

딥러닝 기반 rPPG 모델 사용을 위한 경량 모델 연구

A Study on the Lightweight Model for the Use of Deep Learning-Based rPPG Model

Daeyeol Kim Ph.D*, Kwangkee Lee Ph.D*

* AI research team, TVstorm, Seoul 137875, wagon0004@tvstorm.com, kklee@tvstorm.com

Abstract: The photoplethysmography(PPG) signal contains various information related to cardiovascular diseases, such as information on acute heart attack and arrhythmia, and renin antiosin system variables. The remote PPG(rPPG) method is a method that can measure PPG using a face image taken with a camera without a PPG device. The rPPG method is being studied using deep learning to achieve performance close to PPG. However, it is difficult to run rPPG deep learning models on devices used in real life due to the problem of using large resources and requiring pre/post processing. This study proposes a network concept of the Axis Projection method. The proposed method proposes a method to solve the problem that the STmap method requires a lot of resources for preprocessing and that the 3D CNN model consumes a large amount of memory. In addition, notch learning was made possible by using fft loss and neg-pearson loss at the same time, and compared with the representative rPPG, model, the size of the model was reduced by 1.9~4MB so that it can be used with less resources in the end device.

Keyword: rPPG, notch, axis projection, end device rPPG concept

I. 배경

rPPG기술의 발달로 PPG 장치 없이 카메라로 환자의 건강 상태를 지속적으로 모니터링 가능한 연구가 진행되고 있다. PPG는 빛의 투과 및 반사로 발생하는 혈액 내 헤모글로빈의 물리적 특징을 분석하여 Photoplethysmogram(PTG)를 측정하는 방법이다. 정확도 높은 PPG 정보를 활용하면 맥박, 산소포화도, 혈압, 호흡 수 등 다양한 활력 신호를 취득 할 수 있을 뿐만 아니라 주파수 분석을 통해 뇌졸중 증세 감지, 레닌-안티시오톤계 변수 측정, 교감-부교감 신경계 성분 검출, 호흡 부정맥 검출 등 다양한 신체 정보를 얻을 수 있다[1-2]

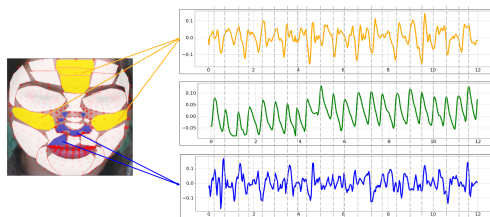


Figure 1. rPPG inference result

Figure 1은 rPPG의 결과 이미지이다. rPPG는 oximeter의 측정 방법을 벤치마킹한 방법으로, 얼굴 내의 혈류 변화로 발생하는 빛의 난반사 정보의 변화량을 분석하여 혈류의 변화를 측정하는 비디오 기반 측정 방법이다[3]. rPPG는 카메라를 이용하여 측정이 가능하므로, 모바일 기기 혹은 카메라가 부착된 IOT 기기의 보급에 따라 실생활 적용이 용이하다.

딥러닝 기술의 발전으로 다양한 방식으로 rPPG 모델의 연구가 진행되어 왔다. 다양한 rPPG 모델 중 baseline으로 이용되는 대표적인 방법으로 Physnet[4], Rhythmnet[5], Deepphys[6]의 방법이 있으며 각각 다른 접근 방식을 갖고 있다. Physnet은 비디오 픽셀 변화의 시공간적 특징을 3D 컨볼루션 네트워크로 분석하는 방법을 제안하였으며, Rhythmnet은 비디오를 STmap으로 변환하는 전처리 과정을 도입하여 PPG의 시계열 특징을 학습하는 방법을 제안하였다. 마지막으로 Deepphys는 DRM 모델을 기반으로 정반사 정보를 없애고 PPG 파형의 변화량을 학습하는 차분 이미지 학습 방안을 제안하였다. 실제 end device 환경에서 rPPG 모델을 적용하기 위해서는 세 가지 문제의 해결이 필요하다.

첫째, 3D 컨볼루션을 이용하기 때문에 큰 크기의 GPU 메모리를 필요로 한다.

둘째, STmap과 같은 전처리 과정이 필요하기 때문에 발생하는 고성능의 CPU, 많은 전처리 시간을 필요로 한다.

셋째, 2D 컨볼루션을 이용하면서 장기적인 시계열 특징을 학습하며, 후처리가 필요한 문제가 발생한다.

넷째, 다양한 주파수 분석을 위해서는 notch가 학습되는 손실 함수가 필요하다[7].

이러한 문제점은 rPPG가 저사양의 end device에서 사용되는 것에 상당한 어려움을 초래할 수 있다. 향후 rPPG 모델이 실생활에 보급되기 위해서 저용량의 메모리, 전/후처리가 최소로 수행되는 모델이 필요하다.

