
rPPG-Toolbox: Deep Remote PPG Toolbox

Xin Liu¹, Girish Narayanswamy^{1*}, Akshay Paruchuri^{2*}, Xiaoyu Zhang³, Jiankai Tang³,
Yuzhe Zhang³, Roni Sengupta², Shwetak Patel¹, Yuntao Wang³, Daniel McDuff¹

University of Washington Seattle ¹

University of North Carolina at Chapel Hill ²

Tsinghua University ³

{xliu0, girishvn, dmcdff}@cs.washington.edu, akshay@cs.unc.edu

* Equal Contribution

Abstract

Camera-based physiological measurement is a fast growing field of computer vision. Remote photoplethysmography (rPPG) utilizes imaging devices (e.g., cameras) to measure the peripheral blood volume pulse (BVP), and enables cardiac measurement via webcams and smartphones. However, the task is non-trivial with important pre-processing, modeling, and post-processing steps required to obtain state-of-the-art results. Replication of results and benchmarking of new models is critical for scientific progress; however, as with many other applications of deep learning, reliable codebases are not easy to find or use. We present a comprehensive toolbox, rPPG-Toolbox, that contains unsupervised and supervised rPPG models with support for public benchmark datasets, data augmentation, and systematic evaluation: <https://github.com/ubicomplab/rPPG-Toolbox>

1 Introduction

The vision of ubiquitous computing is to embed computation into everyday objects to enable them to perform useful tasks. The sensing of physiological vital signs is one such task and plays an important role in how health is understood and managed. Cameras are both ubiquitous and versatile sensors, and the transformation of cameras into accurate health sensors has the potential to make the measurement of health signals more comfortable and accessible. Examples of the applications of this technology include systems for monitoring neonates [1], dialysis patients [2], and the detection of arrhythmias [3].

Building on advances in computer vision, camera-based measurement of physiological vitals signs has developed into a research field of its own [?]. Researchers have developed methods for measuring cardiac and pulmonary signals by analyzing skin pixel changes over time. Recently, several companies have been granted FDA De Novo status for products that use software algorithms to analyze video and estimate pulse rate, heart rate, respiratory rate and/or breathing rate¹².

There are hundreds of computational architectures that have been proposed for the measurement of cardiopulmonary signals. Unsupervised signal processing methods leverage techniques such as Independent Component Analysis (ICA) or Principal Component Analysis (PCA) and assumptions about the periodicity or structure of the underlying blood volume pulse waveform. Neural network architectures can be trained in a supervised fashion using videos with synchronized gold-standard ground truth signals [4, 5, 6, 7]. Innovative data generation [8] and augmentation [9], meta-learning for personalization [10, 11], federated learning [12], and unsupervised pretraining [13, 14, 15, 16] have been widely explored in the field of camera-based physiological sensing and have led

¹https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN200019.pdf

²https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN200038.pdf

rPPG-PLAN: 深度远程PPG PLAN

刘新、Girish Narayanswamy、Akshay Paruchuri、张晓宇、唐建凯、张裕哲、Roni Sengupta、Shwetak Patel、王云涛、Daniel McDuff
华盛顿大学西雅图北卡罗来纳大学教堂山分校
清华大学

{xlio0, girishvn, dmcduff}@cs.washington.edu, akshay@cs.unc.edu
等额的贡献

摘要

基于摄像机的生理测量是计算机视觉的一个快速发展的领域。远程光电体积描记(rPPG)利用成像设备(例如,通过摄像头测量外周血容量脉搏(BVP),并通过网络摄像头和智能手机进行心脏测量。然而,这项任务是不平凡的,需要重要的预处理,建模和后处理步骤,以获得最先进的结果。结果的复制和新模型的基准测试对于科学进步至关重要;然而,与深度学习的许多其他应用一样,可靠的代码库不容易找到或使用。我们提出了一个全面的工具箱,rPPG-RNN,其中包含无监督和监督的rPPG模型,支持公共基准数据集,数据增强和系统性

evaluation: <https://github.com/ubicomplab/rPPG-Toolbox>

1 介绍

普适计算的愿景是将计算嵌入到日常物品中,使它们能够执行有用的任务。生理生命体征的感测就是这样一项任务,并且在如何理解和管理健康方面发挥着重要作用。摄像头是无处不在的多功能传感器,将摄像头转变为精确的健康传感器有可能使健康信号的测量更加舒适和方便。该技术的应用示例包括用于监测新生儿[1]、透析患者[2]和心律失常检测的系统[3]。

