

RPPG-HIBA: HIERARCHICAL BALANCED FRAMEWORK FOR REMOTE PHYSIOLOGICAL MEASUREMENT

A INTRODUCTION

This document presents the supplementary materials omitted from the main paper due to the space limitation. In Section B, details about the five datasets are presented. In Section C, details of ablation experiment results for parameters α and M . In Section D, details about experimental results for the *Domain Balance Analysis on BUAA, VIPL* section of the main paper are presented. In Section E, details about experimental results for the *Phys-label Balance Analysis on V4V, VIPL* section of the main paper are presented. In Section F, Experimental results for the *HRV Estimation* section of the main paper are presented.

B DATASET DETAILS

PURE [8], the dataset contains 10 objects that were recorded in 6 different setups resulting in a total number of 60 sequences of 1 minute each, and the video was captured in a low-illumination environment. Ground-truth BVP signal is included.

BUAA [12], this data set focuses on lighting conditions and varied illuminance in the range: $\{10^0, 10^{0.2}, 10^{0.4}, 10^{0.6}, 10^{0.8}, 10^{1.0}, 10^{1.2}, 10^{1.4}, 10^{1.6}, 10^{1.8}, 10^{2.0}\}$ lux (equivalent to 1.0, 1.6, 2.5, 4.0, 6.3, 10.0, 15.8, 25.1, 39.8, 63.1, 100.0 lux). Since the face region is hardly extracted in low-light videos below 10lux, we used 10-100 lux data as the research subject. Contains ground-truth BVP signals.

UBFC-rPPG [1], contains measurement videos of 42 subjects simulating realistic situations. These subjects also have large differences in appearance, such as skin color, whether they wear glasses, whether they have beards, etc. Contains ground-truth BVP signals.

VIPL [5], the data set contains 107 subjects, and each object takes videos of 9 scenarios, including v1 (stable scenario), v2 (motion scenario), v3 (talking scenario), v4 (dark scenario), v5 (bright scenario), v6 (long distance scenario), v7 (exercise scenario), v8 (phone stable scenario), v9 (phone motion scenario). Nine scenarios were collected with four different acquisition devices, named as source1-4. Notably, this paper does not study the NIR data sources, and only sources1-3 are used. Besides, in this paper, we do not use the uncalibrated BVP signal of the dataset for the time being and only use the HR value labels.

V4V [7]. There are 179 subjects in total, and each subject has at most 10 experimental tasks. These tasks are designed to obtain the changes in physiological indicators of each subject. Unfortunately, this dataset does not collect BVP signals.

C ABLATION STUDY OF THE ADL AND CCR ALGORITHMS

We performed ablation studies on α and M shared by the ADL and CCR algorithms. α is set to 0.3, 0.5, 0.7, 0.9. M is set to 16, 32, 64, 128. We can observe from Tab. 1 and Tab. 2 that the ADL and CCR algorithms are optimal when $\alpha=0.7$, $M=64$. This is because a larger α causes the embeddings of differently distributed domains in the memorybank to cluster more quickly towards their respective centers, thus strengthening the tightness of the features of differently distributed domains. This improves the discriminative properties of the embeddings. In addition, when M takes a larger value, the performance is not improved more, but the computational consumption is increased. When M takes a smaller value, there are fewer embeddings in the memorybank, which does not accurately calculate the compactness between embeddings, and it is also more challenging to learn the center of a more unbiased domain.

D DOMAIN BALANCE ANALYSIS ON BUAA, VIPL

In the main paper, we presented the MAE results on different domains. This section details the results of MAE, RMSE, and Pearson correlation coefficients on VIPL and BUAA for different domains, as shown in Tab. 3 and Tab. 4. In addition, to further validate the effectiveness of anti-spurious domain center learning, we perform experiments on the consistency of training features center with testing features center on cross-dataset testing on VIPL protocol. In general, the smaller the distance between the training features center and the testing features center under the same phys-label indicates that the features learned by the model are more consistent and more capable of generalization between the train and test data. Then, the model is more likely to make accurate predictions for new and unseen data, which can mitigate the negative impact of spurious correlation. As shown in Fig. 1, rows (a) and (b) show the visualization results of the feature centers without and with ADL, respectively. It can be seen that the training and testing feature centers are closer to each other when ADL is employed. The results show that ADL can extract the domain-agnostic features against domain imbalance.

