Supplementary Material of Continual Learning for Remote Physiological Measurement: Minimize Forgetting and Simplify Inference

Qian Liang, Yan Chen, and Yang Hu*

University of Science and Technology of China
qianliang@mail.ustc.edu.cn, {eecyan,eeyhu}@ustc.edu.cn

A Feasibility and Effectiveness of Prefix Finetuning

Since the last two stages of Uniformer [4] are implemented with self attention, it is feasible to utilize prefix [5] to finetune these parts of our backbone. Hence, we perform a transfer learning experiment to evaluate the effectiveness of this strategy for rPPG measurement. Specifically, the backbone is per-trained on the VIPL [7,8] Fold1-4 task and then finetuned on the MMPD [10] task. The results of HR estimation on the target task are presented in Tab. 1.

It can be observed that compared to the pre-trained backbone, prefix finetuning only slightly improves the performance and still significantly lags behind the adapter [1] finetuning. This suggests that prefix finetuning fails to effectively adapt the backbone to the challenging MMPD task. We believe this is because prefix is unable to finetune the convolutional parts of Uniformer and consequently has limited plasticity. Therefore, we argue that prefix is unsuitable to finetune Uniformer in the context of rPPG measurement.

B HRV Estimation

In addition to HR estimation, we also conduct a testing for HRV features on the rPPG DIL protocol, including low-frequency power (LF) in normalized units

* Corresponding author

Table 1: HR estimation results of the transfer evaluation. "Freezing" means we directly evaluate the pre-trained backbone on the MMPD task. The "10, 20, 50, 100" indicate the length of prefixes. The best results are in bold.

Method	Std↓	MAE↓	RMSE↓	R↑
Freezing	13.59	9.06	13.66	0.51
Prefix10	13.35	8.95	13.35	0.53
Prefix20	13.32	8.99	13.32	0.54
Prefix50	13.15	9.01	13.18	0.54
Prefix100	13.36	8.89	13.36	0.53
Adapter	**10.49**	**6.72**	**10.53**	**0.74**

Table 2: LF and HF estimation results on the rPPG DIL protocol. Std_N represents the final incremental Std and the same applies to MAE_N, RMSE_N as well as R_N.

Methods	$\mathrm{Std}_N \downarrow$	$\mathrm{MAE}_N \downarrow$	$\mathrm{RMSE}_N \downarrow$	$\mathrm{R}_N \uparrow$
Upper Bound	0.134±0.003	0.154±0.022	0.189±0.021	0.526±0.021
EWC [3]	0.148±0.010	0.159±0.007	0.195±0.012	0.455±0.054
ANCL-EWC [2]	0.154±0.001	0.171±0.011	0.207±0.014	0.419±0.065
LwF [6]	0.148±0.001	0.183±0.008	0.219±0.008	0.407±0.015
ANCL-LwF [2]	0.148±0.007	0.155±0.004	0.196±0.004	0.407±0.011
S-Prompts [11]	0.147±0.005	0.166±0.003	0.200±0.004	0.458±0.039
ADDP	**0.145**±0.003	**0.150**±0.003	**0.186**±0.005	**0.498**±0.019

Table 3: LF/HF estimation results on the rPPG DIL protocol.

Methods	$\mathrm{Std}_N \downarrow$	$\mathrm{MAE}_N \downarrow$	$\mathrm{RMSE}_N \downarrow$	$\mathrm{R}_N \uparrow$
Upper Bound	1.046±0.058	0.810±0.048	1.267±0.065	0.559±0.028
EWC [3]	1.125±0.044	0.893±0.012	1.366±0.038	0.465±0.074
ANCL-EWC [2]	1.187±0.069	0.953±0.051	1.437±0.105	0.415±0.118
LwF [6]	1.107±0.052	0.959±0.026	1.425±0.046	0.487±0.051
ANCL-LwF [2]	1.098±0.057	0.831±0.015	1.322±0.046	0.492±0.065
S-Prompts [11]	1.134±0.068	0.925±0.015	1.394±0.062	0.466±0.072
ADDP	**1.066**±0.024	**0.826**±0.019	**1.281**±0.030	**0.500**±0.055

(n.u.), high-frequency power (HF) in normalized units (n.u.), and the ratio of LF and HF power (LF/HF). Since the calculation of ground truth HRV features relies on high quality rPPG signal labels, we only perform this evaluation on the PURE, UBFC, BUAA and MMPD tasks.