基于计算机视觉的进步,基于相机的生理生命体征测量已经发展成为一个自己的研究领域[?]。研究人员已经开发出通过分析皮肤像素随时间的变化来测量心脏和肺信号的方法。最近,几家公司已被授予FDA De Novo状态,用于使用软件算法分析视频和估计脉搏率,心率,呼吸率和/或呼吸率的产品。

已经提出了数百种用于测量心肺信号的计算架构。无监督信号处理方法利用诸如独立分量分析(伊卡)或主分量分析(PCA)的技术以及关于基础血容量脉搏波形的周期性或结构的假设。神经网络架构可以使用具有同步黄金标准地面真实信号的视频以监督方式进行训练[4, 5, 6, 7]。创新的数据生成[8]和增强[9],用于个性化的元学习[10, 11],联邦学习[12]和无监督预训练[13, 14, 15, 16]已经在基于相机的生理传感领域得到了广泛的探索,并导致了

¹https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN200019.pdf

²https://www.accessdata.fda.gov/cdrh_docs/reviews/DEN200038.pdf

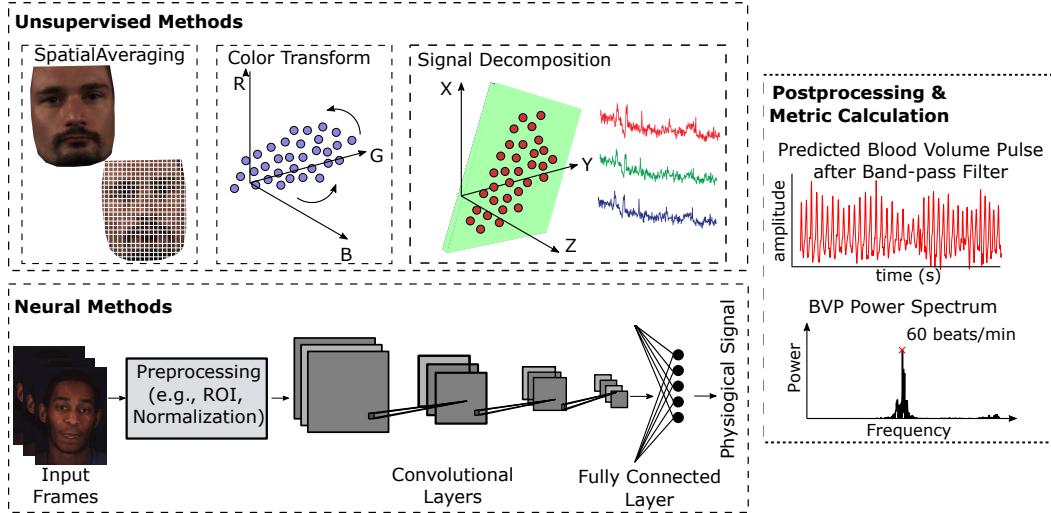


Figure 1: **rPPG Pipeline.** An example of the components of an rPPG pipeline including preprocessing, training, inference, and evaluation.

to significant improvements in state-of-the-art performance. Further information regarding the background, algorithms, and potential applications of rPPG are included in the Appendix-B and C.