E PHYS-LABEL BALANCE ANALYSIS ON V4V, VIPL

In the main paper, we presented the MAE results on disjoint subsets. This section presents detailed results on disjoint subsets on VIPL and V4V, including MAE, RMSE, and Pearson correlation coefficient, as shown in Tab. 5. It is worth noting that the Pearson correlation coefficients of all methods in Tab. 5 are generally smaller on the many-shot subset than on the medium-shot and few-shot subsets. The main reason for this abnormality is that Pearson correlation coefficients correlate more with the trend between ground truth(GT) and prediction. They are less sensitive to the consistency between ground truth and prediction. We demonstrate this in Fig. 2. In row (a), we plot the values of GT and prediction for different subsets. In row (b), we smoothed the values of GT and prediction to show the trend between them more clearly. The dashed boxes mark the cases where the trends of GT and prediction are opposite in the many-shot subset, which is also the main reason for the low Pearson correlation coefficient. In addition, we plotted Bland-Altman plots of ground truth heart rate (phys-label) and

Method	α	M	v1/s			v2/s			v3/s			all		
			MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* + \mathcal{L}_{ADL}	0.3	64	8.18	10.95	0.56	7.91	10.58	0.64	8.76	12.22	0.52	8.10	11.12	0.55
	0.5	64	8.20	10.96	0.57	7.93	10.60	0.65	8.69	12.19	0.54	8.02	11.02	0.58
	0.7	64	7.82	10.49	0.60	7.52	10.23	0.66	8.37	11.72	0.57	7.67	10.60	0.60
	0.9	64	8.20	11.03	0.58	8.06	10.74	0.64	8.64	12.15	0.55	8.10	11.17	0.57
	0.7	16	8.17	10.86	0.58	8.04	10.69	0.64	8.53	12.04	0.56	8.00	10.97	0.58
	0.7	32	7.84	10.61	0.60	7.55	10.22	0.68	8.28	11.74	0.58	7.72	10.73	0.60
	0.7	128	7.96	10.72	0.58	7.64	10.29	0.66	8.57	11.96	0.54	7.82	10.80	0.58
(a) Cross-dataset testing on v1/s, v2/s, and v3/s domains on VIPL.														
Method	α	M	v4/s			v5/s			v6/s			all		
			MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* + \mathcal{L}_{ADL}	0.3	64	7.72	10.38	0.58	7.99	11.24	0.51	7.99	11.06	0.59	8.10	11.12	0.55
	0.5	64	7.44	10.00	0.63	8.00	11.25	0.52	7.99	10.90	0.62	8.02	11.02	0.58
	0.7	64	7.00	9.60	0.65	7.61	10.79	0.56	7.58	10.48	0.64	7.67	10.60	0.60
	0.9	64	7.63	10.32	0.61	8.01	11.33	0.52	7.98	11.03	0.62	8.10	11.17	0.57
	0.7	16	7.38	10.02	0.64	7.88	11.02	0.55	8.07	10.92	0.61	8.00	10.97	0.58
	0.7	32	7.15	9.78	0.65	7.72	10.98	0.55	7.71	10.64	0.64	7.72	10.73	0.60
	0.7	128	7.20	9.82	0.63	7.76	10.96	0.54	7.70	10.69	0.62	7.82	10.80	0.58
(b) Cross-dataset testing on v4/s, v5/s, and v6/s domains on VIPL.														
Method	α	M	v7/s			v8/s			v9/s			all		
			MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* + \mathcal{L}_{ADL}	0.3	64	7.66	10.86	0.53	8.09	10.17	0.51	9.56	13.24	0.38	8.10	11.12	0.55
	0.5	64	7.46	10.59	0.56	7.98	10.26	0.54	9.35	13.06	0.43	8.02	11.02	0.58
	0.7	64	7.37	10.36	0.58	7.64	9.64	0.56	8.93	12.60	0.43	7.67	10.60	0.60
	0.9	64	7.62	10.84	0.55	8.22	10.56	0.51	9.60	13.42	0.41	8.10	11.17	0.57
	0.7	16	7.51	10.62	0.56	8.05	10.30	0.53	9.24	12.97	0.44	8.00	10.97	0.58
	0.7	32	7.34	10.51	0.57	7.76	9.92	0.57	9.06	12.82	0.45	7.72	10.73	0.60
	0.7	128	7.48	10.60	0.55	7.82	9.91	0.55	9.02	12.76	0.43	7.82	10.80	0.58
(c) Cross-dataset testing on v7/s, v8/s, and v9/s domains on VIPL.														
Method	α	M	s1/v			s2/v			s3/v			all		
			MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* + \mathcal{L}_{ADL}	0.3	64	8.17	11.43	0.55	7.92	10.85	0.57	8.27	11.20	0.52	8.10	11.12	0.55
	0.5	64	8.25	11.45	0.57	7.77	10.69	0.60	8.15	11.06	0.55	8.02	11.02	0.58
	0.7	64	7.80	10.96	0.60	7.50	10.38	0.61	7.78	10.57	0.58	7.67	10.60	0.60
	0.9	64	8.27	11.52	0.57	7.89	10.87	0.59	8.22	11.23	0.55	8.10	11.17	0.57
	0.7	16	8.20	11.36	0.58	7.77	10.68	0.60	8.12	11.00	0.56	8.00	10.97	0.58
	0.7	32	7.89	11.11	0.60	7.51	10.42	0.62	7.85	10.78	0.58	7.72	10.73	0.60
	0.7	128	7.97	11.14	0.58	7.66	10.57	0.60	7.89	10.79	0.56	7.82	10.80	0.58
(d) Cross-dataset testing on s1/v, s2/v, and s3/v domains on VIPL.														

