

논문 2023-18-07

얼굴 영상 기반의 심박수 추정을 위한 딥러닝 모델의 경량화 기법 (Lightweight Deep Learning Model for Heart Rate Estimation from Facial Videos)

황 규 태, 박 명 근, 이 상 준*

(Gyutae Hwang, Myeonggeun Park, Sang Jun Lee)

Abstract : This paper proposes a deep learning method for estimating the heart rate from facial videos. Our proposed method estimates remote photoplethysmography (rPPG) signals to predict the heart rate. Although there have been proposed several methods for estimating rPPG signals, most previous methods can not be utilized in low-power single board computers due to their computational complexity. To address this problem, we construct a lightweight student model and employ a knowledge distillation technique to reduce the performance degradation of a deeper network model. The teacher model consists of 795k parameters, whereas the student model only contains 24k parameters, and therefore, the inference time was reduced with the factor of 10. By distilling the knowledge of the intermediate feature maps of the teacher model, we improved the accuracy of the student model for estimating the heart rate. Experiments were conducted on the UBFC-rPPG dataset to demonstrate the effectiveness of the proposed method. Moreover, we collected our own dataset to verify the accuracy and processing time of the proposed method on a real-world dataset. Experimental results on a NVIDIA Jetson Nano board demonstrate that our proposed method can infer the heart rate in real time with the mean absolute error of 2.5183 bpm.

Keywords : Digital health care, Deep learning, Model compression, Remote photoplethysmography, Heart rate estimation

1. 서 론

인구 고령화와 건강에 대한 관심 증대로 디지털 헬스케어 분야에 대한 관심이 커지는 추세이다. 스마트폰과 스마트워치 등의 엣지 디바이스를 통하여 사용자의 행동 패턴 및 생체신호를 기록하고, 이를 기반으로 다양한 헬스케어 관련 서비스가 연구개발 되고 있다. 특히, 생체신호는 심박수, 심박변이도, 혈압, 산소포화도 등의 임상적으로 의미 있는 생체정보를 측정하는 데에 활용할 수 있다. 또한, 이와 같은 생체정보는 사용자의 활력도나 스트레스 지수와 같은 사용자 상태 정보를 추정할 수 있다는 면에서 활용도가 높다. 대표적 생체신호인 PPG 신호는 혈류에 대한 흡수도가 높은 광원을 피부 표면에 투영하고 반사되어 돌아오는 광량을 측정하는 원리를 기반으로 혈류량의 변화를 측정한다. 하지만, PPG 센서는 사용자의 피부 표면과 접촉해야만 생체신호 측정이 가능하며, 활용가능한 분야 및 기기가 한정적이라는 한계를 가진다.

비접촉 방식으로 생체신호를 추출하기 위하여 카메라 영상정보를 활용하는 연구가 다수 시도되었으며, 이와 같은 생체신호를 rPPG (remote PPG)라고 한다. PPG 신호와 유

사한 생체신호를 비접촉 방식으로 획득 가능하다는 점이 rPPG 측정 기술의 특징이다. 기존의 rPPG 추정 방법들은 입력 영상 시퀀스에서 얼굴과 같은 피부 영역을 찾아내는 단계와 피부 영역의 밝기변화를 정제하여 rPPG 신호를 추출하는 단계로 구성된다. 피부 영역의 밝기변화를 기반으로 생체신호를 추정하는 기존 방법들은 수학적인 알고리즘을 기반으로 하는 방법과 딥러닝 모델을 활용한 학습기반 방법으로 분류할 수 있다.

영상 생체신호 연구의 초기에는 피부 밝기에 대한 수학적 모델과 주파수 분석을 이용한 알고리즘이 다수 제안되었다. Poh et al.는 얼굴 영역을 검출하고 얼굴 영역의 RGB 채널 밝기에 대한 독립 성분 분석을 적용하여 생체신호 정보를 추출하는 방법을 제안하였다 [1]. 얼굴 영역에서 추출한 생체신호에서 0.7 Hz에서 4 Hz까지의 신호를 통과시키는 bandpass filtering 적용함으로써 잡음에 강건한 알고리즘을 구성하였다. Lewandowska et al.는 rPPG 신호를 얼굴 전체 영역과 이마 영역에서 추출하고, 다양한 조합의 채널에서 획득한 신호의 성질을 분석함으로써 개선된 성능의 rPPG 추정 방법을 제안하였다 [2]. Haan et al.의 연구에서는 피부 표면에서 정 반사된 빛의 성분과 혈관까지 투영된 빛의 성분 차이를 분석하여 움직임이 있는 영상에서도 기존 방법대비 강건한 알고리즘을 제안하였다 [3]. Wang et al.의 연구는 rPPG 추정을 위하여 이전에 발표된 수학적인 모델의 피부 반사의 광학 및 생리학적 특성을 통합했다는 면에서 의미가 있다 [4].

최근 시계열 데이터 분석을 위한 딥러닝 모델의 발달로 rPPG 추정을 위한 다양한 딥러닝 기법이 제안되었다. Chen

*Corresponding Author (sj.lee@jbnu.ac.kr)

Received: Jan. 9, 2023, Revised: Feb. 8, 2023, Accepted: Mar. 2, 2023.

G. Hwang: Jeonbuk National University. (M.S.)

M. Park: Polaris3D Research (Researcher)

S. J. Lee: Jeonbuk National University (Asst. Prof.)

* This work was supported by the National Research Foundation of Korea and funded by a grant from the Korean government (No. 2021R1G1A1009792).

et al.의 연구에서는 인접한 두 프레임 영상의 차이를 통하여 PPG 신호의 변화량을 추정함과 동시에 피부영역에 가중치를 높이는 공간 어텐션 기반의 CAN 알고리즘을 제안하였다 [5]. Yu et al.은 3D 컨볼루션 기반의 딥뉴럴네트워크 구조를 제안하고 Pearson correlation을 기반으로 딥러닝 모델을 학습시키는 방법을 제안하였다 [6]. Nowara et al.은 CAN 모델의 공간 어텐션 정보를 역으로 활용하여 피부와 관련 없는 영역에서의 노이즈 신호를 추정하고 LSTM을 기반으로 하는 노이즈 처리 방법을 제안하였다 [7]. Yu et al.의 후속연구로, 3D 컨볼루션을 통한 지역적인 특징과 비디오의 시간적 차이에 기반한 트랜스포머의 전역적인 특징 추출방법을 이용한 심박수 예측 모델이 제안되었다 [8].

