egoPPG: Heart Rate Estimation from Eye-Tracking Cameras in Egocentric Systems to Benefit Downstream Vision Tasks

Supplementary Material

10. Related datasets

Tab. 10 gives a comparison of the dataset size and activities of some related remote photoplethysmography (rPPG) datasets. In terms of hours of recordings and recorded frames, *egoPPG-DB* is among the largest dataset. Furthermore, we see that all comparable rPPG datasets only include activities with very little motion and heart rate (HR) changes such as watching videos, head rotations or talking. In contrast, *egoPPG-DB* features a wide variety of challenging everyday activities, such as kitchen work, dancing and riding an exercise bike, which induce significant motion artifacts and HR changes.

11. Excluded tasks

For all participants and activities, we checked the mean absolute error (MAE) between the predicted HR from our custom contact PPG sensor on the nose and the gold standard ECG from the chest belt. We excluded all tasks with an MAE over 3.0 beats per minute (bpm), which can happen, for example, when the PPG sensor loses alignment with the angular artery due to movement. In this way, we ensured that the photoplethysmography (PPG) signal from the nose, which we used as the target signal to train our model, is highly accurate. As a result, we had to exclude 20 out of the 150 tasks (13%), which we list in Tab. 6. We can see that this applied only to tasks with more motion (dancing, exercise bike, and walking). Since the participants had to walk multiple stairs throughout the data recording, this mostly happened during walking.

Activity	Excluded participants
Watch video	—
Office work	—
Kitchen work	—
Dancing	012, 015
Exercise bike	009, 012, 014, 015, 016, 023
Walking	004, 012, 013, 014, 018, 021, 022

Table 6. Detailed table of all excluded tasks.

12. Detailed description of activities

Tab. 11 gives a comprehensive description of the actions for each activity during our recording. Generally, participants were free to talk during the entire duration of the recording and conduct the tasks as they would do it normally. For example, during the kitchen work, the participants were completely free to prepare the sandwich and if they would like to eat or drink while doing it.

13. Data recording

In Fig. 8, we show a variety of different images and people of our data recording from a third person view to visualize the apparatus and capture protocol. All participants visible in these images explicitly agreed to be visualized.

14. Initial signal verification

In Fig. 5, we show the raw mean intensity values after spatial cropping of the skin region and the eye region (see Fig. 4) compared to the ground truth contact PPG signal from the nose. We can clearly see that the blood volume pulse is present both in the eyes and skin region with the skin region having a higher signal-to-noise ratio (SNR) compared to the eyes.

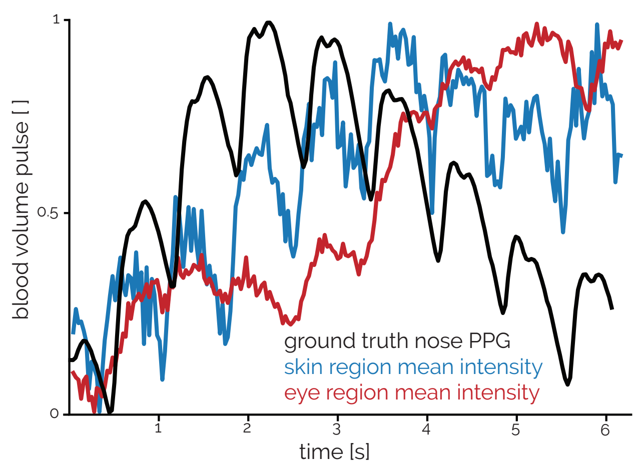

Figure 5. Example raw mean intensity of the skin and eye region, showing the higher SNR for the skin region around the eyes compared to the eyes.

15. Variance of results

In Fig. 6 we show the boxplot of the MAEs of the predictions of *PulseFormer* on *egoPPG-DB* by split. The interquartile range across all splits is between 1.7 and 10.5 bpm.

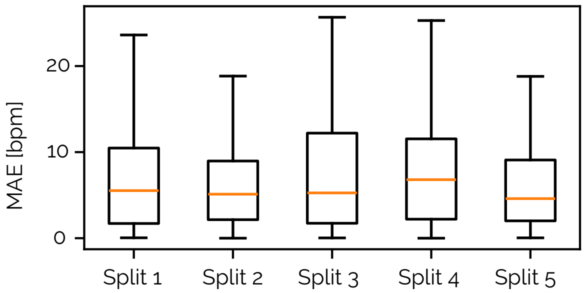

Figure 6. Boxplot of the MAEs of the predictions of *PulseFormer*.

16. Cross-dataset evaluation

We evaluated *PulseFormer* and the two strongest baselines when training on three conventional rPPG datasets (MMPD [82], UBFC-rPPG [6], and PURE [79]) and testing on *egoPPG-DB* (Tab. 7), and vice versa (Tab. 8). For the rPPG datasets, we extracted the eye region using MediaPipe [47], resized to 48×128, and converted to grayscale. *PulseFormer* consistently outperforms the baselines across all scenarios and datasets (except one case), showing strong generalization to unseen data. Please note that we can only evaluate *PulseFormer* w/o MITA as conventional rPPG datasets do not contain IMU data from the participants' heads.

