method, which could remove motion artifacts, and only extracts skin information. Hyperparameters of the deep neural networks are optimized through an automatic approach, Bayesian Optimization and Hyperband

※ 본 연구는 과학기술정보통신부 및 정보통신기술진흥센터의 SW중심대학지원사업의 연구결과로 수행되었음 (2017-0-00096)

◆ First Author : Kwangwoon University Department of Software, rhkd865@gmail.com, 학생회원

◦ Corresponding Author : Kwangwoon University Department of Computer Engineering, parkcheolsoo@kw.ac.kr, 정회원

* Kwangwoon University Computer Engineering Department, yulisun@telkomuniversity.ac.id

** Kwangwoon University Department of Intelligent Information System and Embedded Software Engineering, seungwoohan@kw.ac.kr

*** Kwangwoon University Department of Software, hyunbinkong@gmail.com

**** Emma Healthcare Co., Ltd., chief@emmahc.com, CEO

논문번호 : 202011-276-0-SE, Received November 6, 2020; Revised November 22, 2020; Accepted November 23, 2020

I. 서 론

여러 기술을 활용하여 도시 문제를 해결하려는 시도가 있었었는데, 최근 AI 국가전략에 스마트시티가 포함되어 더욱 많은 관심을 받고 있다. 이에 따라, 최근 학계에도 스마트시티에 대한 연구가 도시 계획 및 설계뿐만 아니라 법률, 제도, 거버넌스, ICT 플랫폼, 서비스, 도시 경쟁력, 시민 삶의 질, 시민 참여를 통한 혁신 등 매우 다양한 관점에서 활발히 진행되고 있다.^[1,2] 시민 참여에 대한 분야별 정의와 필요성이 대두되어 이제는 스마트시티에서 생활하는 사람들의 니즈나 참여에 관심을 두어야 한다.^[3]

개인의 건강에 대한 관심이 증가하고 있는데, 개인 건강의 척도로 심박수가 오래전부터 사용되어지고 있다. 심박수는 건강관리에 필요한 생체 신호 중 하나로 각광받고 있다. 하지만 웨어러블기구나 모바일 디바이스의 컨택트 기반 바이오센서로 생체 물리량을 측정해야 하는 제약이 있는데, 일상생활 중 이러한 장치를 착용하는 것은 사용자에게 불편함을 줄 수 있다.

심박수를 측정하는 방법은 맥박을 세는 것과 두 맥박 사이의 시간 간격을 가지고 추정하는 방법^[4]이 있다. 전자의 경우 정확히 시간을 재면서 세지 않을 경우 값 사이에 일정한 간격을 가지게 된다. 후자의 경우 실수 형태를 가질 수 있으나, 반올림을 통해 근사할 수 있다. 더 적은 시간으로 알맞게 심박수를 추정하고자 후자의 방식이 주로 활용되고 있다.

현재까지 대부분 컨택트 기반 바이오센서를 이용하여 생체 신호를 추정하였는데, 이미지 기반 기술이 발전함에 따라 이미지로부터 많은 정보들을 추출할 수 있게 되었다. 이로 인하여 사람 영상으로부터 생체 신호를 얻는 연구도 활발히 진행되고 있다. 과거의 경우 얼굴의 작은 움직임만을 특징으로 이용하여 전통적인 영상처리나 신호처리 기법을 통해 맥박 신호를 추정하고자 했다.^[5,6] 하지만 최근에 얼굴의 색과 심장 박동이 높은 상관관계가 있다는 것이 입증되었다.^[7] 이로 인해 딥러닝을 활용하여 얼굴 영상으로부터 맥박 신호를 추정하는 연구가 활발히 진행되고 있는데, 그 중 맥박 신호가 아니라 심박수를 바로 예측하는 연구도 진행되고 있다.^[8]

본 논문에서는 영상 안정화 알고리즘을 사용하여 움직임 정보를 손실 시키고, 오로지 색 정보만을 사용하여 학습한다. 또한 시멘틱 세그멘테이션 기법을 통해 얼굴의 피부를 제외한 나머지 영역을 제거하여 필요한 특징 부분만을 남긴다. 또한 베이지안 최적화와 하이퍼 밴드 기법을 통해 하이퍼 파라미터를 최적화

한 모델을 제시한다.

2장에서는 스마트 시티에 대한 소개와 실증에 대해 설명하고, 3장에서는 실험에 필요한 데이터 전처리 및 하이퍼 파라미터 최적화에 사용된 연구들을 소개한다. 4장에서는 실험에 대한 전처리 및 방법에 대한 과정과 결과를 보여준다. 마지막 5장에서 결론을 통해 끝맺음을 짓는다.