본 논문에서는 500 GFLOPs 미만의 연산 성능을 가진 저전력 임베디드 시스템에서 얼굴 영상으로부터 rPPG와 심박수 (HR, heart rate)를 추정하기 위한 딥러닝 기법을 제안한다. 기존의 딥러닝 방법들은 입력 영상의 spatio-temporal 특징을 분석하고 생체정보를 추정하는 데에는 효과적이거나, 계산복잡도 면에서는 임베디드 시스템에서 활용되기 어렵다는 한계가 있다. 이러한 한계를 극복하고자 지식증류 (knowledge distillation) 기법을 적용하여 rPPG 추정에 효과적인 기존의 딥러닝 모델을 경량화하였다. 딥러닝 모델의 경량화 과정에서는 다른 크기의 수용영역 (receptive field)에서 추출된 공간 어텐션 정보를 distillation 하는 손실함수를 적용하여, 모델 복잡도를 줄이는 동시에 성능 하락을 최소화 하였다. 공개 데이터셋인 UBFC-rPPG [9]를 이용하여 제안하는 알고리즘을 구현하였으며, 실제 데이터셋을 수집하여 젯슨나노 임베디드 보드에서 정확도 및 처리속도 측면의 성능을 검증하였다.

II. 본 론

1. 데이터셋 구성

영상으로부터 rPPG 신호를 추정하는 딥러닝 모델을 개발하고 평가하기 위하여 공개 데이터셋 UBFC-rPPG를 사용하였다. UBFC-rPPG는 피실험자들의 얼굴을 촬영한 영상과 손가락 말단에서 측정된 PPG 신호의 데이터로 구성된다. 이 공개 데이터셋은 43명의 피실험자들로부터 획득되었으며, 각 동영상 데이터는 640 x 480 크기 및 30 FPS의 프레임 속도로 2분 동안 촬영되었다. 데이터 수집 과정에서 피실험자들은 카메라로부터 1m 가량 떨어진 거리에서 촬영되었으며, PPG 생체신호는 영상과 동기화되어 기록되었다. 그림 1은 UBFC-rPPG 데이터셋에 있는 피실험자들의 영상 예시를 보여준다.

공개 데이터셋에서 개발한 딥러닝 모델의 심박수 추정 정확도를 한국형 데이터셋에 검증하기 위하여 실제 데이터를 수집하고 알고리즘의 성능을 평가하였다. 그림 2는 동기화된 영상-PPG 데이터 수집을 위한 젯슨나노, 아두이노, CSI

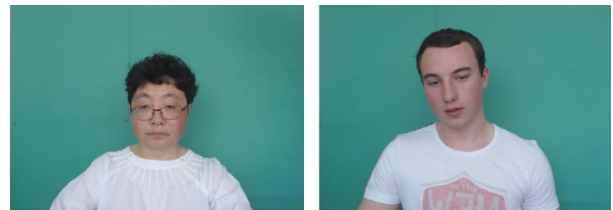


그림 1. UBFC-rPPG 데이터 영상 예시
Fig. 1. Example image of UBFC-rPPG dataset

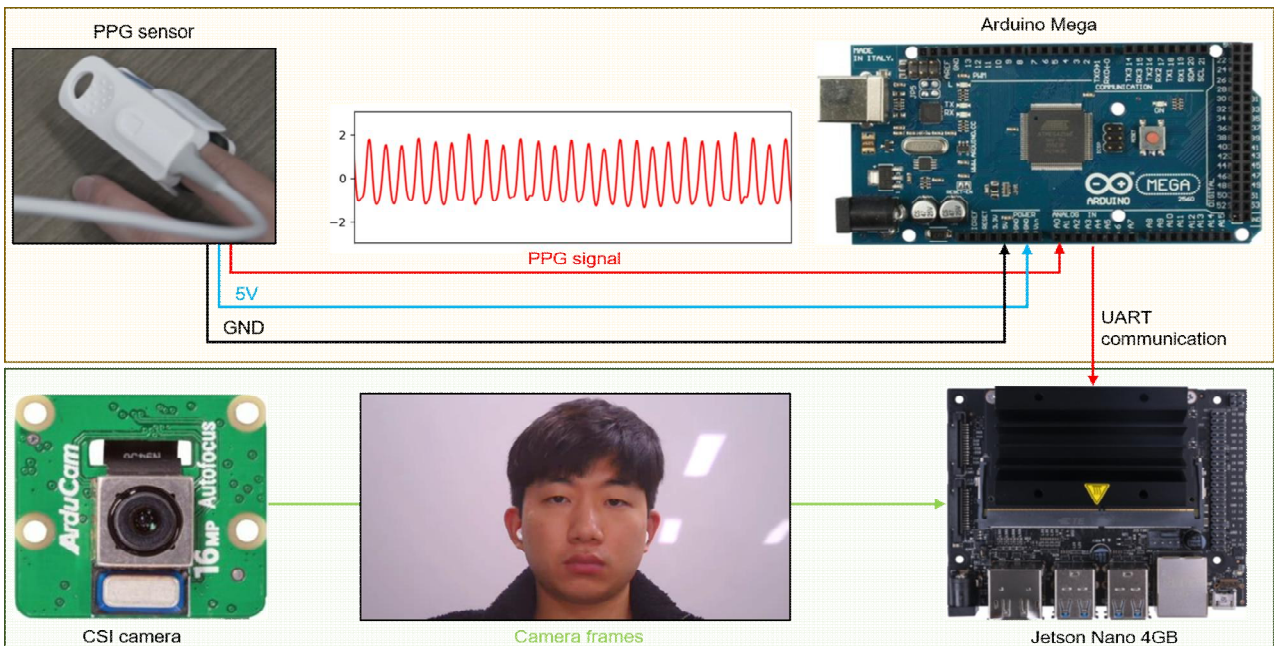


그림 2. 영상-PPG 데이터 수집 시스템 구성도
Fig. 2. Configuration diagram of video-PPG data acquisition system