Table 1: Cross-dataset testing on different domains on VIPL for domain imbalance. The dataset contains nine different scenarios (v1-v9), which are captured with three different devices (s1-s3). We consider the data of different acquisition devices for the same scenario as one domain, e.g., the v1/s domain represents all the data captured by the s1,s2, and s3 devices in the v1 scenario. The same acquisition devices for different scenarios are also considered as one domain, e.g., the s1/v domain, which represents all the data of v1-v9 captured by the s1 device. Bold indicates the best ablation results for the baseline combined with ADL on parameters α and M .

rPPG-HiBa's predicted heart rate in Fig. 2(c) to analyze the consistency of the predictions on the subsets. The lines represent the mean and the 95% consistency bounds. In Fig. 2(c), rPPG-HiBa shows the best predictive consistency on many-shot. Hence, despite rPPG-HiBa's lower Pearson correlation coefficient on the many-shot subset, the predictive consistency is higher and still achieves satisfactory MAE and RMSE results on many-shot.

Method	α	M	NI			MI			HI			all		
			MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* + \mathcal{L}_{ADL}	0.3	64	3.08	3.98	0.91	2.83	3.72	0.94	2.80	3.75	0.95	2.92	3.82	0.94
	0.5	64	3.01	3.90	0.92	2.84	3.71	0.94	2.78	3.77	0.96	2.87	3.79	0.94
	0.7	64	2.93	3.84	0.94	2.81	3.67	0.96	2.77	3.71	0.96	2.84	3.74	0.95
	0.9	64	2.99	3.95	0.92	2.86	3.74	0.94	2.80	3.73	0.95	2.88	3.81	0.94
	0.7	16	3.04	3.98	0.92	2.85	3.73	0.94	2.83	3.76	0.94	2.90	3.82	0.94
	0.7	32	2.96	3.86	0.93	2.83	3.70	0.94	2.75	3.69	0.96	2.85	3.75	0.95
	0.7	128	2.99	3.95	0.93	2.87	3.78	0.93	2.78	3.72	0.96	2.88	3.82	0.95

Table 2: Cross-dataset testing on different domains on BUAA for domain imbalance. The dataset contains rich "illumination" domain information. We set the samples with 10.0lux and 15.8lux as "normal illumination" (NI), the samples with 25.1lux and 39.8lux as "medium illumination" (MI), and the samples with 63.1 lux and 100 lux as "high illumination" (HI). Bold indicates the best ablation results for the baseline combined with CCR on parameters α and M .