II. 배 경

도시화와 경제 성장으로 인해, 도시의 인구가 빠르게 증가하고 있다. 이는 도시가 수용 가능한 범위를 초과하여 행정, 주택, 교통, 에너지 등의 자원 부족 문제들을 발생시킬 수 있다. 최근에는 도시 거주자들의 삶의 질이 문제시되면서 도시 사회 문제가 날로 증가하고 있다. 해당 문제들을 해결하기 위해 많은 시도가 이 진행되고 있는데, 그중 지속 가능한 도시 모델로서 스마트 시티가 주목을 받고 있다.^[9]

스마트 시티에 대한 개념은 최근 생겨나, 그 정의가 기술적, 사회적, 국가별로 다르다. ITU-T(국제 전기통신연합 전기통신 표준화 부문)은 스마트 시티를 정보통신기술(ICT) 및 기타 수단을 사용하여 삶의 질, 도시 운영 및 서비스의 효율성 및 경쟁력을 향상시키는 혁신적인 도시로 정의하였다. 이는 기술적 관점으로 사회적 관점을 해결하려는 정의이므로, 기술을 통한 문제 해결이 중요하게 여겨진다. 이러한 기술로는 교육, 보안, 에너지, 인프라, 교통, 헬스케어, 건축 등이 있다. 관련 스마트 시티 연구 문헌 출판 현황^[10] 볼 때, 컴퓨터 분야와 엔지니어링 분야가 50% 넘는 상황을 볼 수 있다. 기존의 도시 문제를 ICT 분야와 융합하여 해결하고자 하는 것으로 알 수 있다.

스마트 시티의 시장이 최근 빠르게 성장 중인데, 그 중 스마트 헬스케어는 연 7% 정도의 성장률을 보여준다. 스마트 헬스케어의 가치 중 하나는 개인 맞춤형 건강관리 프로그램인데 특히 50대 이상일수록 이러한 서비스를 많이 이용한다^[11]. 이는 개인의 건강에 대한 관심이 증가하면서, 건강을 모니터링할 수 있는 서비스에 대한 수요가 높아졌기 때문이다. 특히 심장은 이상이 생길 시 긴급한 처치가 필요한데, 개인의 생체신호를 모니터링하면서 응급상황을 대비할 수 있는 장점이 있다.

생체신호는 주로 사물 인터넷 기반 웨어러블 디바이스의 컨택트형 바이오센서를 이용하여 얻을 수 있다. IoT 기반 헬스케어 시장 또한 꾸준히 증가하고 있다.^[12] 하지만 이는 컨택트형 디바이스에 종속적이기

때문에 착용하지 않는다면, 이상 유무를 파악하지 못하는 단점이 생긴다. 그리하여 최근에는 비접촉 생체 신호 측정 시스템이 개발 중이다. 비접촉 심장 박동수를 추정하는 기술은 운동 중 손목에 대한 부담을 제거할 수 있고, 원격 의료 같은 분야에도 활용할 수 있다.

과거 심장박동수를 추정하기 위해 주파수를 통한 전통적인 신호처리 방식을 사용하였다. 그 이후에 얼굴의 미세한 움직임을 주파수 분석을 통해 얻으려는 시도가 존재했다. 하지만 최근 딥러닝이 화제가 되면서 많은 시도가 진행 중이다. 연구자들은 혈액량에 따라 얼굴의 혈색이 의미 있는 수준으로 변하는 것을 입증하게 되었고, 이로써 얼굴의 혈색을 특징으로 사용할 수 있게 되었다. 즉 얼굴의 특정 영역만을 사용했던 과거에서 얼굴 전체를 사용할 수 있게 되었다.

과거 심박수는 이미지에 대한 수학적 모델링 통해 추정을 해왔다. 하지만 이는 정확하지 않아서 딥러닝을 통해 해결하고자 한다. 딥러닝을 사용하는 데 있어서 중요한 것은 데이터, 구조, 하이퍼 파라미터이다. 성능 개선에 있어서, 사람의 손이 많이 가는 것은 하이퍼 파라미터이다. 개발자의 경험에 의존할 수밖에 없고, 아무리 구조가 좋아도 적절한 하이퍼 파라미터를 설정하지 않으면, 모델의 성능이 떨어질 수 있다.