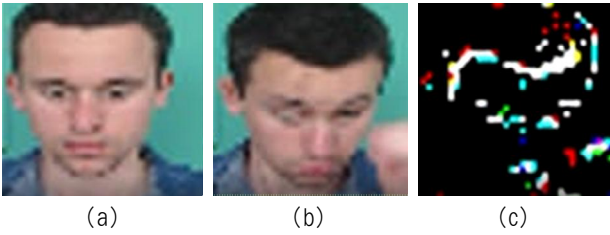


그림 3. 전처리 영상 시각화 예시
Fig. 3. Visualization of pre-processed image

카메라, PPG센서로 이루어진 시스템 구성도와 수집된 실제 데이터의 예시를 보여준다. 젯슨나노에 CSI 카메라를 연결하여 피실험자의 얼굴을 640 x 360의 크기로 촬영하였으며, 동시에 손가락 말단에서 PPG 센서로 피실험자의 맥파를 측정하였다. PPG 센서와 연결된 아두이노는 UART 시리얼 통신을 통하여 젯슨나노에 PPG 신호를 전송하였으며, 이는 25 Hz로 샘플링 후 영상 신호와 동기화되어 기록되었다. 영상 촬영 환경은 하얀색의 배경과 정면에 백색광을 비추어 얼굴에 드리워지는 그림자를 최소화하였다. 실제 데이터를 구성하기 위하여 한국 성인 남성 4명으로부터 약 12분 길이의 동영상 35개를 획득하였으며, 긴 시간의 동영상 촬영을 통하여 졸음으로 인한 심박수의 감소를 유도하였다. 학습

모델의 일반화 정도를 평가하기 위하여 획득된 실제 데이터셋을 피실험자가 겹치지 않도록 학습 및 테스트 셋으로 분할하였다. 실제 데이터셋 중 3명으로부터 수집된 30개의 동영상 데이터는 딥러닝 모델의 학습에 활용하였으며, 나머지 1명으로부터 획득된 5개의 동영상 데이터에 대하여 알고리즘 성능을 평가하였다.

2. 데이터 전처리 과정

딥러닝 모델의 효과적인 심박수 예측을 위한 전처리 과정으로서 얼굴영역 검출과 PPG 신호의 노이즈를 제거하기 위한 필터링이 수행되었다. 얼굴 영역의 검출을 위하여 MTCNN 모델이 사용되었으며 [10], 영상의 첫 번째 프레임에서 검출된 바운딩 박스를 포함하는 1.6배 크기의 영역을 이후 프레임에서도 사용하였다. 검출된 얼굴 영역을 36 x 36으로 크기를 변환하였으며, 0과 1사이의 값으로 normalize된 appearance data와 인접 프레임 사이의 차이를 계산한 motion data로 변환하였다. PPG 신호의 노이즈를 제거하고, 사람의 심박수 정보가 포함되어있는 주파수 대역 정보를 활용하기 위하여, 0.7 Hz부터 4.0 Hz까지 범위의 bandpass filtering을 수행하였다. 그림 3의 (a)는 전처리 이후 영상의 밝기 정보를 포함하는 t시점의 appearance data이고, (b)는 t+1시점의 영상이며, (c)는 인접 프레임 사이에서 계산한