However, standardization in the field is still severely lacking. Based on our review of literature in the space, we identified four issues that have hindered the interpretation of results in many papers. First, and perhaps most obviously, a number of the published works are not accompanied by public code. While publishing code repositories with papers is now fairly common in the machine learning and computer vision research communities, it is far less common in the field of camera-based physiological sensing. While there are reasons that it might be difficult to release datasets (e.g., medical data privacy), we cannot find good arguments for not releasing code. Second, many papers do not compare to previously published methods in an “apples-to-apples” fashion. This point is a little more subtle, but rather than performing systematic side-by-side comparisons between methods, the papers compare to numerical results from previous work, even if the training sets and/or test sets are not identical (e.g., test samples were filtered because they were deemed to not have reliable labels). Unfortunately, this often makes it unclear if performance differences are due to data, pre-processing steps, model design, post-processing, training schemes and hardware specifications, or a combination of the aforementioned. Continuing this thread, the third flaw is that papers use pre- and post-processing steps that are not adequately described. Finally, different researchers compute the “labels” (e.g., heart rate) using their own methods from the contact PPG or ECG time-series data. Differences in these methods lead to different labels and a fundamental issue when it comes to benchmarking performance. When combined, the aforementioned issues make it very difficult to draw conclusions from the literature about the optimal choices for the design of rPPG systems.

Open source codes allow researchers to compare novel approaches to consistent baselines without ambiguity regarding the implementation or parameters used. This transparency is important as subsequent research invariably builds on prior state-of-the-art. Implementing a prior method from a paper, even if clearly written, can be difficult. Furthermore, it is an inefficient use of time for many researcher to re-implement all baseline methods. In an effort to address this, several open source toolboxes have been released for camera-based physiological sensing. These toolboxes have been a significant contribution to the community and provide implementations of methods and models [17, 18, 19]; however, they are also incomplete. McDuff and Blackford [17]³ implemented a set of source separation methods (Green, ICA, CHROM, POS) and Pilz [19] published the PPGI-Toolbox⁴ containing implementations of Green, SSR, POS, Local Group Invariance (LGI), Diffusion Process (DP) and Riemannian-PPGI (SPH) models. These toolboxes are implemented in MATLAB (e.g., [17]) and do not contain examples of supervised methods. Python and supervised neural models are now the focus of a large majority of computer vision and deep learning research. There are