II. 방법

본 연구에서는 end device에서 rPPG 모델을 구동시키기 위해 Axis Projection 방식을 통한 rPPG 추론 모델의 개념을 제안한다. Axis Projection 방식을 이용하기 위한 모델 구조도와 notch 학습을 위한 손실함수를 제안하며, 마지막으로 한정된 데이터 셋을 활용하여 학습 가능성을 제시한다.

1. Axis Projection 방식의 모델 구조도

본 연구에서는 전/후처리가 필요 없는 중단 학습 모델인 Axis Projection 방식의 모델을 제안한다.

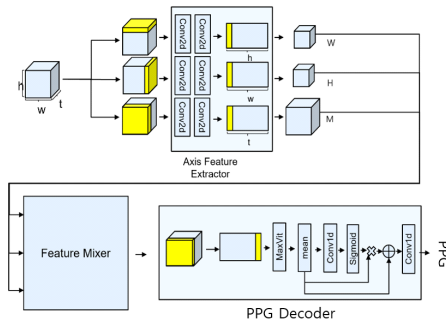


Figure 2. Proposed Model architecture

Figure 2은 제안하는 모델의 구조도이다. 3차원 형태의 데이터인 비디오를 3가지 방향에서 2차원 정보로 슬라이싱 하여 이용하였다. 슬라이싱 된 데이터의 축에 기반한 특징을 뽑아내는 Axis Feature Extractor와 그 특징을 조합하여 PPG를 뽑아 낼 수 있는 Feature로 가공하는 feature mixer, 마지막으로 압축된 정보로 PPG를 생성해 내는 PPG Decoder로 구성하였다.

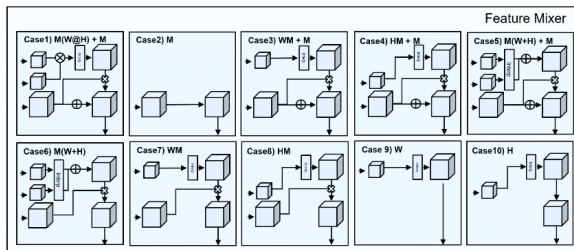


Figure 3. 10 Types of feature mixer

Figure 3는 Feature mixer의 종류이다. 각 축에 특화된 정보들은 개별적으로 이용했을 때 좋은 성능을 기대하기 어렵다. 각 축의 정보를 조합하여 Attentive Feature를 만들게 되면 PPG Decoding 시 우수한 성능을 기대할 수 있다.

2. 손실 함수

일반적으로 rPPG task는 심박 수를 체크하기 위한 목적으로 진행 된다. neg-pearson loss는 대략적인 추

세와 피크를 찾기에 용이한 손실 함수로 rPPG task에서 이용된다[8]. 하지만 neg-pearson loss는 notch를 학습하지 않고 큰 피크를 학습하는 경향이 있으므로 FFT loss와 neg-pearson loss를 함께 사용하였다.

3. 데이터 셋

V4V 데이터 셋은 여성참여자 82명, 남성 참여자 58명으로 구성된 학습 데이터를 제공한다[8]. 각 참여자는 노이즈 생성을 위해 웃거나, 눈을 찡그리는 등 10가지 작업을 수행한다. ground truth는 영상 촬영과 동시에 1000kHz로 샘플링된 PPG 정보, 심장박동, 호흡수가 제공된다. 실제 챌린지에서는 학습데이터로 학습된 모델을 테스트 셋으로 평가를 진행하였지만, 해당 데이터 셋에는 test 셋에 대한 label 정보가 포함되어 있지 않기에 train set을 80:20 비율로 나누어 데이터를 구성하였다.

4. 평가지표

실험 평가를 위해 3가지 평가 지표를 이용하였다.

- Pearson-Correlation(PCC; R) : 주어진 두 신호 사이의 선형관계를 해석하는 방법. 1에 가까울수록 양의 선형적인 관계를 보인다.
- HR-MAE : HR의 절대적인 정확도를 확인 하는데 이용한다.
- HR-RMSE : HR의 표준 분포를 확인하는데 이용한다.

III. 결과

V4V 데이터 셋을 활용하여 10가지 Feature Mixer에 대해 실험을 진행하였다.

Table 1. Experiment result of 10 types of feature mixer

Feature Mixer	HR-MAE	HR-RMSE	R
M(W+H)+M	81.04	83.05	0.00
M(WxH)+M	10.63	14.01	0.24
WM+M	8.49	11.74	0.53
HM+M	8.27	11.35	0.57
M	8.72	11.67	0.53
WM+HM	5.37	8.48	0.67
WM	6.96	9.88	0.65
HM	5.79	8.46	0.64
H	11.13	14.55	0.33
W	11.81	14.81	0.38

Table 1은 10가지 feature mixer에 따른 실험 결과이다. 10가지 종류 중 WM+HM이 좋은 HR-MAE, R 두 가지 측정 지표에서 좋은 성능을 보였다.