III. 제안 알고리즘

3.1 컨볼루션 신경망

컨볼루션 필터는 이미지의 주요 특징을 추출하기 위해 사용되는 필터이다. 해당 필터를 좀 더 다양하고 용도에 알맞게 사용하고자 필터의 크기나 연산 방법을 조정할 수 있는 컨볼루션 신경망이 나왔다. 배치 사이즈와 필터 개수를 제외한 차원의 개수에 따라 n 차원 컨볼루션 필터로 나누어진다. 또한 가중치가 필터로 국한되기 때문에 적은 가중치를 사용하고, 희소 연결성을 확보하여 데이터를 순회할 수 있다.^[13]

컨볼루션 신경망은 이미지 인식 대회에서 인간의 수준을 뛰어넘는 성과를 보였는데, 이를 계기로 많은 분야에서 적용하였다.^[14,15] 컨볼루션 신경망의 층이 증가할수록 원본 이미지 대비 필터가 인식하는 부분도 커지게 된다. 초기의 컨볼루션 층은 낮은 추상화 특징들을 보게 되고, 층이 깊어질수록 고차원적으로 추상화된 특징들을 보게 된다. 초기의 낮은 특징들은 다른 문제의 데이터에 대해서도 의미가 있을 수 있는데, 이러한 경우 전이 학습^[16]을 진행하기도 한다.

3.2 얼굴 영상 안정화 알고리즘

영상을 촬영하는 데 있어 카메라나 모델의 불필요한 움직임은 노이즈로서 제거해야 할 대상이다. 이동에 의해 연속된 2개의 이미지는 서로 달라지게 되는데, 이중 어떤 것을 기준으로 정할지가 중요하다. 기존에는 코너 검출기를 통해 영상의 특징점을 추출하고 이를 기준으로 이미지를 매칭 시킨다. 그 후 이동 거리 및 방향 등을 파악하여 거리를 좁히는 방식으로 영상을 안정화 시킨다.^[17]

딥러닝 또한 특징 추출이 가능한 기술로, 얼굴의 랜드마크 등을 추출할 수 있어 해당 분야에 적용되었다. 얼굴 영상 안정화 알고리즘은 객체를 얼굴로 한정하여 얼굴이 이미지의 한 가운데에 놓이는 것을 목표로 한다. MTCNN^[18]은 얼굴의 중요한 특징점 중 하나인 눈, 코, 입을 추정할 수 있는데, 이 점들을 이용하여 프레임 간의 간격을 최소화하는 방식으로 안정화시킬 수 있다. 최근에는 얼굴의 랜드마크를 추출하고, 최적화 문제로 바꾸어 흔들림을 최소화하는 연구가 진행 중이다.^[19]

3.3 얼굴 시맨틱 세그멘테이션

이미지 인식 분야에서 사람의 수준을 뛰어넘기 시작하자 이미지에서 객체가 어디에 있는지까지 추출하는 탐지 기술이 연구되고 있다. Fast RCNN^[20]과 Faster RCNN^[21]의 큰 차이점은 객체가 있을 지역을 제안하는 방식이다. 기존 영상 처리 기법은 시간과 연산이 많이 소모되었는데, 딥러닝으로 대체하면서 시간이 줄어들고 있다. 이들은 물체가 존재하는 지역을 잘 찾도록 하고 있다.

기본적인 물체 탐지 기술은 바운딩 박스(사각형) 형태로 지역을 분할했는데, 이는 물체의 모양을 신경 쓰지 않은 형태이다. 그 후 시맨틱 세그멘테이션^[22] 기법이 등장했다. 이는 물체의 위치를 픽셀 단위로 구별하여 물체의 모양에 알맞은 지역을 제안하고자 한다. 초기에 물체의 유무만을 판단하여 물체에 대한 분류가 힘들었지만 Faster RCNN에 FCN 층을 추가한 Mask RCNN^[23]이 등장하면서 각 물체에 대한 구별도 가능해졌다.

3.4 베이지안 최적화와 하이퍼밴드

기존의 하이퍼 파라미터를 최적화하는 방식은 사람이 경험에 기반하여 손으로 파라미터를 수정하면서 최적화하거나 무작위나 체계적으로 선택하여 성능을 비교하는 방식으로 진행되었다. 그 이후 베이지안 최적화^[24]가 등장하였는데, 이는 미지의 함수를 최대

하는 최적 해를 찾는 기법이다. 알고 있는 점을 가지고 함수를 추정하는 방식이 위 방식들과 큰 차이점을 보인다. 이 기법을 충분히 반복하면 빠르게 최적해를 근사할 수 있다.

딥러닝의 학습 횟수 또한 하이퍼 파라미터에 속하는 데, 정확한 학습 횟수를 사용하지 않으면 모델의 정확한 성능을 얻어내기 어렵게 된다. 이 문제를 해결하기 위해 하이퍼 밴드^[25]가 등장하였는데, 이것은 임의의 하이퍼 파라미터를 설정하고 짧은 학습 후 높은 성능을 가지는 파라미터에 대해 길게 학습하면서 최종적으로 한 개의 파라미터만 남기는 기법이다.