³<https://github.com/danmcduff/iphys-toolbox>

⁴<https://github.com/partofthestars/PPGI-Toolbox>

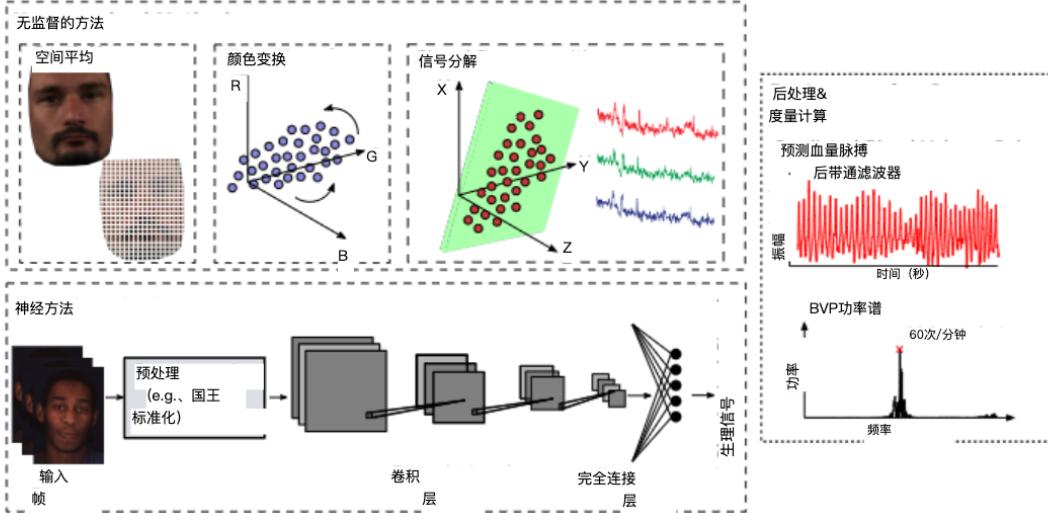


图1: rPPG管道。rPPG管道组件的示例，包括预处理、训练、推理和评估。

到最先进性能的显著改进。关于rPPG的背景、算法和潜在应用的更多信息，请参见EPDix-B和C。

然而，该领域的标准化仍然严重缺乏。根据我们对该领域文献的回顾，我们确定了四个问题，这些问题阻碍了许多论文中结果的解释。首先，也许是最重要的，许多已出版的作品没有附带公共代码。虽然在机器学习和计算机视觉研究社区中发布代码库和论文现在相当普遍，但在基于相机的生理传感领域却不太常见。虽然有一些原因可能很难发布数据集（例如，医疗数据隐私），我们找不到不发布代码的好理由。其次，许多论文没有以“苹果对苹果”的方式与以前发表的方法进行比较。这一点有点微妙，但论文并没有在方法之间进行系统的并排比较，而是与以前工作的数值结果进行比较，即使训练集和/或测试集不相同（例如，测试样品被过滤，因为它们被认为没有可靠的标签）。不幸的是，这通常使人不清楚性能差异是由于数据、预处理步骤、模型设计、后处理、训练方案和硬件规格，还是上述因素的组合。继续这条线索，第三个缺陷是论文使用了没有充分描述的预处理和后处理步骤。最后，不同的研究人员计算“标签”（例如，心率）使用他们自己的方法从接触PPG或ECG时间序列数据中获得。这些方法的差异导致了不同的标签和一个基本的问题，当谈到基准性能。当结合时，上述问题使得很难从文献中得出关于rPPG系统设计的最佳选择的结论。

开放源代码允许研究人员将新方法与一致的基线进行比较，而不会对所使用的实施或参数产生歧义。这种透明度是很重要的，因为后续的研究总是建立在现有的最先进的基础上。从论文中实施现有的方法，即使写得很清楚，也是很困难的。此外，对于许多研究人员来说，重新实现所有基线方法是对时间的低效利用。为了解决这个问题，已经发布了几个开源工具箱用于基于相机的生理感测。这些工具箱对社区做出了重大贡献，并提供了方法和模型的实现[17, 18, 19]；然而，它们也是不完整的。

McDuff和布莱克福德[17]实现了一组源分离方法（绿色、伊卡、CHROM、POS），Pilz [19]发布了包含绿色、SSR、POS、局部群不变性 (LGI)、扩散过程 (DP) 和黎曼PPGI (SPH) 模型实现的PPGI工具箱。这些工具箱在MATLAB中实现（例如，[17]），并且不包含监督方法的示例。Python和监督神经模型现在是大多数计算机视觉和深度学习研究的焦点。有

³<https://github.com/danmcduff/iphys-toolbox>

⁴<https://github.com/partofthestars/PPGI-Toolbox>

Table 1: **Comparison of rPPG Toolboxes.** Comparison of rPPG-Toolbox with existing toolboxes in camera-based physiological measurement.

Toolbox	Dataset Support	Unsup. Eval	DNN Training	DNN Eval
iPhys-Toolbox [20]	✗	✓	✗	✗
PPG-I Toolbox [19]	✗	✓	✗	✗
pyVHR [18, 21]	✓	✓	✗	✓
rPPG-Toolbox (Ours)	✓	✓	✓	✓

Unsup. = Unsupervised learning methods, DNN = Deep neural network methods.

several implementations of popular signal processing methods in Python: Bob.rrpg.base⁵ includes implementations of CHROM, SSR and Boccignone et al. [18] released code for Green, CHROM, ICA, LGI, PBV, PCA, and POS. Several published papers have included links to code; however, often this is only inference code and not training code for neural models. Without providing training code for neural networks, it is challenging for researchers to conduct end-to-end reproducible experiments and build on existing research.