HRV estimation is a more challenging problem that requires accurately locating systolic peaks in rPPG signals. Consequently, only the methods that can provide relatively accurate HR estimation are evaluated and the results are shown in Tab. 2 and Tab. 3.[1] We observe that our method (ADDP) surpasses other baselines once again and significantly reduces the performance gap with the upper bound, further highlighting the superiority of our method.

C Sensitivity Analysis on Hyperparameters

We conduct a sensitivity analysis on the selection of hyperparameters. Specifically, we perform an evaluation on the rPPG DIL protocol while keeping one hyperparameter varied and fixing the remained hyperparameters constant. As presented in Tab. 4, the selection of the number of domain prototypes K and the probability for augmentation p has only a minor effect on the performance, indicating that our method is robust to the values of these two hyperparameters. For the sake of simplicity, we choose $K = 8$ and $p = 0.5$ as the default

[1] The results of LF and HF estimation are identical.

Table 4: HR estimation results of the sensitivity analysis on the rPPG DIL protocol.

Hyperparameters	$\text{Std}_N \downarrow$	$\text{MAE}_N \downarrow$	$\text{RMSE}_N \downarrow$	$\text{R}_N \uparrow$
$K=8, p=0.5, \alpha=9$	5.56±0.04	3.70±0.07	5.59±0.02	0.83±0.01
$K=6, p=0.5, \alpha=9$	5.52±0.02	3.70±0.08	5.65±0.03	**0.84**±0.01
$K=7, p=0.5, \alpha=9$	**5.50**±0.07	**3.67**±0.04	**5.56**±0.05	**0.84**±0.00
$K=9, p=0.5, \alpha=9$	5.65±0.13	3.71±0.17	5.72±0.17	**0.84**±0.02
$K=10, p=0.5, \alpha=9$	5.62±0.23	3.73±0.07	5.74±0.20	0.83±0.01
$K=8, p=0.3, \alpha=9$	5.64±0.14	3.69±0.09	5.71±0.10	**0.84**±0.01
$K=8, p=0.4, \alpha=9$	5.60±0.04	3.70±0.11	5.66±0.08	**0.84**±0.01
$K=8, p=0.6, \alpha=9$	5.57±0.08	**3.67**±0.02	5.61±0.05	**0.84**±0.00
$K=8, p=0.7, \alpha=9$	5.57±0.07	3.73±0.08	5.67±0.10	0.83±0.01
$K=8, p=0.5, \alpha=7$	5.52±0.02	3.70±0.08	5.65±0.03	**0.84**±0.01
$K=8, p=0.5, \alpha=8$	5.51±0.15	3.73±0.07	5.62±0.12	**0.84**±0.01
$K=8, p=0.5, \alpha=10$	5.59±0.19	3.72±0.09	5.69±0.14	**0.84**±0.01
$K=8, p=0.5, \alpha=11$	5.58±0.07	3.74±0.10	5.70±0.15	**0.84**±0.00

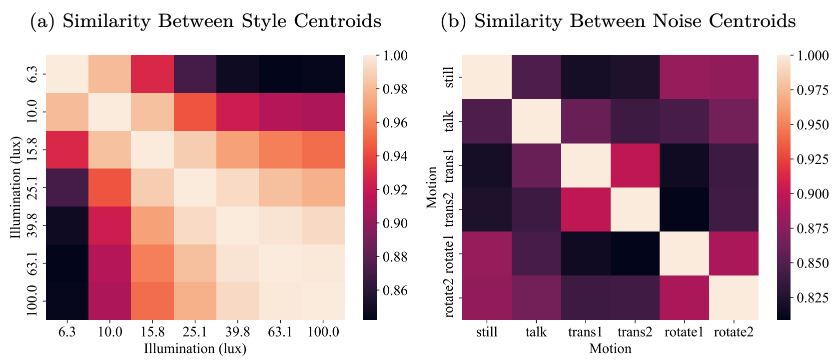

Fig. 1: Pairwise cosine similarity between feature centroids. "Trans1", "trans2", "rotate1" and "rotate2" denote "slow translation", "fast translation", "small rotation" and "medium rotation" respectively.

setting, as they provide satisfactory performance without significantly impacting the results.