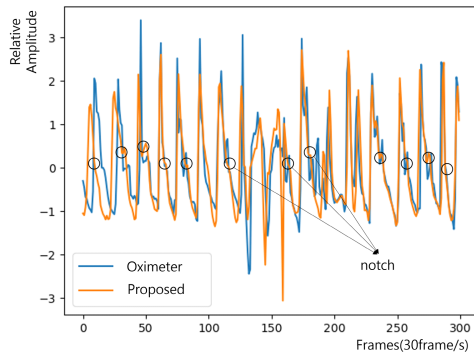


Figure 3. Inference result of WM+HM type feature mixer

그림 3은 WM+HM 타입의 추론 결과이다. 추론 결과 ground truth인 oximeter와 유사한 파형을 가짐을 확인했고, 추가로 notch도 학습이 되었음을 확인하였다.

Table 2. Comparison of Memory Size with Representative Algorithm

Model Name	Memory Size
Deepphys	5.6MB
PhysNet	3.0MB
Proposed Method	1.1MB

Table 2는 대표적인 rPPG 알고리즘의 모델 사이즈이다. Deepphys와 비교 하였을 때 4MB 가량, PhysNet과 비교하였을 때 1.9MB 가량 메모리 하락을 보였다.

IV. 결 론

본 연구에서는 End Device에서 rPPG 모델 활용을 위한 Axis Projection method를 제안하였다. End Device로 흔히 활용되는 스마트폰 그리고 IOT 기기들은 한정된 자원을 활용하기 때문에 적은 리소스를 활용하여 기존의 모델과 유사한 성능을 내는 것이 중요하다. 제안하는 방법의 성능을 극대화시키기 위해 10가지 Feature Mixer에 대해 실험을 진행하였다. End Device에서 활용도를 높이기 위하여 중단간 학습 기법을 채택하였으며, 추후 다양한 질병 모니터링 활용도를 높이기 위해 notch를 학습 할 수 있도록 손실 함수를 설계하였다. Pearson Correlation, HR-MSE, HR-RMSE 세 가지 지표를 이용하여 학습 타당성을 살펴보았다.

V4V 데이터 셋을 활용하여 제안하는 모델을 평가하였을 때, 실제 기기와 유사한 결과를 낼 수 있음을 확인하였다. 하지만 해당 데이터 셋은 검정 배경에 한명의 인물이 단독으로 촬영을 진행하였으므로, 향후 실제 환경에 rPPG 모델을 적용하기 위해선 다양한 노이즈에 강인한 데이터 셋 제작하여 타당성 입증에 추가로 필요하다. 또한, rPPG 분야를 대표하

는 다른 모델과 메모리 사이즈를 비교하였을 때 적게는 1.9MB에서 많게는 5MB 까지 크기를 줄일 수 있으므로 end device에서 사용함이 타당해 보인다.

본 연구를 통하여 end device에서 활용 가능한 rPPG 모델의 타당성을 보였고, 현재 rPPG를 대표하는 모델의 문제점을 보완하였다. 또한 notch가 학습됨을 보임으로 rPPG 알고리즘이 실생활 활력징후 모니터링 관점에서 사용성을 가질 수 있음을 확인했다.

참고문헌

- Kim, D., Kim, J., & Lee, K. (2021). Real-time Vital Signs Measurement System using Facial Image Data. *Journal of Broadcast Engineering*, 26(2), 132 - 142. <https://doi.org/10.5909/JBE.2021.26.2.132>
- Shaffer, F., & Ginsberg, J. P. (2017). An overview of heart rate variability metrics and norms. *Frontiers in public health*, 258.
- Kim, D. Y., Lee, K., & Sohn, C. B. (2021). Assessment of ROI Selection for Facial Video-Based rPPG. *Sensors*, 21(23), 7923.
- Yu, Z., Li, X., & Zhao, G. (2019). Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks. *arXiv preprint arXiv:1905.02419*.
- Niu, X., Shan, S., Han, H., & Chen, X. (2019). Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing*, 29, 2409-2423.
- Chen, W., & McDuff, D. (2018). Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the european conference on computer vision (ECCV)* (pp. 349-365).
- Nara, S., Kaur, M., & Verma, K. L. (2014). Novel notch detection algorithm for detection of dirotic notch in PPG signals. *International Journal of Computer Applications*, 86(17), 36-39.
- Schober, P., Boer, C., & Schwarte, L. A. (2018). Correlation coefficients: appropriate use and interpretation. *Anesthesia & Analgesia*, 126(5), 1763-1768.
- Revanur, A., Li, Z., Ciftci, U. A., Yin, L., & Jeni, L. A. (2021). The first vision for vitals (v4v) challenge for non-contact video-based physiological estimation. In *Proceedings of the IEEE/CVF International Conference on Computer Vision* (pp. 2760-2767).

사사

이 논문은 2021년도 정부(과학기술정보통신부)의 재원으로 정보통신기획평가원의 지원을 받아 수행된 연구임(No. 2021-0-00900)