In this paper, we present an end-to-end toolbox⁶ for camera-based physiological measurement. This toolbox includes: 1) support for six public datasets, 2) pre-processing code to format the datasets for training neural models, 3) implementations of six neural model architectures and six unsupervised learning methods, 4) evaluation and inference pipelines for supervised and unsupervised learning methods for reproducibility and 5) enabling advanced neural training and inference such as weakly supervised pseudo labels, motion augmentation and multitask learning. We use this toolbox to publish clear and reproducible benchmarks that we hope will provide a foundation for the community to compare methods in a more rigorous and informative manner.

2 Related Work

In the field of remote PPG sensing, there are three significant open-source toolboxes (documented in Table 1):

iPhys-Toolbox [17]: An open-sourced toolbox written in MATLAB that is comprised of implementations of numerous algorithms for rPPG sensing. It empowers researchers to present results on their datasets using public, standard implementations of baseline methods, ensuring transparency of parameters. This toolbox incorporates a wide range of widely employed baseline methods; however, it falls short on Python support, public dataset data loaders, and neural network training and evaluation.

PPG-I Toolbox [19]: This toolbox provides MATLAB implementations, specifically for six unsupervised signal separation models. It incorporates four evaluation metrics, including Pearson correlation, RMSE/MSE, SNR, and Bland-Altman plots. However, similar to the iPhys-Toolbox, it lacks support for public dataset data loading and neural network training and evaluation.

pyVHR [21]: The most recent in the field, this toolbox adopts Python instead of MATLAB. While it offers ample support for numerous unsupervised methods, its capabilities are limited when it comes to modern neural networks. Notably, pyVHR supports only two neural networks for inference, and none for model training. This omission can be a roadblock for researchers aiming to reproduce and further advance state-of-the-art neural methods.

3 The rPPG-Toolbox

To address the gaps in the current tooling and to promote reproducibility and clearer benchmarking within the camera-based physiological measurement (rPPG) community, we present an open-source toolbox designed to support six public datasets, six unsupervised methods and six neural methods for data preprocessing, neural model training and evaluation, and further analysis.

3.1 Datasets

The toolbox includes pre-processing code that converts six public datasets into a form amenable for training with neural models. The standard form for the videos we select includes raw frames and

⁵<https://pypi.org/project/bob.rrpg.base/>

⁶<https://github.com/ubicomplab/rPPG-Toolbox>

表1：rPPG工具箱的比较。rPPG-100与基于摄像机的生理测量中现有工具箱的比较。

	数据集支持不支持。[20]第20话我的世界	[19]第十九话	PyVHR [18, 21]	rPPG-100 (我们的)	—	—	—
	X	✓	✓	✓	X	X	X
[19]第十九话	X	✓	✓	✓	X	X	X
PyVHR [18, 21]	✓	✓	✓	✓	X	✓	✓
rPPG-100 (我们的)	✓	✓	✓	✓	✓	✓	✓

—不好=无监督学习方法，DNN=深度神经网络方法。

Python中流行的信号处理方法的几种实现：Bob.rppg. base包括CHROM、SSR和Boccignone等的实现。[18]发布了绿色、CHROM、伊卡、LGI、PBV、PCA和POS的代码。一些已发表的论文包含了代码的链接；然而，这通常只是推理代码，而不是神经模型的训练代码。如果不提供神经网络的训练代码，研究人员就很难进行端到端的可重复实验，并在现有研究的基础上进行研究。

在本文中，我们提出了一个端到端的工具箱基于摄像头的生理测量。该工具箱包括：1) 支持六个公共数据集，2) 预处理代码以格式化用于训练神经模型的数据集，3) 实现六种神经模型架构和六种无监督学习方法，4) 用于监督和无监督学习方法的评估和推理管道，以实现可重复性，以及5) 实现高级神经训练和推理，例如弱监督伪标签，运动增强和多任务学习。我们使用这个工具箱来发布清晰和可复制的基准，我们希望这些基准将为社区提供一个基础，以更严格和更翔实的方式比较方法。

