The Way to my Heart is through Contrastive Learning: Remote Photoplethysmography from Unlabelled Video

John Gideon* Simon Stent*

TOYOTA RESEARCH INSTITUTE

2021 ICCV OCTOBER 11-17 VIRTUAL

TL;DR You can train a neural network to estimate heart rate from videos without any ground truth data

Introduction

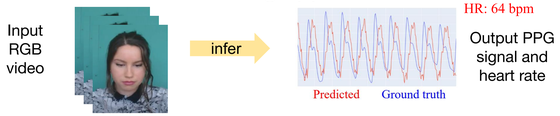
What?

Remote **P**hoto**P**lethysmo**G**raphy (“rPPG”)

from a distance — using light — fluctuations — write down measurements

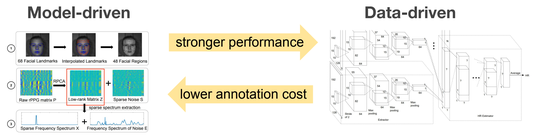
Why?

Health monitoring applications



What’s new?

Show how one can train a deep neural network to detect this signal without ground truth training data



e.g. Self-Adaptive Matrix Completion [Tulyakov et al, 2016], SparsePPG [Nowara et al, 2018] (above)

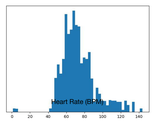
e.g. HR-CNN [Špetlík et al, 2018] (above), RhythmNet [Niu et al, 2019], Meta-rPPG [Lee et al, 2020]

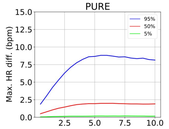
- Strong task performance with zero annotation cost

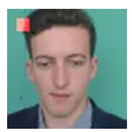
Our assumptions

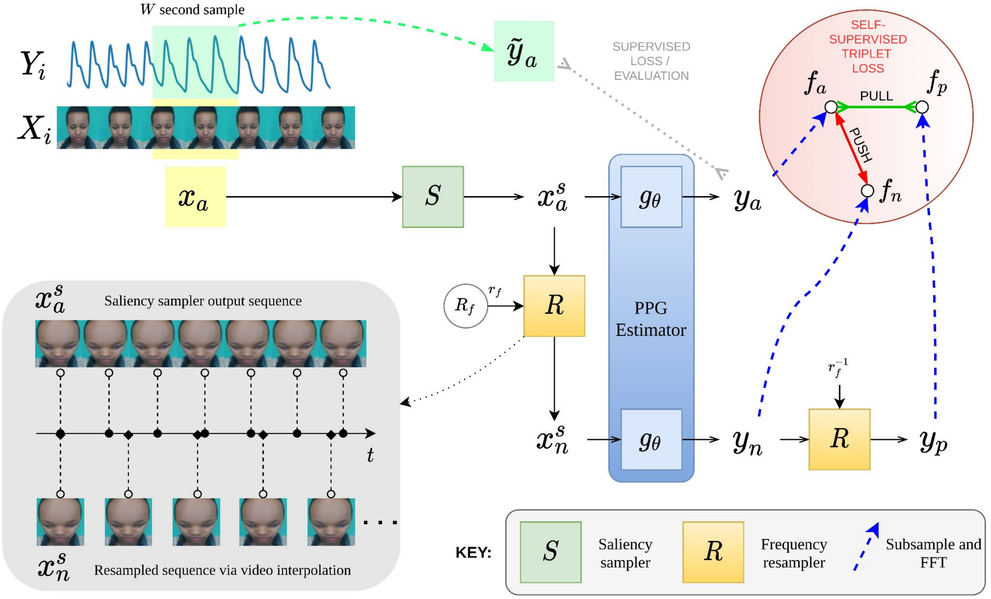
We assume that the signal of interest:

1. lies within a certain frequency range (40-250 bpm)
2. does not vary rapidly over short time intervals (10s)
3. is the dominant visual signal within the target freq. band









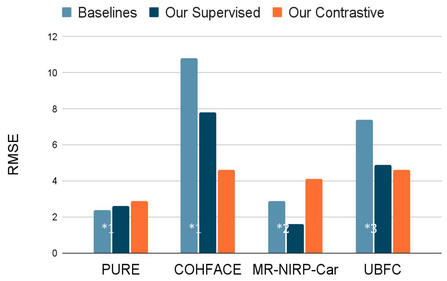
7 steps to understand our method

1. 10 second input video → x_a
2. Video optionally passed through saliency sampler (S), which learns a task-salient mask while resampling input according to that mask → x_a^s

3. Video resampled in time using a factor R_f → negative sample x_n^s
4. Both x_a^s and x_n^s are fed to a PPG estimator (3D CNN) → y_a and y_n
5. y_n resampled to original rate via R_f^{-1} → positive sample y_p
6. PPGs y_a, y_n, y_p are randomly subsampled to 5s views and FFTed → f_a, f_n, f_p
7. Tripet loss encourages dist(f_a, f_n) to be large and dist(f_a, f_p) to be small

Results

Baselines
1) HR-CNN [Špetlík et al, 2018]
2) SparsePPG [Nowara et al, 2018]
3) Meta-rPPG [Lee et al, 2020]



- Compares favorably to supervised approaches
- Saliency sampler improves interpretability

We will present our winning entry, based on this contrastive approach, to the Vision 4 Vitals workshop challenge on Saturday. Join us to find out more!

Conclusions

- A self-supervised neural network to estimate heart rate from video
- Our approach performs well on existing datasets and is interpretable
- Future extensions: stabilize for input motion and estimate uncertainty

Further results, paper links, and code: https://github.com/ToyotaResearchInstitute/RemotePPG