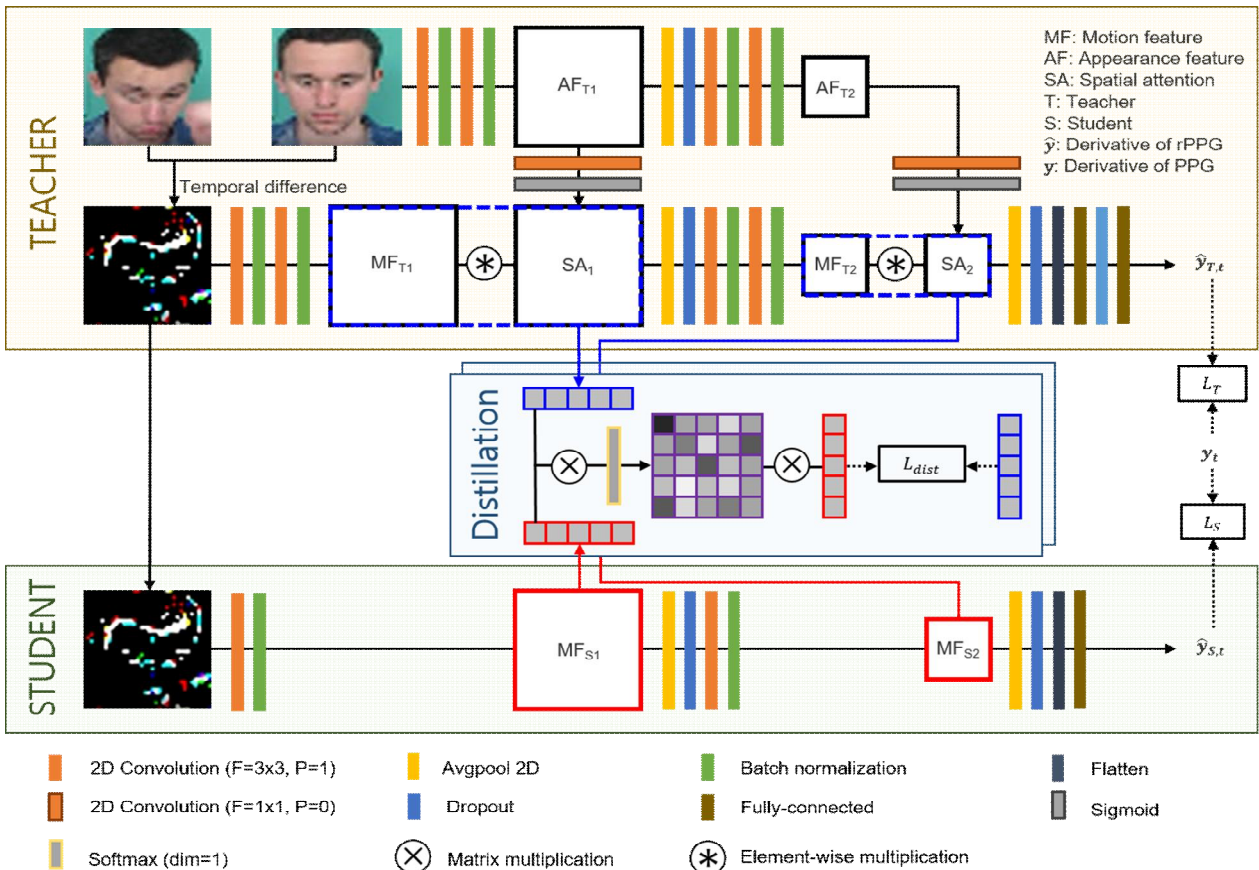


그림 4. Teacher 모델과 Student 모델 구조 및 경량화 과정
Fig. 4. Pipeline of Knowledge distillation with structure of teacher and student model

motion data의 시각화 예시를 보여준다. 그림 3 (c)에서 변하지 않는 배경픽셀의 경우 검은색으로 표현되었으며, 각 채널에서의 밝기변화는 움직임이 발생한 영역을 의미한다.

3. 심박수 추정을 위한 딥러닝 모델 구조 및 학습 방법

영상 신호로부터 rPPG 정보를 추론하기 위하여 컨볼루션과 어텐션 모듈 기반의 teacher와 student 모델을 구성하였다. 그림 4는 임베디드 시스템에서 실시간으로 심박수를 추정하기 위한 knowledge distillation 파이프라인을 보여준다. Teacher 모델은 기존 연구인 CAN [5]의 네트워크 구조를 기반으로 하고 있으며, motion data와 appearance data를 별도로 처리하기 위한 두 개의 모델로 구성된다. Motion data를 처리하는 모델에서는 2D 컨볼루션과 batch normalization을 활용하여 프레임 사이에 발생하는 모션 노이즈와 피부의 혈색 변화 정보를 추출한다. Appearance data를 처리하는 모델에서는 특징맵 AF_T 로부터 피부 영역에서 높은 가중치 값을 가지고, 0과 1 사이의 값들로 이루어진 공간 어텐션 맵 SA를 계산한다. AF_{T1} 과 AF_{T2} 는 각각 appearance 모델의 중간단 및 말단의 특징맵을 의미하며, SA_1 과 SA_2 는 이로부터 계산된 어텐션 맵을 의미한다. Motion 모델 특징맵 MF_{T1} 과 MF_{T2} 는 동일한 층에서 계산된 SA_1 과 SA_2 와의 요소 별 곱셈을 통하여 혈색 변화 정보가 있는 공간적 영역에 집중한다. Motion 모델의 마지막 특징맵은 2차원 벡터로 형태가 변환되어 2개의 fully-connected layer를 통하여 인접 프레임 영상 사이의 PPG 변화량을 예측한다. 그림 4에서 $\hat{y}_{T,t}$ 는 t 시점과 t+1 시점의 인접 영상을 입력받았을 때 teacher 모델에 의해서 추론된 t 시점에서의 PPG 변화량을 의미한다. Teacher 모델을 학습하기 위한 손실함수 L_T 는 아래와 같은 평균제곱오차에 기반한 수학적 식으로 정의된다.

$$L_T = \frac{1}{N} \sum_t^N (\hat{y}_{T,t} - y_t)^2. \quad (1)$$

위 수학적식에서 N 는 총 인접영상의 수를 의미하며, y_t 는 t 시점의 ground truth PPG에 대한 변화량을 나타낸다.

저전력 임베디드 시스템에서 실시간으로 심박수를 추정하기 위한 딥러닝 모델의 경량화 방법을 적용하였다. 전처리 과정 및 teacher 모델의 파라미터 수와 연산량을 고려하였을 때, 임베디드 보드에서 PPG 변화량의 실시간적인 예측을 기대하기 어렵다. 따라서 임베디드 보드에서 teacher 모델의 최적화를 위한 knowledge distillation 기반의 경량화를 수행하였다. 심박수 추정 딥러닝 모델은 rPPG 변화량 값을 출력하므로, Hinton et al. [11]이 제안한 출력단의 불확실성을 이용한 경량화에는 어려움이 있어 특징맵 기반의 경량화 기법을 활용한다. 본 논문에서 사용한 TaT (Target-aware Transformer)는 컨볼루션 기반의 딥러닝 모델을 경량화하는 경우에 teacher-student 사이의 수용영역 불일치를 해소하기 위하여 제안되었다 [12]. Student 모델은 motion data만을 입력으로 2개의 컨볼루션 층과 1개의 fully-connected layer

로 구성된 간단한 모델이다. MF_{S1} 과 MF_{S2} 는 경량화를 통하여 teacher 모델의 중간 및 말단의 특징맵을 모방하게 되며, 이를 통해 student 모델의 작은 수용영역과 MF_S 의 피부 영역 가중치 정보의 부재를 해소하였다.

rPPG 변화량을 추정하는 딥러닝 모델의 경량화를 위한 손실함수는 teacher-student 모델의 특징맵 임베딩에 기반하여 정의된다.

$$\tilde{E}_S = \text{Softmax}(E_S \times E_T^\top) \times E_S, \quad (2)$$

$$L_{dist} = \|\tilde{E}_S - E_T\|_2. \quad (3)$$

수학적식 (2)에서 E_T 와 E_S 는 teacher-student 모델 특징맵의 2차원 임베딩 벡터이며, \top 는 전치행렬을 의미한다. 손실함수로서, E_T 와 E_S 의 관계성을 가중치화하여 계산된 \tilde{E}_S 는 수학적식 (3)에 기반하여 E_T 와의 유클리디안 거리가 계산된다. Student 모델이 경량화를 위하여 참고한 teacher 모델의 특징맵은 공간 어텐션 가중치를 포함하여 중간 및 말단 층에서 출력되며, 학습 과정에서 L_{dist1} 과 L_{dist2} 를 통하여 E_S 가 E_T 를 모방하도록 한다. Student 모델을 학습하기 위한 손실함수 L_S 는 $\hat{y}_{S,t}$ 와 y_t 의 수학적식 (1)과 같은 평균제곱오차에 기반하여 정의되고, 경량화를 위한 전체 손실함수 L_{total} 은 아래 수식과 같이 L_{dist1} , L_{dist2} 와 L_S 의 합으로 정의된다.