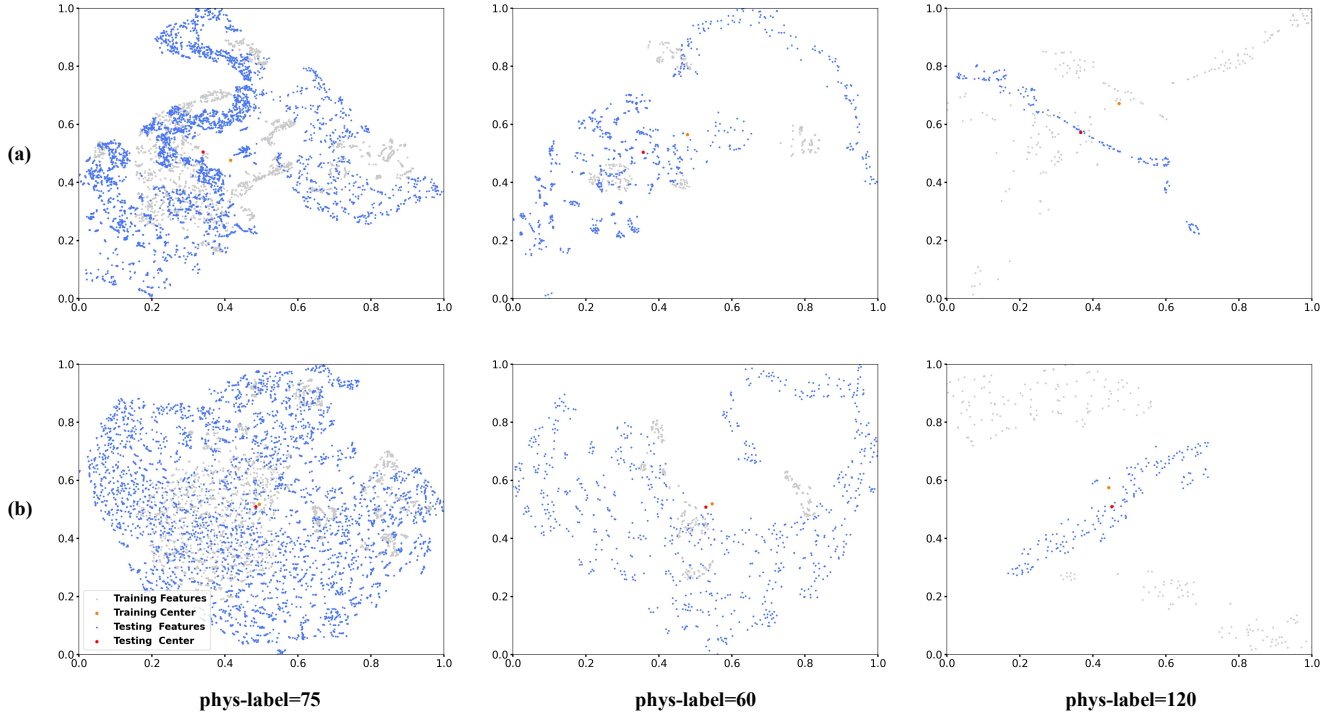


Figure 1: Cross-dataset testing on VIPL protocol. Examples of visualization of features and feature centers under different phys-labels. (a) without ADL. (b) with ADL.

F HRV ESTIMATION

In addition, we also use the HRV and HR-bmp metrics to evaluate the model's performance on data with BVP signals. We use two datasets for training (any two of UBFC, PURE, BUAA) and one other for testing following [4]. As shown in Tab. 6, compared with the traditional methods GREEN, CHROM, POS, and Baseline, our proposed method has achieved the best evaluation on LF, HF, LF/HF, and HR-bmp indicators. The results show that alleviating the hierarchical imbalance problem improves the physiological signal measurement task.