2 相关工作

在远程PPG感测领域中，存在三个重要的开源工具箱（记录在表1中）：

iPhys-Pro [17]：一个用MATLAB编写的开源工具箱，由许多rPPG传感算法的实现组成。它使研究人员能够使用基线方法的公共标准实现在其数据集上呈现结果，确保参数的透明度。这个工具箱包含了广泛使用的基线方法；然而，它在Python支持、公共数据集数据加载器以及神经网络训练和评估方面存在福尔斯不足。

PPG-I工具箱[19]：该工具箱提供MATLAB实现，专门用于六个无监督信号分离模型。它集成了四个评估指标，包括Pearson相关性，RMSE/MSE，SNR和Bland–Altman图。然而，与iPhys-EXP类似，它缺乏对公共数据集数据加载和神经网络训练和支持。

pyVHR [21]：该领域的最新工具箱采用Python而不是MATLAB。虽然它为许多无监督方法提供了足够的支持，但在现代神经网络方面，它的能力有限。值得注意的是，pyVHR只支持两个神经网络进行推理，而不支持模型训练。这种遗漏可能会成为旨在复制和进一步推进最先进神经方法的研究人员的障碍。

3 rPPG-RP-R

为了解决当前工具中的差距，并促进基于相机的生理测量（rPPG）社区中的可重复性和更清晰的基准测试，我们提出了一个开源工具箱，旨在支持六个公共数据集，六种无监督方法和六种神经方法，用于数据预处理，神经模型训练和评估以及进一步分析。

3.1 数据集

该工具箱包括预处理代码，可将六个公共数据集转换为适合神经模型训练的形式。我们选择的视频的标准形式包括原始帧和

⁵<https://pypi.org/project/bob.rppg.base/>

⁶<https://github.com/ubicomplab/rPPG-Toolbox>

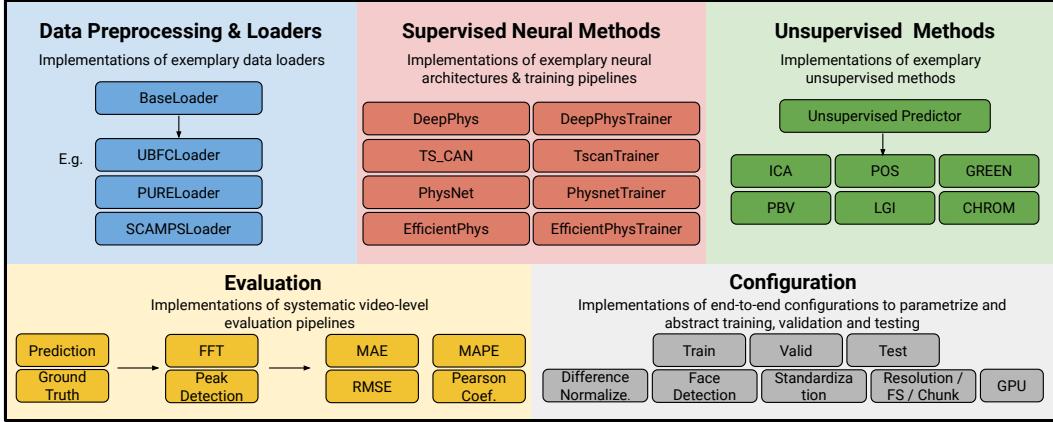


Figure 2: **Overview.** An overview of the rPPG-Toolbox codebase.

difference frames (the difference between each pair of consecutive frames) stored as numpy arrays in a [N, W, H, C] format. Where N is the length of the sequence, W is the width of the frames, H is the height of the frames, and C is the number of channels. There are six channels in this case, as both the raw frames and difference frames account for three color channels each. For faster data loading, all videos in the datasets are typically broken up into several “chunks” of non-overlapping N (e.g., 180) frame sequences. All of these parameters (N, W, H, C) are easy to change and customize. The PPG waveform labels are stored as numpy arrays in a [N, 1] format. The entire pre-processing procedure is supported with multi-thread processing to accelerate the data processing time.