Train Set	**Model**	MAE	MAPE
MMPD	PhysFormer	20.56	27.06
	FactorizePhys	Not converging	
	***PulseFormer* w/o MITA**	**13.66**	**16.64**
UBFC-rPPG	PhysFormer	18.32	23.63
	FactorizePhys	18.58	24.46
	***PulseFormer* w/o MITA**	**14.83**	**18.57**
PURE	PhysFormer	24.39	24.94
	FactorizePhys	13.20	15.44
	***PulseFormer* w/o MITA**	**12.99**	**13.46**

Table 7. Results (MAE) when training on conventional rPPG datasets and testing on *egoPPG-DB*.

Model	**MMPD**		**UBFC-rPPG**		**PURE**	
	MAE	MAPE	MAE	MAPE	MAE	MAPE
PhysFormer	11.76	14.57	16.80	16.46	23.89	37.50
FactorizePhys	12.06	15.11	**14.28**	**14.98**	26.10	40.62
***PulseFormer* (ours)**	**11.48**	**15.08**	15.09	15.81	**23.56**	**36.71**

Table 8. Results (MAE) when training on *egoPPG-DB* and testing on conventional rPPG datasets.

17. HR distribution

egoPPG-DB exhibits the widest HR range (44–164 bpm, see Fig. 7) and significantly more motion (e.g., dancing, exercise bike) than other evaluated rPPG datasets, where participants typically sit calmly at a table.

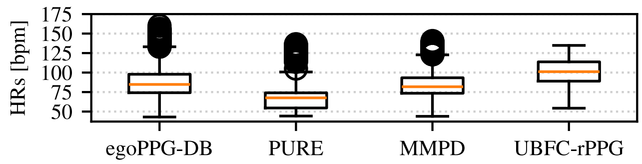

Figure 7. Boxplot of HRs of *egoPPG-DB* and three rPPG datasets.

18. Downstream performance comparison

HR features from the other evaluated baselines perform progressively worse than those from *PulseFormer* when used for proficiency estimation on EgoExo4D, highlighting the importance of accurate HR estimation for downstream tasks (see Tab. 9).

Model	**Ego+HR**	**Exo+HR**	**Ego+Exo+HR**
FactorizePhys	44.62	36.72	40.13
PhysFormer	44.39	36.66	43.07
***PulseFormer* (ours)**	**45.29**	**37.67**	**43.94**

Table 9. Downstream performance (accuracy) on EgoExo4D using the HR predictions from the three best baseline models.

Dataset	Part.	Frames	Hours	Tasks
PURE [79]	10	110 K	1	Resting, talking, small head movements
MAHNOB-HCI [77]	27	2.6 M	12	Watching videos
MMPD [82]	33	1.2 M	11	Resting, head rotation, selfie videos
MMSE-HR [95]	40	310 K	2	Talking, watching videos, experiencing different emotions
UBFC-rPPG [6]	43	150 K	1.5	Gaming on a computer
UBFC-PHYS [70]	56	2.4 M	19	Resting, Trier Social Stress Test
OBF [44]	106	3.8 M	18	Resting with varying HR levels
VIPL-HR [63]	107	**4.3 M**	**20**	Resting, talking, head rotation, different lighting conditions
SCAMPS (synthetic) [54]	**2800**	1.7 M	16	Different facial actions
egoPPG-DB (ours)	25	1.4 M	13	Watching videos, office and kitchen work, dancing, biking, walking

Table 10. Summary of existing datasets for rPPG.

Activity	Actions	Description
Watch video	Watch a documentary	Watch a relaxing documentary on a computer.
Office work	Work on a computer	Randomly browse through websites and type text from a PDF into Word.
	Write on a paper	Write a text from a PDF on a computer onto a piece of paper.
	Talk to the experimenter	Have a free, unscripted conversation with the experimenter.
Walking	Walk to the kitchen	Walk along a hallway, down the stairs into the kitchen.
Kitchen work	Get ingredients	Get all ingredients for a sandwich from the fridge.
	Cut vegetables	Get a cutting board, knife and a plate and cut vegetables.
	Prepare a sandwich	Put the bread into the toaster and afterward freely prepare sandwich.
	Eat sandwich/drink	Participants are free to eat the sandwich or drink during the recording.
	Wash the dishes	Wash everything used while preparing the sandwich.
Walking	Walk to the dancing room	Walking along a hallway into a new room for dancing and biking.
Dancing	Follow random dance video	Choose a dance video and afterward follow it.
Exercise bike	Ride an exercise bike	Ride an exercise bike with moderate to high intensity.
Walking	Walk back to the physical location of the start	Walk back to the physical location of the start either up the stairs or using the elevator.

Table 11. Detailed capture protocol and action descriptions of the *egoPPG-DB* dataset.

Figure 8. Additional images of the data recording showing the variety of everyday activities our dataset includes.