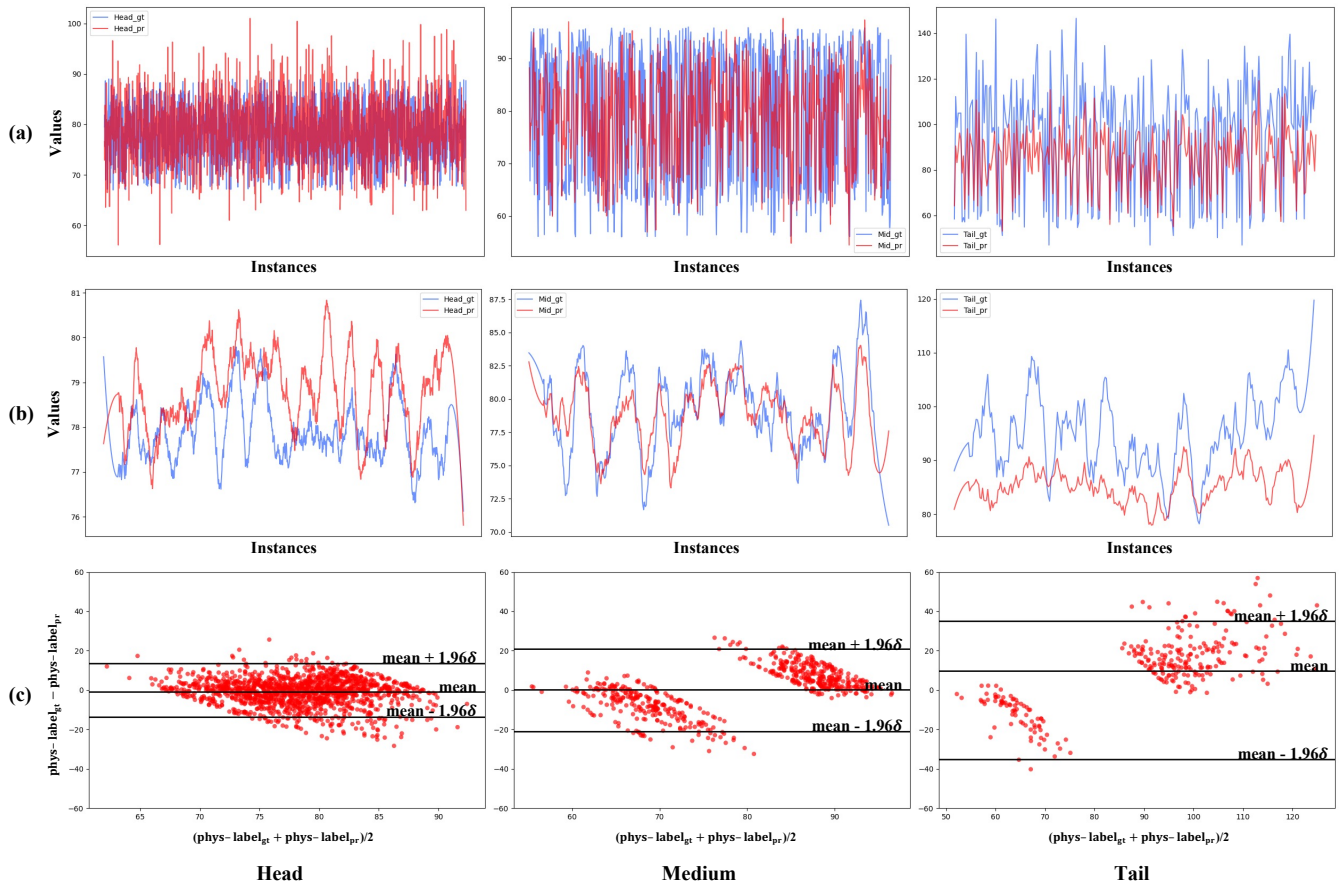


Figure 2: Cross-dataset testing on VIPL protocol. (a) Values of GT and prediction for different subsets. (b) Smoothed values of GT and prediction. (c) Bland-Altman plots demonstrating the consistency between GT and prediction across different subsets.

Method	v1/s			v2/s			v3/s			all		
	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* [6]	9.53	12.56	0.49	8.28	11.09	0.54	11.61	14.53	0.46	8.97	12.16	0.49
Dual-GAN* [3]	8.81	11.22	0.54	9.16	12.15	0.44	9.74	12.81	0.50	8.88	11.69	0.50
NEST* [4]	7.05	10.13	0.62	7.33	10.48	0.65	8.21	11.80	0.57	7.57	10.87	0.60
DOHA* [9]	7.94	10.53	0.60	7.61	10.81	0.63	8.49	11.83	0.57	7.89	11.13	0.56
Baseline* + \mathcal{L}_{ADL}	7.82	10.49	0.60	7.52	10.23	0.66	8.37	11.72	0.57	7.67	10.60	0.60
rPPG-HiBa**	7.20	10.17	0.62	7.26	10.33	0.66	7.81	10.95	0.65	7.34	10.41	0.61

(a) Cross-dataset testing on v1/s, v2/s, and v3/s domains on VIPL.

Method	v4/s			v5/s			v6/s			all		
	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* [6]	7.62	10.25	0.56	8.29	11.88	0.46	8.36	11.92	0.53	8.97	12.16	0.49
Dual-GAN* [3]	8.70	11.19	0.48	7.78	10.42	0.56	8.11	10.68	0.61	8.88	11.69	0.50
NEST* [4]	5.82	9.07	0.69	7.94	10.85	0.55	7.12	10.24	0.63	7.57	10.87	0.60
DOHA* [9]	6.21	9.46	0.66	8.22	11.38	0.51	7.25	10.39	0.63	7.89	11.13	0.56
Baseline* + \mathcal{L}_{ADL}	7.00	9.60	0.65	7.61	10.79	0.56	7.58	10.48	0.64	7.67	10.60	0.60
rPPG-HiBa**	6.34	9.61	0.65	7.73	10.67	0.58	7.32	10.45	0.64	7.34	10.41	0.61

(b) Cross-dataset testing on v4/s, v5/s, and v6/s domains on VIPL.

Method	v7/s			v8/s			v9/s			all		
	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* [6]	7.95	11.06	0.49	9.12	12.67	0.44	12.21	15.23	0.29	8.97	12.16	0.49
Dual-GAN* [3]	9.29	12.44	0.39	8.63	12.16	0.44	10.87	13.28	0.37	8.88	11.69	0.50
NEST* [4]	8.07	11.02	0.50	9.39	12.84	0.43	10.11	14.77	0.31	7.57	10.87	0.60
DOHA* [9]	8.38	11.52	0.50	9.15	12.39	0.48	10.06	14.60	0.32	7.89	11.13	0.56
Baseline* + \mathcal{L}_{ADL}	7.37	10.36	0.58	7.64	9.64	0.56	8.93	12.60	0.43	7.67	10.60	0.60
rPPG-HiBa**	7.29	10.25	0.59	7.33	10.41	0.53	8.68	11.59	0.48	7.34	10.41	0.61

(c) Cross-dataset testing on v7/s, v8/s, and v9/s domains on VIPL.