베이지안 최적화와 하이퍼 밴드^[26] 기법은 이름 그대로 위 2가지 기법을 조합한 것으로 계산 효율을 증가시킨다. 하이퍼 밴드를 통해 딥러닝의 성능을 평가하고, 베이지안 최적화를 통해 최적의 하이퍼 파라미터를 찾는 과정을 반복한다. 이는 기존의 방법보다 학습 속도가 빠르고, 최적해에 가까운 하이퍼 파라미터를 좀 더 빨리 찾을 수 있다.

IV. 실험 방법

본 논문에서는 얼굴 영상과 맥박 신호를 제공하는 공개 데이터 셋을 사용하였다. 데이터 전처리로 얼굴 영상 안정화 기법과 시멘틱 세그멘테이션 기법을 사용하였다. 마지막으로 모델 최적화를 위하여 베이지안 최적화와 하이퍼 밴드 기법 적용을 제안한다.

4.1 데이터 셋

이미지 영역에서 빛은 번거로운 특징 중 하나에 속한다. 모델이 학습할 때 빛의 영향을 적게 받지 않기 때문에, 서로 다른 데이터 셋을 사용하고자 할 때는 많은 주의가 필요하다. 학습 데이터와 테스트 데이터에서의 데이터 형태가 서로 달라 성능이 나빠질 수 있기 때문이다. 오픈된 데이터 셋은 5가지 정도가 존재하는데, 카메라 프레임과 빛에 대한 상황이 모두 다르다.

MANHOB-HCI 데이터 셋^[27]은 30명(13 남성, 17 여성)의 앉은 자세를 촬영했다. 대체적으로 어두운 영상이지만, 카메라가 초당 61프레임을 촬영하여 많은 정보를 지니고 있다. 영상과 동시에 초당 256 샘플 개수로 측정된 심전도 신호를 통해 정확한 심박수 정보를 예측할 수 있다.

ECG-Fitness database^[28]는 17명(14 남성, 3 여성)의 실험 대상자들이 대화 및 운동을 하는 영상을 초당 30프레임으로 촬영하였다. 영상 데이터는 3가지 조명

으로 나눌 수 있는데, 자연광, 400W 할로겐 조명과 30W LED 조명의 데이터를 가진다. 데이터베이스 내 실험 대상자들은 최소 심박수는 56BPM (bit per minute), 최고 심박수는 159BPM이며, 평균 108.96 BPM의 심박수를 갖고 있다.

COHFACE 데이터 셋^[29]은 40명(28 남성, 12 여성) 실험 대상자들의 앉은 자세를 1분 간격으로 촬영했다. 크게 그림자가 없는 영상 2개와 그림자가 존재하는 영상 2개로 나누어 녹화했다. 또한 카메라의 성능이 20Hz로 조금 낮은 프레임 수를 가지고 있다. 초당 256 샘플로 기록된 PPG 신호 정보가 포함되어 있으며, 심박수를 얻기 위해서는 추가적인 PPG 신호 peak detection 과정이 필요하다.

PURE 데이터 셋^[30]의 경우 10명의 실험자(8 남성, 2 여성)를 1분 간격으로 초당 30프레임, 녹화 성능을 갖는 카메라로 녹화했다. 정자세, 대화, (느린, 빠른) 머리 움직임, (느린, 빠른) 머리 회전을 포함한 6가지 상황을 가정했다. 전체적으로 다른 데이터 셋에 비해 영상이 어둡다.

UBFC-RPPG 데이터 셋^[31]의 경우 42명의 실험자를 대상으로 1분 간격으로 초당 30프레임 녹화성능을 갖는 카메라로 녹화했다. 앉은 상태에서 심박수를 높이고자 수학 문제를 푸는 상황을 설정하였다. 조명이 풍부한 실내에서 녹화를 수행하여 전체적으로 영상이 밝다.

4.2 데이터 전처리

영상을 10초 단위로 자른 후, 이 10초 동안의 영상을 입력으로 사용하였고, 해당 시간 동안의 평균 심박수를 목표 데이터로 사용하였다. 얼굴의 색 변화를 사용하는 데 있어서 얼굴의 움직임은 노이즈에 속한다. 얼굴 안정화 알고리즘을 사용하여 얼굴을 정중앙에 위치시키도록 하여 움직임을 최소화시켰다. 색 변화이



그림 1. 전처리된 이미지
Fig. 1. Preprocessed image

기 때문에 피부가 가장 큰 특징이므로 얼굴 시맨틱 세그멘테이션 기법을 사용해 배경을 삭제하고, 그림 1과 같이 피부 영역만을 남겨 두었다. 정수형 이미지를 실수 형태로 바꾸면서 최대 픽셀 값으로 나누어 주었다.

