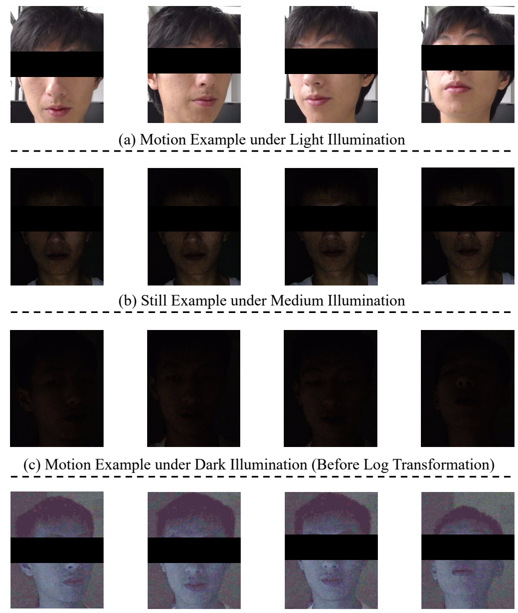
Spatial Alignment and Temporal Matching Adapter for Video-Radar Remote Physiological Measurement

Supplementary Material

Parameters	Value
Start frequency	77GHz
Frequency slope	65MHz/μs
Idle time	10μs
Ramp end time	60μs
Sample points	256
Sample rate	5MHz
Frame periodicity	1.67ms

Table 8. Radar Parameters Settings.



(a) Motion Example under Light Illumination

(b) Still Example under Medium Illumination

(c) Motion Example under Dark Illumination (Before Log Transformation)

(d) Motion Example under Dark Illumination (After Log Transformation)

Figure 4. Examples of the self-collected MMRPM dataset.

6. Details of MMRPM Dataset

RGB videos are recorded with a Logitech C930c camera at a frame rate of 30 fps. Radar data is collected using a Texas Instruments AWR1843BOOST development board at a distance of approximately 0.5 meters. We activate 3 transmitters and 4 receivers of the radar to achieve a virtual 2D antenna array with 12 channels. The detailed configuration of the radar chirp and frame parameters is provided in Tab. 8. Ground-truth PPG signals are acquired using a CONTEC CMS50E sensor at a sampling rate of 60 Hz. During preprocessing, the radar data and PPG signals are downsampled to 120 Hz and 30 Hz respectively.

Examples of the dataset are shown in Fig. 4.

7. Details of Unimodal Pre-training

7.1. Video Pre-training

We utilize the following datasets to pre-train our video UniFormer:

VIPL [31, 32]. This dataset contains 2,378 RGB videos captured by 4 cameras in different scenarios, including various head movements and illumination conditions.

MMPD [40]. It comprises 660 mobile phone videos of subjects with different skin tokens. Similarly to VIPL, it also covers a wide range of lighting conditions and subject activities.

UBFC [1]. This dataset includes 42 uncompressed videos recorded at varying levels of sunlight and indoor illumination.

PURE [39]. It consists of 60 videos of 10 subjects under six different activities, namely steady, talking, slow translation, fast translation, slow rotation, and medium rotation.

BUAA [47]. It records 165 videos under varying illumination levels ranging from 1.0 to 100.0 lux. Since video modality struggles to tackle low-light condition alone, we only use videos with illumination greater than or equal to 6.3 lux for video pre-training.

We train the model on these datasets jointly with a batch size of 8 and utilize the Adam [54] optimizer with a learning rate of 1e-4. Random horizontal flipping, spatially resized crop, random intensity noise, and temporal resample introduced in [33, 50] are used for data augmentation.

7.2. Radar Pre-training

We utilize the dataset introduced in [53] to pre-train our radar UniFormer. This dataset records radar data and synchronized ECG signals from 6,222 subjects. Due to the significant waveform differences between ECG and PPG signals, we transform the ECG signals into Gaussian heartbeat signals following [43] as ground truths during pre-training. The batch size and learning rate are set to 8 and 1e-3 respectively.