Method	s1/v			s2/v			s3/v			all		
	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* [6]	9.39	12.60	0.47	8.47	11.75	0.50	9.26	12.31	0.47	8.97	12.16	0.49
Dual-GAN* [3]	8.84	11.55	0.56	8.74	11.53	0.50	9.12	12.02	0.48	8.88	11.69	0.50
NEST* [4]	7.66	10.82	0.60	7.10	10.36	0.62	8.08	11.29	0.53	7.57	10.87	0.60
DOHA* [9]	7.85	10.79	0.60	7.67	10.97	0.57	8.20	11.61	0.53	7.89	11.13	0.56
Baseline* + \mathcal{L}_{ADL}	7.80	10.96	0.60	7.50	10.38	0.61	7.78	10.57	0.58	7.67	10.60	0.60
rPPG-HiBa**	7.31	10.34	0.62	7.33	10.40	0.62	7.39	10.49	0.59	7.34	10.41	0.61

(d) Cross-dataset testing on s1/v, s2/v, and s3/v domains on VIPL.

Table 3: Cross-dataset testing on different domains on VIPL. The dataset contains nine different scenarios (v1-v9), which are captured with three different devices (s1-s3). We consider the data of different acquisition devices for the same scenario as one domain, e.g., the v1/s domain represents all the data captured by the s1, s2, and s3 devices in the v1 scenario. The same acquisition devices for different scenarios are also considered as one domain, e.g., the s1/v domain, which represents all the data of v1-v9 captured by the s1 device.

Method	NI			MI			HI			all		
	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* [6]	3.84	5.49	0.83	3.30	5.12	0.84	3.02	4.89	0.83	3.38	5.17	0.84
Dual-GAN* [3]	3.88	5.53	0.83	3.07	5.00	0.84	3.30	5.16	0.85	3.41	5.23	0.84
NEST* [4]	3.11	4.89	0.84	2.45	3.30	0.94	2.47	3.31	0.93	2.67	3.89	0.90
DOHA* [9]	3.29	5.05	0.84	2.94	4.38	0.89	3.09	4.42	0.90	3.05	4.62	0.87
Baseline* + \mathcal{L}_{ADL}	2.93	3.84	0.94	2.81	3.67	0.96	2.77	3.71	0.96	2.84	3.74	0.95
rPPG-HiBa**	2.56	3.44	0.96	2.38	3.16	0.99	2.41	3.25	0.98	2.45	3.28	0.98

Table 4: Cross-dataset testing on different domains on BUAA. The dataset contains rich "illumination" domain information. We set the samples with 10.0lux and 15.8lux as "normal illumination" (NI), the samples with 25.1lux and 39.8lux as "medium illumination" (MI), and the samples with 63.1 lux and 100 lux as "high illumination" (HI).

Method	head >50			medium ≤50 & ≥23			tail < 23			overall		
	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* [6]	5.76	8.04	0.30	11.24	13.02	0.51	21.35	23.37	0.69	8.97	12.16	0.49
NEST* [4]	4.81(+0.95)	6.50(+1.54)	0.38(+0.08)	9.05(+2.19)	11.58(+1.44)	0.59(+0.08)	19.21(+2.14)	22.12(+1.25)	0.65(-0.04)	7.57(+1.40)	10.87(+1.29)	0.60(+0.11)
LDS** [13]	6.07(-0.31)	8.54(-0.50)	0.28(-0.02)	8.52(+2.72)	11.24(+1.78)	0.61(+0.10)	16.78(+4.57)	20.79(+2.58)	0.71(+0.02)	7.94(+1.03)	11.35(+0.81)	0.56(+0.07)
Baseline* + \mathcal{L}_{CCR}	5.34(+0.42)	7.18(+0.86)	0.31(+0.01)	8.35(+2.89)	10.71(+2.31)	0.64(+0.13)	17.31(+4.04)	20.67(+2.70)	0.73(+0.04)	7.51(+1.46)	10.58(+1.58)	0.60(+0.11)
rPPG-HiBa**	5.15(+0.61)	6.97(+1.07)	0.35(+0.05)	8.33(+2.91)	10.67(+2.35)	0.63(+0.12)	16.95(+4.40)	20.36(+3.01)	0.74(+0.05)	7.34(+1.63)	10.41(+1.75)	0.61(+0.12)