4.3 평가 지표

딥러닝 모델의 성능을 평가하기 위해 아래 식(1)과 (2)와 같이 절대 오차의 평균 및 오차 제곱의 평균을 사용하였다.

$$Mean\ Squared\ Error = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2 \quad (1)$$

$$Mean\ Absolute\ Error = \frac{1}{n} \sum_{i=1}^n |Y_i - \hat{Y}_i| \quad (2)$$

식(1)을 미분할 경우 오차가 반영되기 때문에 손실 함수로 사용하였고, 식(2)의 경우 오차의 차이를 명확히 알 수 있어서 평가 지표로 사용하였다. 식에서 n은 데이터 개수, Y는 실제 값, \hat{Y} 은 예측된 값을 뜻한다.

4.4 모델

비디오 영상을 분석하며 시계열 데이터를 처리하기 때문에 3D CNN으로 진행되는 방법과 각 시간에 따라 2D CNN으로 진행되는 방법이 존재한다. 전자의 경우 파라미터의 개수가 후자보다 증가하게 되지만, 시간에 따른 값 차이도 계산할 수 있다는 장점이 존재한다.

본 논문에서는 VGG-16^[32]의 모델 구조에서 2D CNN을 3D CNN으로 변경하고 마지막 Fully-Connected 층 하나를 제거한 구조를 사용하였다. 베이지안 최적화와 하이퍼 밴드 기법으로 탐색할 하이

퍼 파라미터 구성은 표 1과 같은데, 최대 필터 개수는 제한된 그래픽카드의 메모리에 의해 결정되었다. Pooling의 경우 첫 번째 CNN 다음에서만 선택하도록 하였고, 나머지는 Average Pool를 사용하였다. 최적화 방법은 아담을 사용했고, 활성화 함수는 Leaky-ReLU를 사용하였다. 학습률의 경우 0.00001을 사용하였다.

표 1. 하이퍼 파라미터 탐색 공간
Table 1. Search spaces of the hyper-parameters

Layer	Search Space
Conv1 Filter	8 ~ 16
Conv2 Filter	16 ~ 64
Conv3 Filter	32 ~ 128
Conv4 Filter	32 ~ 256
Conv5 Filter	32 ~ 512
Conv6 Filter	2 ~ 64
Pooling	Max Pool, Average Pool
Dropout	0.1, 0.2, 0.3
Fully Connected Unit	8 ~ 128

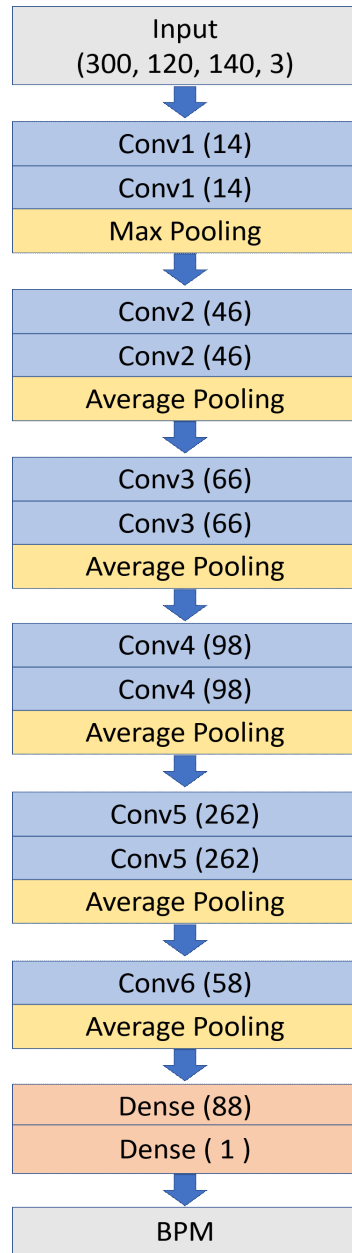


그림 2. 모델 구조
Fig. 2. Architecture of network

Nvidia RTX 2080 Super로 대략 6일 동안 300개 정도의 하이퍼 파라미터를 탐색하였다. 하이퍼 파라미터를 탐색할 경우 연산량 문제로 인해 한 가지 방법의 데이터 셋만 사용이 가능하다. 하지만 이는 일반화 능력이 떨어질 수 있기 때문에 최종적으로 잘 나온 모델들 중 일부를 선택하여 학습 횟수를 늘린 후 성능을 체크하였다. 그리고 해당 학습에서 일정 부분을 줄인 후 cross-validation 기법을 적용하여 모델의 평균 성능을 평가하였다. 최종적으로 가장 높은 성능을 보여주는 하이퍼 파라미터 조합을 얻을 수 있었고, 해당 하이퍼 파라미터는 그림 2와 같다.

컨볼루션이 진행될 때마다 필터의 개수가 증가하는 양상을 보여주고, 마지막 컨볼루션에서만 필터의 개수가 감소하였다.