$$L_{total} = L_{dist1} + L_{dist2} + L_S. \quad (4)$$

본 논문에서 제안하는 딥러닝 모델은 인접 시점 샘플의 rPPG 변화량만을 추정하므로, 심박수에 대하여 평가를 위한 후처리 방법이 필요하다. 딥러닝 모델이 순차적으로 인접 시점의 영상을 처리하고, 출력하는 변화량의 누적합으로 비디오 영상에 대한 rPPG 신호를 계산한다. 이후 노이즈 처리를 위한 PPG 신호의 전처리 방법과 같이 0.7 Hz부터 4.0 Hz까지 범위의 bandpass filtering을 수행하였다. 주기성이 있는 심박수의 계산을 위해 FFT (Fast Fourier Transform)에 기반한 주파수 분석으로 파워스펙트럼이 계산되었다. 본 논문에서는 필터링이 수행된 rPPG 및 PPG 신호를 10초의 슬라이딩 윈도우에서 파워스펙트럼을 계산하고, 피크에 해당하는 주파수의 60배를 슬라이딩 윈도우의 심박수로 정의하였다. 전체 신호의 심박수는 각 슬라이딩 윈도우 심박수의 평균치이다.

4. 실험 결과

본 논문에서 제안하는 알고리즘의 구현과 성능평가를 위하여 Intel core i9-10900K CPU, NVIDIA RTX 3090, 64GB DDR4 RAM 등으로 이루어진 하드웨어 환경에서 실험하였다. 저전력 임베디드 시스템은 Quad-core ARM A57 CPU, 128-core Maxwell GPU, 4GB 64-bit RAM 등으로 이루어진 NVIDIA Jetson Nano 4GB를 활용하였다. 알고리즘의 구현에는 파이썬과 Pytorch 소프트웨어를 사용하였다. 알고리즘의 정확도를 평가하는 과정에서는 알고리즘에 의한 심박

표 1. UBFC-rPPG 데이터 셋에서의 비교실험 결과
Table 1. Performance comparison on the UBFC-rPPG dataset

	MAE	RMSE
Meta-rPPG [13]	5.97	7.42
SynRhythm [14]	5.59	6.82
3D CNN [15]	5.45	8.64
Chrominance-rPPG [3]	4.70	6.61
AHRE [16]	4.20	5.78
PhysNet [6]	3.63	5.29
PulseGAN [17]	1.19	2.10
rPPGNet [18]	0.72	1.47
Proposed method	3.4444	7.2979

수 추론 값과 실제 심박수 값 사이의 절대평균오차 (MAE, mean absolute error)와 평균제곱근오차 (RMSE, root mean squared error)를 평가지표로 사용하였다.

제안하는 알고리즘의 성능을 기존알고리즘의 성능과 비교하기 위하여 공개데이터셋인 UBFC-rPPG에서 실험하였다. 표 1은 제안하는 알고리즘을 최신 알고리즘들과 비교실험한 결과를 보여준다. 실험 과정에서는 기존 알고리즘들과 동일한 학습 및 검증 방법을 사용하였다. 학습 및 검증 데이터셋에 동일한 피실험자의 정보가 포함되지 않도록 피실험자를 기준으로 5개의 그룹으로 구성하여 교차검증의 방법으로 알고리즘을 학습 및 검증하였으며, 5번의 검증 결과의 평균값을 알고리즘의 정확도로 정의하였다. 실험 결과 최근에 제안된 Meta-rPPG [13] 방법보다 MAE와 RMSE 적인 측면에서 제안하는 알고리즘이 우수한 성능을 보이는 것을 확인할 수 있다. 다른 알고리즘들인 SynRhythm [14], 3D CNN [15], Chrominance-rPPG [3], AHRE [16], Physnet [6]의 방법들과 비교해보면 RMSE 측면에서는 성능이 조금 떨어졌지만 MAE 측면에서는 더 높은 정확도를 보였다. 이는 제안하는 알고리즘이 일부 이상 추론치 값을 제외하고는 실제 값에 가까운 심박수를 추론했음을 의미한다. 또 다른 최신 기법들인 PulseGAN [17]과 rPPGNet [18]의 기법들은 제안하는 알고리즘 보다 정확도적인 면에서 좋은 성능을 보였지만, 딥러닝 모델이 필요로 하는 연산량으로 인하여 임베디드 보드에서 영상정보를 실시간으로 처리하기 어렵다는 한계를 가지고 있다.

제안하는 knowledge distillation에 기반한 학습 파이프라인의 효과를 보이기 위하여, teacher와 student 모델의 비교 실험을 수행하였다. 표 2는 UBFC-rPPG 데이터셋에서 teacher와 student 모델의 MAE와 RMSE를 보여준다. Teacher 모델의 경우 MAE와 RMSE의 두 평가지표 측면에서 student 모델보다 월등히 높은 정확도를 보이지만, 본 논문에서 대상으로 하는 임베디드 보드에서 실시간 추론이 불가능하다는 한계가 있다. 본 논문에서 제안하는 knowledge distillation 기법을 student 모델에 적용함으로써, 처리시간의 실시간성을 확보하는 동시에 student 모델의 정확도가

표 2. Teacher와 student 모델의 성능 비교 실험
Table 2. Performance comparison of the teacher and student models (KD: knowledge distillation)

	MAE	RMSE
Teacher	0.9563	1.5795
Student (without KD)	4.5589	9.6856
Student + KD	3.4444	7.2979

표 3. 실제 데이터셋에 대한 심박수 추정 성능 검증
Table 3. Evaluation of the proposed method on the real-world dataset (KD: knowledge distillation)

	MAE	RMSE
Teacher	2.1803	5.9514
Student (without KD)	2.8000	6.7063
Student + KD	2.5183	6.2080

표 4. 젯슨나노 임베디드 보드에서의 처리시간 비교
Table 4. Processing time of the teacher and student models in Jetson Nano embedded board

	Teacher	Student
Number of parameters	795,811	24,769
Inference time (sec/frame)	0.50	0.05

떨어지는 문제를 어느 정도 보완할 수 있는 것을 실험적으로 확인하였다.