(a) Cross-dataset testing on VIPL

Method	head >14			medium ≤14 & ≥8			tail < 8			overall		
	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
Baseline* [6]	7.23	8.99	0.26	11.82	15.39	0.39	20.03	27.19	0.44	10.16	14.57	0.34
NEST* [4]	6.46(+0.77)	7.18(+1.81)	0.33(+0.07)	12.17(-0.35)	13.81(+1.58)	0.48(+0.09)	20.70(-0.68)	27.22(-0.03)	0.43(-0.01)	9.87(+0.29)	13.53(+1.04)	0.40(+0.06)
LDS** [13]	8.48(-1.25)	9.73(-0.74)	0.23(-0.03)	11.34(+0.48)	13.41(+1.98)	0.48(+0.09)	19.12(+0.9)	24.61(+2.58)	0.51(+0.07)	10.68(-0.52)	13.70(+0.87)	0.34(+0.00)
Baseline* + \mathcal{L}_{CCR}	6.97(+0.26)	7.85(+1.14)	0.30(+0.04)	11.17(+0.65)	12.91(+2.48)	0.50(+0.11)	19.09(+0.93)	24.71(+2.48)	0.52(+0.08)	9.72(+0.44)	12.84(+1.73)	0.40(+0.06)
rPPG-HiBa**	6.90(+0.33)	7.78(+1.21)	0.30(+0.04)	11.29(+0.53)	13.20(+2.19)	0.49(+0.10)	18.98(+1.04)	23.39(+3.80)	0.57(+0.13)	9.69(+0.47)	12.54(+2.03)	0.42(+0.08)

(b) Cross-dataset testing on V4V

Table 5: Cross-dataset testing on VIPL and V4V datasets. Results of different methods on three disjoint subsets. For VIPL, We classify phys-labels with more than 50 samples as Head, fewer than 23 as tail, and the remaining as medium. **Green** characters indicate how the method performance exceeds the baseline. **Red** characters indicate how the method performs inferior to the baseline.

Target	Method	LF-(u.n)			HF-(u.n)			LF/HF			HR-(bpm)		
		MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑	MAE↓	RMSE↓	r↑
UBFC	GREEN [10]	0.2355	0.2841	0.0924	0.2355	0.2841	0.0924	0.6695	0.9512	0.0467	8.0184	9.1776	0.3634
	CHROM [2]	0.2221	0.2817	0.0698	0.2221	0.2817	0.0698	0.6708	1.0542	0.1054	7.2291	8.9224	0.5123
	POS [11]	0.2364	0.2861	0.1359	0.2364	0.2861	0.1359	0.6515	0.9535	0.1345	7.3539	8.0402	0.4923
	NEST [4]	0.0597	0.0782	0.2017	0.0597	0.0782	0.2017	0.2138	0.2824	0.3179	4.7471	6.8876	0.8546
	Baseline [6]	0.0621	0.0813	0.1873	0.0621	0.0813	0.1873	0.1985	0.2667	0.3043	5.1542	7.4672	0.8165
	rPPG-HiBa w/o \mathcal{L}_{CCR}	0.0611	0.0807	0.2375	0.0611	0.0807	0.2375	0.1950	0.2606	0.3309	4.6789	6.8114	0.8542
	rPPG-HiBa w/o \mathcal{L}_{ADL}	0.0594	0.0783	0.2756	0.0594	0.0783	0.2751	0.1706	0.2570	0.3442	4.3048	6.6309	0.8667
	rPPG-HiBa	0.0575	0.0763	0.2955	0.0575	0.0763	0.2955	0.1668	0.2548	0.3489	4.0014	6.6279	0.8762
PURE	GREEN [10]	0.2539	0.3002	0.0326	0.2539	0.3002	0.0326	0.6525	0.8932	0.0417	10.3247	14.2693	0.4952
	CHROM [2]	0.2096	0.2751	0.1059	0.2096	0.2751	0.0759	0.5404	0.8266	0.1173	9.7914	12.7568	0.3732
	POS [11]	0.1959	0.2571	0.1684	0.1959	0.2571	0.1684	0.5373	0.846	0.1433	9.8273	13.4414	0.3432
	NEST [4]	0.0635	0.0874	0.6422	0.0635	0.0874	0.6422	0.2255	0.3505	0.5734	7.6889	10.4783	0.7255
	Baseline [6]	0.0671	0.0923	0.6046	0.0671	0.0923	0.6046	0.2864	0.4184	0.5526	8.2542	11.1765	0.6832
	rPPG-HiBa w/o \mathcal{L}_{CCR}	0.0639	0.0837	0.6562	0.0639	0.0837	0.6562	0.2384	0.3195	0.6561	7.0837	9.8497	0.8225
	rPPG-HiBa w/o \mathcal{L}_{ADL}	0.0592	0.0790	0.6981	0.0592	0.0790	0.6981	0.2138	0.3341	0.6835	6.6516	9.3256	0.8449
	rPPG-HiBa	0.0574	0.0733	0.7329	0.0574	0.0733	0.7329	0.2017	0.2912	0.7397	6.8458	9.3706	0.8460
BUAA	GREEN [10]	0.3472	0.3951	0.0871	0.3472	0.3951	0.0871	0.6453	0.8632	0.0921	5.8231	7.9882	0.5624
	CHROM [2]	0.3786	0.3237	0.0682	0.3786	0.3237	0.0682	0.6813	0.8836	0.0715	6.0934	8.2938	0.5165
	POS [11]	0.3198	0.3762	0.0962	0.3198	0.3762	0.0962	0.6275	0.8424	0.1127	5.0407	7.1198	0.6374
	NEST [4]	0.1436	0.1665	0.2955	0.1436	0.1665	0.2955	0.5514	0.6884	0.3004	3.3723	5.8806	0.7647
	Baseline [6]	0.1451	0.1681	0.2891	0.1451	0.1681	0.2891	0.5564	0.6904	0.2914	3.7852	6.3237	0.7492
	rPPG-HiBa w/o \mathcal{L}_{CCR}	0.1363	0.1634	0.4464	0.1363	0.1634	0.4464	0.5100	0.6589	0.4818	2.6802	3.9052	0.8922
	rPPG-HiBa w/o \mathcal{L}_{ADL}	0.1440	0.1677	0.3972	0.1440	0.1677	0.3972	0.5488	0.6860	0.4385	2.8753	4.0071	0.8670
	rPPG-HiBa	0.1343	0.1610	0.4512	0.1343	0.1610	0.4512	0.5136	0.6508	0.4920	2.6668	3.8898	0.9014