V. 실험 방법

그림 3은 UBFC-RPPG 데이터 셋을 사용하여 얻은 심박수 예측 결과를 실제 값과 함께 그래프로 표현한 것이다. 빨간색 점은 예측한 값이고, 파란색 점은 실제 값이다. 12번째 데이터 아래에서는 예측한 값이 실제 값과 비슷한 양상을 보여 주었다. 하지만 23번째 데이터부터는 양상이 비슷하지만, 오차는 커졌다. 심박수의 격한 변화는 예측을 잘하지만, 오히려 미세한 변화를 정확히 예측하지는 못하는 양상을 띄었다.

성능 평가를 위하여 UBFC-RPPG 데이터 셋과 PURE 데이터 셋을 사용하였다. cross validation을 통해 얻은 성능들의 평균과 표준편차를 구하여 모델의 실제 성능으로 판단하였다. UBFC-RPPG의 경우 밝은 영상으로 인하여 색 변화를 잘 감지할 수 있어서 평균 성능이 4.128 MAE(mean average error)로 좋게

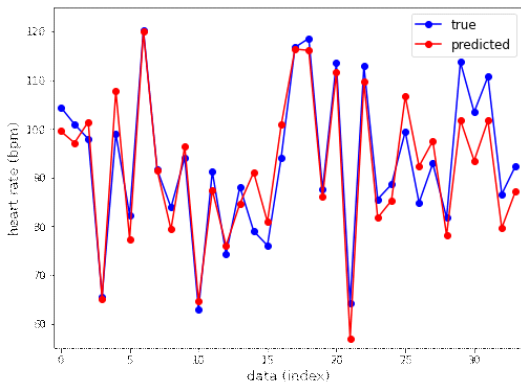


그림 3. 예측된 값의 결과
Fig. 3. Result of predicted data

표 2. 모델 성능
Table 2. Model performance

	Mean of MAE	Std of MAE	Mean of MSE	Std of MSE
UBFC	4.128	1.112	27.594	14.154
PURE	4.918	0.711	39.528	11.685

나왔지만, 표준 편차가 1.112로 성능에 대한 편차가 크게 나타났다. PURE 데이터 셋의 경우 대체로 어두운 이미지가기 때문에 색 변화도 적어서 평균 성능이 4.918 MAE로 적게 나왔지만, 표준 편차가 0.711로 cross validation에 대한 성능 차이가 적었다. PURE 데이터 셋의 경우 다양한 환경이 존재함으로 인하여, 이에 따른 다양한 노이즈가 제공되고, 모델이 이를 잘 학습하여 큰 성능 차이를 보이지 않은 것 같다.

VI. 결과 토의

딥러닝 최적화 방법은 비록 높은 컴퓨터 성능이 필요하지만 개발자의 경험에 의존하는 방법보다 더 좋은 성능을 기대할 수 있다. 사전 지식은 탐색 공간을 정의할 때 쓰일 수 있는데, 이러한 탐색 공간은 개발자의 경험에 의존하여 적절히 정의해야 보다 높은 성능을 얻을 수 있다. 그러나 탐색공간을 매우 넓게 정의한다면, 개발자의 경험을 뛰어 넘는 새로운 공간에서의 최적점을 찾아 더 높은 성능 또한 기대할 수 있을 것이다. 본 논문에서는 고정된 네트워크 구조에 대해 하이퍼 파라미터를 탐색하였는데, Neural Architecture Search와 같이 다양한 구조에 대한 탐색도 진행된다면 더 최적화된 모델을 찾을 수 있을 것이다.

찾은 하이퍼 파라미터의 양상을 보면 마지막 필터의 개수가 56개로 시간 축이 300인 걸 감안하면 많은 입력 값이 남게 되는 것을 볼 수 있어 네트워크의 층을 더 늘려도 좋은 결과를 얻을 수 있을 것이다. 또한 마지막 CNN 층과 Dense 층 사이의 파라미터가 매우 커지게 되는데, 이 부분을 해결하면 더 좋은 성능을 얻을 수 있을 것이다. 그렇지 않고, 원래 VGG-16처럼 Dense 층의 유닛을 더 키워서 특징의 차원을 더 늘려줘도 좋을 것 같다. 이는 특징의 손실이 큰 것 같기 때문이다.

VII. 결론

본 논문에서는 스마트 헬스 시티 구현을 위하여 사용자가 일상생활 중 무구속적으로 손쉽게 심박수를

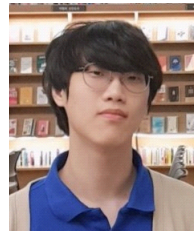
측정할 수 있는 영상기반 센서 연구를 소개하였다. 딥러닝의 최적화를 통해 영상을 통한 심박수 예측의 높은 정확도를 얻어냈으며, 이를 통해 비접촉 건강 모니터링 시스템 구현에 한 발짝 더 나아가게 되었다.