한국형 영상-PPG 신호 데이터로 구성된 수집 데이터셋에서도 제안하는 학습 파이프라인의 효과를 보이기 위하여 teacher와 student 모델의 비교 실험을 수행하였다. 표 3은 수집된 실제 데이터셋에서 teacher와 student 모델의 MAE와 RMSE를 보여준다. Teacher 모델은 경량화를 수행한 student 모델에 비하여 두 가지 평가지표에서 좋은 성능을 보이지만, 성능의 차이에 비하여 더 많은 연산량을 필요로 한다는 점에서 임베디드 보드에서의 적용에 부적합하다는 면이 있다. 경량화를 수행한 student 모델은 그렇지 않은 모델과 비교하여 두 가지 평가지표에서 높은 정확도를 보이며, 이를 통하여 teacher 모델의 경량화를 통한 student 모델의 학습이 의미가 있다는 것을 알 수 있다.

젯슨나노 임베디드 보드에서의 teacher와 student 모델의 처리시간 비교를 통하여 경량화 기법의 효과를 검증하였다. 표 4는 임베디드 보드에서 심박수 추정에 필요한 딥러닝 모델 파라미터 수 및 전처리를 포함한 전체 파이프라인의 추론 시간을 보여준다. Teacher 모델과 student 모델의 파라미터 수는 약 32.12배의 차이를 보이며, 추론 성능의 차이와 비교했을 때 경량화를 통한 student 모델의 효율성을 입증한다. 전체 파이프라인의 처리시간을 비교한 경우에서도 student 모델이 teacher 모델보다 약 10배 빠른 속도를 보이는 것을 확인할 수 있다. 임베디드 보드에서 Teacher 모델은 1초에 약 2샘플의 rPPG 변화량 추론이 가능한 반면,

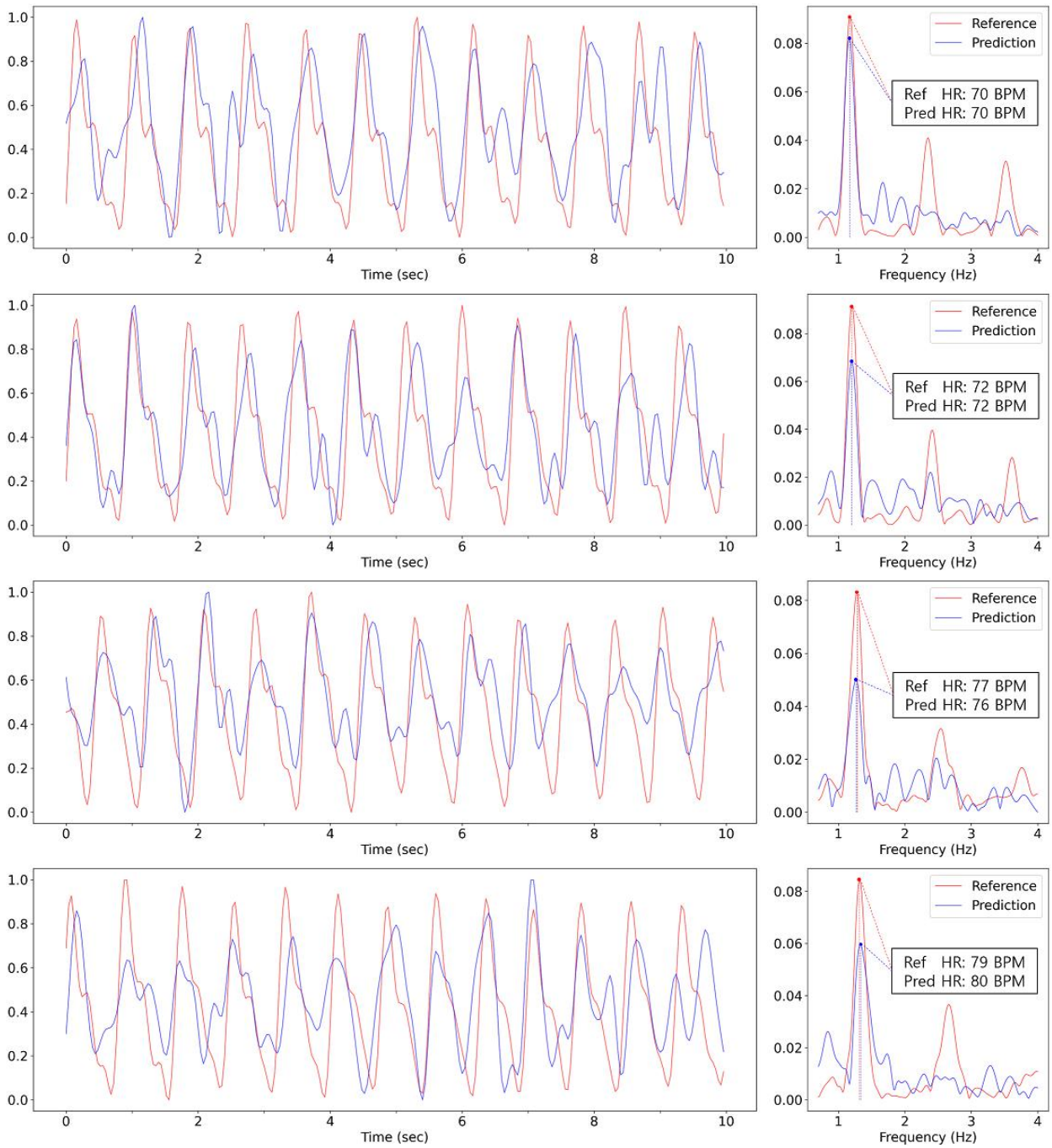


그림 5. 딥러닝 모델이 추론한 rPPG 및 이에 대한 파워스펙트럼 (심박수 추정 MAE가 1 bpm 이하인 경우)
 Fig. 5. Predicted rPPG and its power spectrum with MAE below 1 bpm