Table 6: HRV and HR estimation results on cross-dataset testing (Two for training and the other one as a testing).

REFERENCES

- [1] Serge Bobbia, Richard Macwan, Yannick Benezeth, Alamin Mansouri, and Julien Dubois. 2019. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognit Lett* 124 (2019), 82–90.
- [2] Gerard De Haan and Vincent Jeanne. 2013. Robust pulse rate from chrominance-based rPPG. *IEEE Trans. Biomed. Eng.* 60, 10 (2013), 2878–2886.
- [3] Hao Lu, Hu Han, and S Kevin Zhou. 2021. Dual-GAN: Joint BVP and Noise Modeling for Remote Physiological Measurement. In *Proc. IEEE CVPR*. 12404–12413.
- [4] Hao Lu, Zitong Yu, Xuesong Niu, and Yingcong Chen. 2023. Neuron Structure Modeling for Generalizable Remote Physiological Measurement. *arXiv preprint arXiv:2303.05955* (2023).
- [5] Xuesong Niu, Hu Han, Shiguang Shan, and Xilin Chen. 2018. VIPL-HR: A Multi-modal Database for Pulse Estimation from Less-constrained Face Video. In *Proc. ACCV*. 562–576.
- [6] Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. 2020. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Trans. on Image Process.* 29 (2020), 2409–2423.
- [7] Ambareesh Revanur, Zhihua Li, Umur A Ciftci, Lijun Yin, and László A Jeni. 2021. The first vision for vitals (v4v) challenge for non-contact video-based physiological estimation. In *Proc. CVPR workshop*. 2760–2767.
- [8] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. 2014. Non-contact video-based pulse rate measurement on a mobile service robot. In *Proc. IEEE ISRHIC*. 1056–1062.
- [9] Weiyu Sun, Xinyu Zhang, Hao Lu, Ying Chen, Yun Ge, Xiaolin Huang, Jie Yuan, and Yingcong Chen. 2023. Resolve Domain Conflicts for Generalizable Remote Physiological Measurement. In *Proceedings of the 31st ACM International Conference on Multimedia*. 8214–8224.
- [10] Wim Verkruysse, Lars O Svaasand, and J Stuart Nelson. 2008. Remote plethysmographic imaging using ambient light. *Opt. Express* 16, 26 (2008), 21434–21445.
- [11] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. 2017. Algorithmic principles of remote PPG. *IEEE Trans. Biomed. Eng.* 64, 7 (2017), 1479–1491.
- [12] Lin Xi, Weihai Chen, Changchen Zhao, Xingming Wu, and Jianhua Wang. 2020. Image enhancement for remote photoplethysmography in a low-light environment. In *FG*. IEEE, 1–7.
- [13] Yuzhe Yang, Kaiwen Zha, Yingcong Chen, Hao Wang, and Dina Katabi. 2021. Delving into deep imbalanced regression. In *ICML*. PMLR, 11842–11851.