References

- [1] H.-J. Chang, "A study on the procedure of using big data to solve smart city problems based on citizens' needs and participation," *J. KIIECT*, vol. 13, no. 2, pp. 102-112, Apr. 2020.
- [2] J. Y. Chang, E. Lee, and W. Suh, "An analysis on the smart city research trends," *J. Korean Assoc. Regional Inf. Soc.*, vol. 23, no. 2, pp. 147-170, Jun. 2020.
- [3] Y.-J. Jeon, Y. Choi, M.-J. Kim, J.-W. Lee, and B.-J. Park, "Potentiality of urban forest roads as an environment for enhancing physical fitness," *Korean J. Agricultural Sci.*, vol. 38, no. 1, pp. 109-113, Mar. 2011.
- [4] E. Peper, R. Harvey, I. M. Lin, H. Tylova, and D. Moss, "Is there more to blood volume pulse than heart rate variability, respiratory sinus arrhythmia, and cardiorespiratory synchrony?," *Biofeedback*, vol. 35, no. 2, pp. 54-61, Jan. 2007.
- [5] G. Balakrishnan, F. Durand, and J. Guttag, "Detecting pulse from head motions in video," in *Proc. IEEE Conf. CVPR*, pp. 3430-3437, Portland, U.S.A., Jun. 2013.
- [6] H. Y. Wu, M. Rubinstein, E. Shih, J. Durand, F. Guttag, and W. Freeman, "Eulerian video magnification for revealing subtle changes in the world," *ACM Trans. Graphics (TOG)*, vol. 31, no. 4, pp. 1-8, Jul. 2012.
- [7] M. Sabokrou, M. Pourreza, X. Li, M. Fathy, and G. Zhao, "Deep-hr: Fast heart rate estimation from face video under realistic conditions," *arXiv preprint arXiv:2002.04821*, 2020.
- [8] F. Bousefsaf, A. Pruski, and C. Maaoui, "3D convolutional neural networks for remote pulse rate measurement and mapping from facial video," *Applied Sci.*, vol. 9, no. 20, pp. 4364, Oct. 2019.
- [9] H. Park, D.-H. Kim, and S. Chang, "Research trend analysis on smart city based on structural topic modeling(STM)," *J. Digital Contents Soc.*, vol. 20, no. 9, pp. 1839-1846, Sep. 2019.
- [10] W. J. Lee, "A study on the R&D trends and implications of smart city in the U.S.," *LAND LAW Rev.*, vol. 34, no. 2, pp. 257-291, Dec. 2018.
- [11] J. A. Han and W. S. Na, "A study on the smart healthcare health management system," *J. Convergence for Inf. Technol.*, vol. 10, no. 6, pp. 8-13, Jun. 2020.
- [12] Y. Kim, "IoT-based smart healthcare technology and market trends," *Summer Annu. Conf. IEIE*, pp. 1232-1234, Jeju Island, Korea, Jun. 2019.
- [13] Y. Lee, H.-J. Kim, G. B. Kim, and Namkug Kim, "Deep learning-based feature extraction for medical image analysis," *Korean Soc. Imaging Informatics in Med.*, vol. 20, no. 1, pp. 1-12, 2014.
- [14] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proc. IEEE Conf. CVPR*, pp. 770-778, Las Vegas, U.S.A., Jun. 2016.
- [15] J. Hu, L. Shen, and G. Sun, "Squeeze-and-excitation networks," in *Proc. IEEE Conf. CVPR*, pp. 7132-7141, Salt Lake City, U.S.A., Jun. 2018.
- [16] S. J. Pan and Q. Yang, "A survey on transfer learning," in *IEEE Trans. Knowledge and Data Eng.*, vol. 22, no. 10, pp. 1345-1359, Oct. 2010.
- [17] K. M. Lee and C. H. Lin, "Video stabilization algorithm of shaking image using deep learning," *J. The Inst. Internet, Broadcasting and Commun.*, vol. 19, no. 1, pp. 145-152, Feb. 2019.
- [18] K. Zhang, Z. Zhang, Z. Li, and Y. Qiao, "Joint face detection and alignment using multitask cascaded convolutional networks," in *IEEE Signal Process. Lett.*, vol. 23, no. 10, pp. 1499-1503, Oct. 2016.