student 모델은 약 20샘플의 rPPG 변화량 추론이 가능하여 저전력 임베디드 보드에서도 실시간 심박수 추정이 가능하다. 수집 데이터셋의 테스트 데이터에 대한 ground truth 및 student 모델의 예측결과를 시각화하였다. 그림 5는 심박수 MAE가 1 bpm 이하인 결과 그래프를 나타내며, 빨간색은 ground truth 신호를, 파란색은 예측 신호를 의미한다. 왼쪽의 신호 비교 결과에서 두 신호의 피크가 유사한 시간대에 위치 해있는 것을 볼 수 있다. 오른쪽 파워스펙트럼을 통한 주파수 분석 결과 그래프에서는 x축이 신호의 주파수를 의

미하며, 두 신호의 최대 파워에 해당하는 주파수가 일치하는 모습을 보인다. 그림 6은 심박수의 MAE가 10 bpm 이상인 결과 그래프를 나타낸다. 그림 6 왼쪽의 신호 비교 결과를 살펴보면, 부분적으로 피크가 일치하나, 일부 노이즈가 포함되어있는 것을 확인할 수 있다. 이에 따라 오른쪽의 파워스펙트럼에서도 최대 파워에 해당하는 주파수에 차이가 나타난다. 실험 결과를 통하여 10 bpm 이상 오차가 발생한 경우의 노이즈 패턴을 확인할 수 있었으며, 향후 연구에서 이러한 패턴을 분석하고 개선할 예정이다.

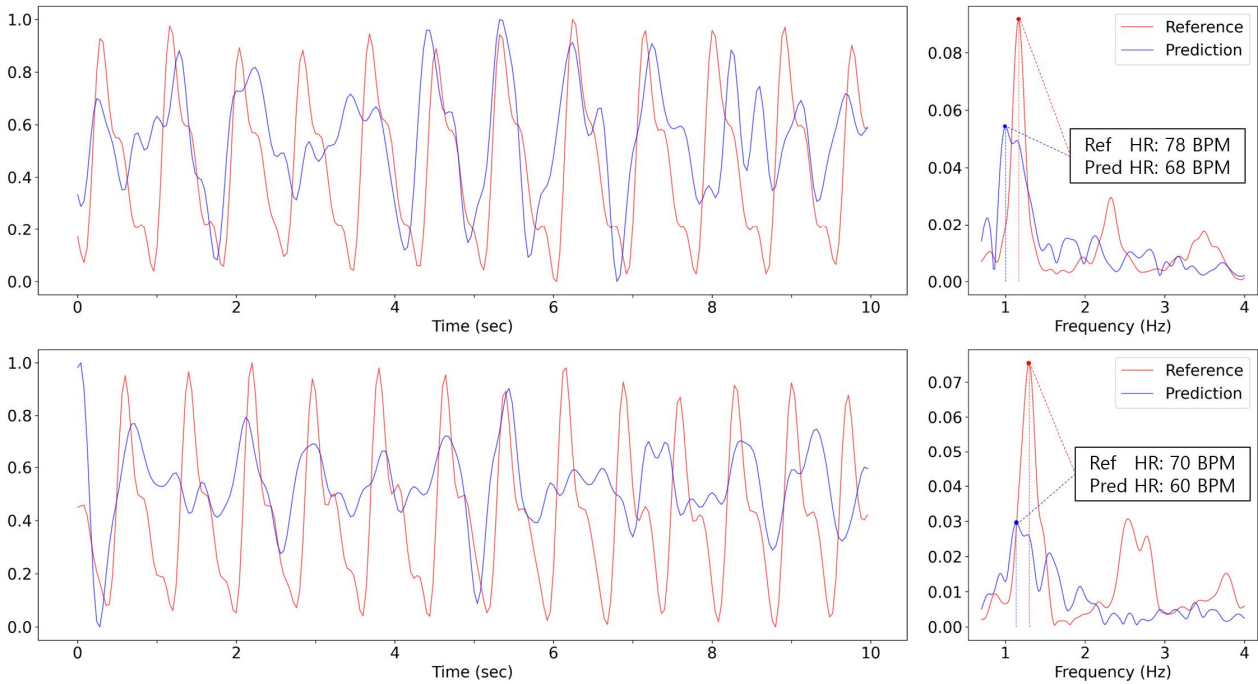


그림 6. 딥러닝 모델이 추론한 rPPG 및 이에 대한 파워스펙트럼 (심박수 추정 MAE가 10 bpm 이상인 경우)
 Fig. 6. Predicted rPPG and its power spectrum with MAE over 10 bpm

III. 결론

본 논문에서는 딥러닝 모델의 경량화를 통한 임베디드 보드에서의 영상 기반 rPPG 신호의 추정방법을 제안한다. 기존의 영상 기반 심박수 추정 모델은 성능이 높지만, 계산 복잡도로 인해 저전력 임베디드 시스템에서 실시간으로 심박수 추정이 불가능하다는 한계가 있다. 이러한 문제를 해결하기 위해서 knowledge distillation 기법으로 teacher 모델의 경량화를 수행하여 임베디드 시스템에서 심박수를 실시간으로 추정할 수 있는 student 모델을 연구개발 하였다. 또한 한국형 데이터에서 모델의 성능을 검증하기 위해 실제 데이터셋을 수집하였으며, 이에 대한 teacher 및 student 모델의 심박수 추정 성능을 보였다. 실험 결과, 본 논문에서 제안하는 student 모델은 젓슨나노 임베디드 보드에서 기존 딥러닝 모델과 유사한 구조의 teacher 모델보다 10배 가량 빠른 속도로 비슷한 정확도의 심박수 추정이 가능함을 확인 하였다. 그러나 피실험자 수의 한계로 인하여 실제 데이터셋의 다양성이 부족하고, teacher 및 student의 성능이 공개 데이터셋에서의 성능과 비교하여 비슷한 양상을 보이나 수치상으로 차이가 있다. 이러한 한계점을 극복하기 위하여 향후 연구에서는 본 연구에서 구축된 영상-PPG 데이터 수집 시스템으로 성별, 나이대 다양성 등을 고려한 실제 데이터셋을 구성할 예정이다. 본 논문에서 제안하는 영상 기반의 심박수 추정 기법은 디지털 헬스케어 분야의 다양한 어플리케이션에서 활용될 수 있을 것이라 기대된다.