8. Additional Results

8.1. Comparison with Unimodal Methods

As shown in Tab. 9, we also compare our method with existing unimodal remote physiological measurement methods on our MMRPM dataset. Notably, our radar Uni-

Methods	Modality	Std↓	MAE↓	RMSE↓	R↑	Trainable Params (M)↓
PhysNet[49]	Video	10.981	4.813	11.280	0.501	0.77
PhysFormer[51]	Video	8.400	2.841	8.473	0.667	7.38
RCG2ECGNet[3]	Radar	6.481	2.879	6.555	0.761	3.61
HRVNet[43]	Radar	6.333	2.363	6.386	0.782	0.86
UniFormer[18]	Video	5.726	1.733	5.726	0.818	22.64
UniFormer	Radar	5.824	1.891	5.845	0.813	22.23
SATM (Ours)	Video-Radar	**5.063**	**1.292**	**5.074**	**0.858**	10.78

Table 9. Additional results of existing unimodal methods on the MMRPM dataset.

Methods	Std↓	MAE↓	RMSE↓	R↑
EquiPleth*	8.787	4.256	9.082	0.629
Fusion-Vital*	8.538	4.457	8.897	0.642
CardiacMamba*	9.048	5.038	9.308	0.603
UniAdapter	8.503	3.930	8.748	0.656
MMA	8.614	4.250	8.821	0.640
BAT	8.455	3.994	8.590	0.647
LAVISH	8.026	3.687	8.232	0.712
SATM (Ours)	**7.542**	**3.423**	**7.656**	**0.738**

Table 10. Detailed results of the cross-dataset evaluation from the EquiPleth dataset to the MMRPM dataset.

Former achieves significant improvements over existing radar-based methods, which can be attributed to its superior spatio-temporal modeling capability. In contrast, existing methods are primarily designed for constrained scenarios and are highly susceptible to subject motions appearing in the MMRPM dataset. Moreover, unimodal UniFormers demonstrate subpar performance due to their sensitivity to light conditions or head motions, while our SATM extracts complementary features for both backbones and showcases superior performance on this challenging dataset.

8.2. Details of Cross-dataset Evaluation

The main differences in radar setup between EquiPleth and MMRPM lie in the number of antennas and the resulting virtual channels. Therefore, for the evaluation from EquiPleth to MMRPM, we only used the MMRPM radar data collected from a pair of antennas for alignment.

Detailed results from EquiPleth to MMRPM are provided in Tab. 10. We also performed the reverse evaluation from MMRPM to EquiPleth, where the virtual channel of the EquiPleth radar data are duplicated to match the number of MMRPM. As shown in Tab. 11, our method consistently outperforms the best baseline, LAVISH.

Methods	Std↓	MAE↓	RMSE↓	R↑
EquiPleth*	2.713	0.722	2.725	0.973
Fusion-Vital*	1.939	0.698	1.944	0.986
CardiacMamba*	3.223	0.929	3.240	0.962
UniAdapter	1.653	0.656	1.657	0.990
MMA	1.843	0.721	1.843	0.987
BAT	2.097	0.696	2.100	0.984
LAVISH	1.630	0.641	1.632	0.990
SATM (Ours)	**1.330**	**0.578**	**1.331**	**0.994**

Table 11. Detailed results of the cross-dataset evaluation from the MMRPM dataset to the EquiPleth dataset.

Methods	Motion	Dark	Dark&Motion
Vision UniFormer	2.087	3.046	3.290
Radar UniFormer	2.595	2.181	2.608
EquiPleth*	1.954	2.111	2.848
Fusion-Vital*	1.786	2.453	2.527
CardiacMamba*	2.334	2.380	2.475
UniAdapter	1.903	2.069	2.431
MMA	1.965	2.107	2.521
BAT	2.377	2.369	2.544
LAVISH	1.762	2.029	2.362
SATM (Ours)	**1.647**	**1.812**	**1.923**

Table 12. MAE results of challenging scenarios in MMRPM.

8.3. Extra results on MMRPM

Tab. 12 shows MAE results on challenging scenarios in MMRPM. It is evident that SATM demonstrates superior performance in both low-light and head-motion conditions.