- [19] F. Shi, S. Tsai, Y. Wang, and C. Liang, "Steadiface: Real-time face-centric stabilization on mobile phones," *IEEE ICIP*, pp. 4599-4603, Taipei, Taiwan, Sep. 2019.
- [20] R. Girshick, "Fast R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 1440-1448, Boston, U.S.A., Jun. 2015.
- [21] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards real-time object detection with region proposal networks," in *Advances in NIPS*, pp. 91-99, Montreal, Canada, Dec. 2015.
- [22] J. Long, E. Shelhamer, and T. Darrell, "Fully convolutional networks for semantic segmentation," in *Proc. IEEE Conf. CVPR*, pp. 3431-3440, Boston, U.S.A., Jun. 2015.
- [23] K. He, G. Gkioxari, P. Dollár, and R. Girshick, "Mask R-CNN," in *Proc. IEEE Int. Conf. Comput. Vision*, pp. 2961-2969, Venice, Italy, Mar. 2017.
- [24] J. Snoek, H. Larochelle, and R. P. Adams, "Practical bayesian optimization of machine learning algorithms," *Advances in NIPS*, vol. 25, pp. 2951-2959, Lake Tahoe, U.S.A., Dec. 2012.
- [25] L. Li, K. Jamieson, G. DeSalvo, A. Rostamizadeh, and A. Talwalkar, "Hyperband: A novel bandit-based approach to hyperparameter optimization," *J. Mach. Learning Res*, vol. 18, no. 1, pp. 6765-6816, Jan. 2017.
- [26] S. Falkner, A. Klein, and F. Hutter, "BOHB: Robust and efficient hyperparameter optimization at scale," *arXiv preprint*, arXiv:1807.01774, 2018.
- [27] M. Soleymani, J. Lichtenauer, T. Pun and M. Pantic, "A Multimodal Database for Affect Recognition and Implicit Tagging," in *IEEE Transactions on Affective Computing*, vol. 3, no. 1, pp. 42-55, Jan.-Mar. 2012.
- [28] R. Špetlík, V. Franc, and J. Matas, "Visual heart rate estimation with convolutional neural network," in *Proc. British Mach. Vision Conf.*, Sep. 2018.
- [29] G. Heusch, A. Anjos, and S. Marcel, "A reproducible study on remote heart rate measurement," *arXiv preprint*, arXiv:1709.00962, 2017.
- [30] R. Stricker, S. Müller, and H. Gross, "Non-contact video-based pulse rate measurement on a mobile service robot," *The 23rd IEEE Int. Symp. Robot and Human Interactive Commun.*, pp. 1056-1062, Edinburgh, Scotland, Aug. 2014.
- [31] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, and J. Dubois, "Unsupervised skin tissue segmentation for remote photoplethysmography," *Pattern Recognition Lett.*, vol. 124, pp. 82-90, Oct. 2017.
- [32] K. Simonyan and A. Zisserman, "Very deep convolutional networks for large-scale image recognition," *arXiv preprint*, arXiv:1409.1556, 2014.

김 광 호 (Gwangho Kim)



2020년~현재 : 광운대학교 소프트웨어학부 학사과정
<관심분야> 머신러닝, 하이퍼 파라미터 최적화

Yuli Sun Hariyani



2010년 : B.S. degree in telecommunication engineering from Telkom University
2013년 : M.S. degree in electrical engineering from Telkom University
2014년~현재 : Ph.D. degree with Computer Engineering Department, Kwangwoon University and Lecturer with Telkom University
<관심분야> pattern recognition, medical image processing and biomedical signal processing.

한 승 우 (Seungwoo Han)



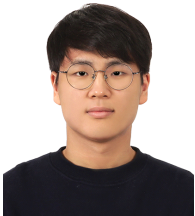
2019년 : 삼육대학교 카메카트
로닉스학과 학사
2019년~현재 : 광운대학교 지능
정보시스템 임베디드SW공
학과 석사과정
<관심분야> 자율주행, 머신러
닝, 강화학습

손 량 희 (Ryanghee Sohn)



2005년 2월 : 연세대학교 의용
전자공학 학사
2010년 8월 : 연세대학교 의공
학 석사
2015년 3월~현재 : 연세대학교
의공학 박사과정
2017년 9월~현재 : (주)엠마헬
스케어 대표이사
<관심분야> 의료공학, 디지털 헬스케어, 인공지능

이 현 빈 (HyeonBin Lee)



2020년~현재 : 광운대학교 소프
트웨어학부 학사과정
<관심분야> 머신러닝, 보안

박 철 수 (Cheolsoo Park)



2004년 : 서강대학교 전자공학
과 학사
2006년 : 서울대학교 협동과정
의용생체공학과 석사
2012년 : 영국 임페리얼 칼리지
런던 박사
2012년~현재 : 광운대학교 컴퓨
터정보공학부 교수
<관심분야> 인공지능, 머신러닝, 생체신호처리