References

- [1] M. Poh, D. J. McDuff, R. W. Picard, "Advancements in Noncontact, Multiparameter Physiological Measurements Using a Webcam," *IEEE Transactions on Biomedical Engineering*, Vol. 58, No. 1, pp. 7–11, 2011.
- [2] M. Lewandowska, J. Rumiński, T. Kociejko, J. Nowak, "Measuring Pulse Rate with a Webcam – A Non-contact Method for Evaluating Cardiac Activity," *Proceedings of 2011 Federated Conference on Computer Science and Information Systems*, pp. 405–410, 2011.
- [3] G. de Haan, V. Jeanne, "Robust Pulse Rate From Chrominance-Based rPPG," *IEEE Transactions on Biomedical Engineering*, Vol. 60, No. 10, pp. 2878–2886, 2013.
- [4] W. Wang, A. C. den Brinker, S. Stuijk, G. de Haan, "Algorithmic Principles of Remote PPG," *IEEE Transactions on Biomedical Engineering*, Vol. 64, No. 7, pp. 1479–1491, 2017.
- [5] W. Chen, D. McDuff, "DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks," *Proceedings of the European Conference on Computer Vision*, pp. 349–365, 2018.
- [6] Z. Yu, X. Li, G. Zhao, "Remote Photoplethysmograph Signal Measurement from Facial Videos Using Spatio-temporal Networks," *Proceedings of the British Machine Vision Conference*, 2019.
- [7] E. M. Nowara, D. McDuff, A. Veeraraghavan, "The Benefit of Distraction: Denoising Camera-Based Physiological Measurements Using Inverse Attention," *Proceedings of the IEEE/CVF International Conference*

- on Computer Vision, pp. 4955-4964, 2021.
- [8] Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. S. Torr, G. Zhao, "PhysFormer: Facial Video-Based Physiological Measurement With Temporal Difference Transformer," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 4186-4196, 2022.
- [9] S. Bobbia, R. Macwan, Y. Benezeth, A. Mansouri, J. Dubois, "Unsupervised Skin Tissue Segmentation for Remote Photoplethysmography," Pattern Recognition Letters, Vol. 124, pp. 82-90, 2019.
- [10] K. Zhang, Z. Zhang, Z. Li, Y. Qiao, "Joint Face Detection and Alignment Using Multitask Cascaded Convolutional Networks," IEEE Signal Processing Letters, Vol. 23, No. 10, pp. 1499-1503, 2016.
- [11] G. E. Hinton, O. Vinyals, J. Dean, "Distilling the Knowledge in a Neural Network," ArXiv:1503.02531, 2015.
- [12] S. Lin, H. Xie, B. Wang, K. Yu, X. Chang, X. Liang, G. Wang, "Knowledge Distillation via the Target-Aware Transformer," Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pp. 10915-10924, 2022.
- [13] E. Lee, E. Chen, Lee, C. Y. Lee, "Meta-rPPG: Remote Heart Rate Estimation Using a Transductive Meta-learner," Proceedings of the European Conference on Computer Vision, pp. 392 - 409, 2020.
- [14] X. Niu, H. Han, S. Shan, X. Chen, "SynRhythm: Learning a Deep Heart Rate Estimator from General to Specific," Proceedings of the International Conference on Pattern Recognition, pp. 3580 - 3585, 2018.
- [15] F. Bousefsaf, A. Pruski, C. Maaoui, "3d Convolutional Neural Networks for Remote Pulse Rate Measurement and Mapping from Facial Video," Applied Sciences 9, No. 20: 4364, 2019.
- [16] P. Gupta, B. Bhowmick, A. Pal, "Accurate Heart-rate Estimation from Face Videos Using Qualitybased Fusion," Proceedings of the 2017 International Conference on Image Processing, pp. 4132 - 4136, 2017.
- [17] R. Song, H. Chen, J. Cheng, C. Li, Y. Liu, X. Chen, "PulseGAN: Learning to Generate Realistic Pulse Waveforms in Remote Photoplethysmography," IEEE Journal of Biomedical and Health Informations, Vol. 25, No. 5, pp. 1373 - 1384, 2021.
- [18] Z. Yu, W. Peng, X. Li, X. Hong, G. Zhao, "Remote Heart Rate Measurement from Highly Compressed Facial Videos: An End-to-end Deep Learning Solution with Video Enhancement," Proceedings of the IEEE/CVF International Conference on Computer Vision, pp. 151 - 160, 2019.

Gyutae Hwang (황규태)



2022 Electronic Engineering from Jeonbuk National University (B.S.)

Career:

2022 Electronics and Telecommunications Research Institute (Researcher)

Field of Interests: Deep learning, Computer vision, Biometric estimation, Biosignal analysis

Email: gyutae741@jbnu.ac.kr

Myeonggeun Park (박명근)



2023 Electronic Engineering from Jeonbuk National University (B.S.)

Career:

2023~ Polaris3D Research (Researcher)

Field of Interests: Deep learning, Computer vision, Object detection, Robotics

Email: jarvis.geun@gmail.com

Sang Jun Lee (이상준)



2011 Electrical Engineering from POSTECH (B.S.)

2018 Electrical Engineering from POSTECH (Ph.D. & M.S.)

Career:

2018~2020 Samsung Advanced Institute of Technology (Senior Researcher)

2020~ Jeonbuk National University (Assis. Prof.)

Field of Interests: Deep learning, Computer vision, Image processing, Robotics

Email: sj.lee@jbnu.ac.kr