Video-Based Pulse Estimation through Spatiotemporal Meta-Learning

Eirini Kateri*, Katayoun Farrahi*

* University of Southampton, Southampton, UK
E.Kateri@soton.ac.uk, K.Farrahi@soton.ac.uk

Accurate and reliable pulse estimation through video-based monitoring systems has captured research's interest by promising to revolutionize the way we view human health. The human pulse serves as a fundamental physiological marker, offering insights into cardiovascular health and overall well-being. Traditionally, pulse measurement has relied on contact methods, which, while effective in clinical settings, are often impractical or unavailable. Video-based pulse estimation leverages the availability of cameras in modern devices, to capture subtle color or motion variations in facial skin generated during the cardiac cycle. Despite significant advancements in remote photoplethysmography technology, several challenges persist, including motion artifacts, data homogeneity and availability, which impact the accuracy and reliability of such solutions.

We introduce a novel framework aimed at addressing some of these challenges. Our methodology involves a stabilization method to mitigate motion artifacts, a spatiotemporal representation of video data and a meta-learning component to optimize estimation accuracy.

We identify the face in the first frame and calculate the central point of the facial bounding box. Subsequently, we employ the persistent independent particles (PIPs) algorithm [1] for precise tracking of this central point through frames, enabling us to extract stabilized sub-frames from the video. These frames undergo a transformation into spatiotemporal images by slicing each frame into 3-pixel wide vertical segments and then concatenating these slices side by side in sequence, ensuring each slice from successive frames forms a continuous, unified image. This method ensures consistent analysis of the same facial region across time. These images are fed into a convolutional neural network (CNN), predicting the the number of beats per spatiotemporal image. To enhance the accuracy of pulse estimation, our framework integrates a meta-learning component. This involves training a multi-layer perceptron (MLP) to distinguish between "good" and "bad" images based on their mean absolute error (MAE) against a set threshold. This selection, refined through a 10-fold cross-validation process, enables our framework to focus on higher-quality data, thereby boosting the framework's accuracy in pulse signal estimation. Our validation encompasses a series of experiments, including blind meta-learning (leave one subject out), few-shot meta-learning, and calibrated meta-learning, to corroborate our framework's effectiveness.

We demonstrate our framework's performance on two benchmark datasets, MMSE-HR [2] and UBFC-Phys [3]. For MMSE-HR we achieve a MAE of 5.94 $\pm$ 4.78 beats per minute (BPM), 5.13 $\pm$ 4.11 BPM and 2.06 $\pm$ 2.35 BPM for blind, few-shot and calibrated experiments respectively. For UBFC-Phys we achieve a MAE of 5.62 $\pm$ 4.76 BPM, 5.24 $\pm$ 3.81 BPM and 3.05 $\pm$ 2.82 BPM for blind, few-shot and calibrated experiments respectively, outperforming previous spatiotemporal approaches and meta learning methods utilizing spatiotemporal features. This research seeks to address common challenges for remote photoplethysmography and to assist in the adoption of these technologies in clinical settings.

REFERENCES:

[1] Adam W Harley, Zhaoyuan Fang, and Katerina Fragkiadaki. Particle video revisited: Tracking through occlusions using point trajectories. In European Conference on Computer Vision, pages 59–75. Springer, 2022.

[2] Zheng Zhang, Jeff M Girard, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canavan, Michael Reale, Andy Horowitz, Huiyuan Yang, et al. Multimodal spontaneous emotion corpus for human behavior analysis. In Proceedings of the IEEE conference on computer vision and pattern recognition, pages 3438–3446, 2016.

[3] Rita Meziati Sabour, Yannick Benezeth, Pierre De Oliveira, Julien Chappe, and Fan Yang. Ubfc-phys: A multimodal database for psychophysiological studies of social stress. IEEE Transactions on Affective Computing, 14(1):622–636, 2021.