We also investigate the influence of the value of α for the noise extraction mask (*i.e.*, the number of "0" in $\mathbf{M}$). As shown in Tab. 4, the performance slightly decreases with variations in α. This makes sense as α directly affects the representation capacity of the noise features. Choosing a large α implies only extracting the feature components with small singular values, which reduces the information contained in the noise features; while selecting a small α may result in the failure to disentangle the rPPG information effectively.

D Further Analysis on Domain Features

As stated in the main body of our paper, BUAA [12] and PURE [9] datasets provide samples with illumination and motion labels. We utilize these two datasets to verify whether our domain features, namely style features and noise features, extract the intended information effectively. Specifically, we calculate the style or noise feature centroids of samples sharing the same illumination or motion labels and measure their pairwise cosine similarity.

The results in Fig. 1 show that the cosine similarity between style centroids decreases as the illumination difference increases. Similarly, the similarity between noise centroids with similar motion labels, such as "small rotation" and "medium rotation", is higher than the one between the noise centroid of "small rotation" and the centroid of "fast translation". This confirms that our style and noise features are capable of capturing the desired domain factors, including but not limited to illumination and motion.

E Responsibility to Human Subjects

Datasets used in our experiment involve human subjects. We have obtained the usage approval from the publishers of these datasets and adhered to their prescribed usage protocols.

References

1. Houlsby, N., Giurgiu, A., Jastrzebski, S., Morrone, B., De Laroussilhe, Q., Gesmundo, A., Attariyan, M., Gelly, S.: Parameter-efficient transfer learning for nlp. In: International Conference on Machine Learning. pp. 2790–2799. PMLR (2019)
2. Kim, S., Noci, L., Orvieto, A., Hofmann, T.: Achieving a better stability-plasticity trade-off via auxiliary networks in continual learning. In: Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition. pp. 11930–11939 (2023)
3. Kirkpatrick, J., Pascanu, R., Rabinowitz, N., Veness, J., Desjardins, G., Rusu, A.A., Milan, K., Quan, J., Ramalho, T., Grabska-Barwinska, A., et al.: Overcoming catastrophic forgetting in neural networks. Proceedings of the national academy of sciences **114**(13), 3521–3526 (2017)
4. Li, K., Wang, Y., Gao, P., Song, G., Liu, Y., Li, H., Qiao, Y.: Uniformer: Unified transformer for efficient spatiotemporal representation learning. arXiv preprint arXiv:2201.04676 (2022)
5. Li, X.L., Liang, P.: Prefix-tuning: Optimizing continuous prompts for generation. arXiv preprint arXiv:2101.00190 (2021)
6. Li, Z., Hoiem, D.: Learning without forgetting. IEEE transactions on pattern analysis and machine intelligence **40**(12), 2935–2947 (2017)
7. Niu, X., Han, H., Shan, S., Chen, X.: Vipl-hr: A multi-modal database for pulse estimation from less-constrained face video. In: Computer Vision–ACCV 2018: 14th Asian Conference on Computer Vision, Perth, Australia, December 2–6, 2018, Revised Selected Papers, Part V 14. pp. 562–576. Springer (2019)

8. Niu, X., Shan, S., Han, H., Chen, X.: Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. IEEE Transactions on Image Processing **29**, 2409–2423 (2019)
9. Stricker, R., Müller, S., Gross, H.M.: Non-contact video-based pulse rate measurement on a mobile service robot. In: The 23rd IEEE International Symposium on Robot and Human Interactive Communication. pp. 1056–1062. IEEE (2014)
10. Tang, J., Chen, K., Wang, Y., Shi, Y., Patel, S., McDuff, D., Liu, X.: Mmpd: Multi-domain mobile video physiology dataset. arXiv preprint arXiv:2302.03840 (2023)
11. Wang, Y., Huang, Z., Hong, X.: S-prompts learning with pre-trained transformers: An occam's razor for domain incremental learning. Advances in Neural Information Processing Systems **35**, 5682–5695 (2022)
12. Xi, L., Chen, W., Zhao, C., Wu, X., Wang, J.: Image enhancement for remote photoplethysmography in a low-light environment. In: 2020 15th IEEE International Conference on Automatic Face and Gesture Recognition (FG 2020). pp. 1–7. IEEE (2020)