We have provided pre-processing code for UBFC-rPPG [22], PURE [23] SCAMPS [24], MMPD [25], BP4D+ [26], and UBFC-Phys [27]. Each of these datasets encompasses a diverse array of real-world conditions, capturing variations in factors such as motion, lighting, skin tones/types, and backgrounds, thus presenting robust challenges for any signal processing and machine learning algorithm. Tools (python notebooks) are provided for quickly visualizing pre-processed datasets and will be detailed further in Appendix-J. We also support the pre-processing and usage of augmented versions of the UBFC-rPPG [22] and PURE [23] datasets, a feature which we describe further in Section 4.2.

UBFC-rPPG [22]: This dataset features RGB videos recorded using a Logitech C920 HD Pro webcam at 30Hz. The videos have a resolution of 640x480, and they are stored in an uncompressed 8-bit RGB format. Reference PPG data was obtained using a CMS50E transmissive pulse oximeter, thereby providing the gold-standard validation data. The subjects were positioned approximately one meter away from the camera during the recording sessions. The videos were captured under indoor conditions with a combination of natural sunlight and artificial illumination.

PURE [23]: This dataset consists of recordings from 10 subjects, including 8 males and 2 females. The video footage was captured with an RGB eco274CVGE camera from SVS-Vistek GmbH, with a frequency of 30Hz and a resolution of 640x480. Subjects were positioned approximately 1.1 meters from the camera and were illuminated from the front by ambient natural light filtering through a window. The gold-standard ground truth of PPG and SpO₂ were obtained at 60Hz with a CMS50E pulse oximeter affixed to the subject’s finger. Each participant completed six recordings under varied motion conditions, thereby contributing to a range of data reflecting different physical states.

SCAMPS [24]: This dataset encompasses 2,800 video clips, comprising 1.68M frames, featuring synthetic avatars in alignment with cardiac and respiratory signals. These waveforms and videos were generated by employing a sophisticated facial processing pipeline, resulting in high-fidelity, quasi-photorealistic renderings. To provide robust test conditions, the videos incorporate various confounders such as head motions, facial expressions, and changes in ambient illumination.

MMPD [25]: This dataset includes 660 one-minute videos recorded using a Samsung Galaxy S22 Ultra mobile phone, at 30 frames per second with a resolution of 1280x720 pixels and then compressed to 320x240 pixels. The ground truth PPG signals were simultaneously captured using an HKG-07C+ oximeter, at 200 Hz and then downsampled to 30 Hz. It contains Fitzpatrick skin types 3-6, four different lighting conditions (LED-low, LED-high, incandescent, natural), four distinct

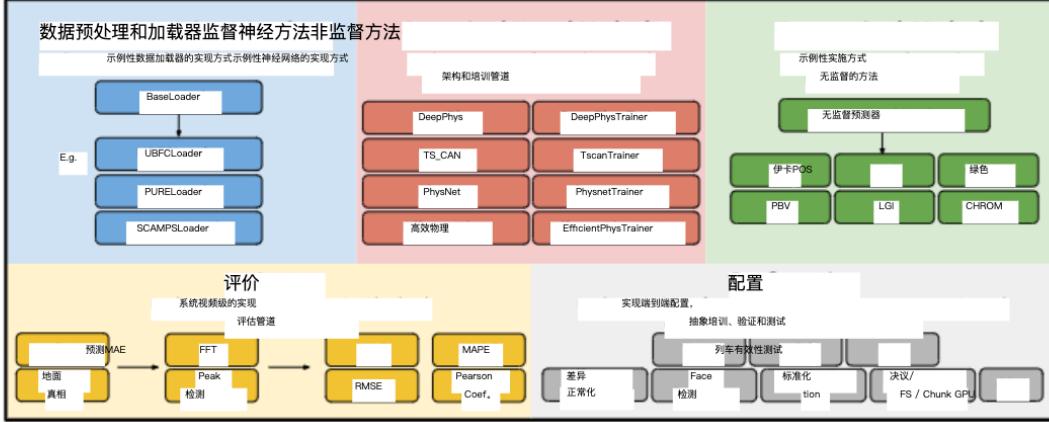


图2：概述。rPPG-P2P代码库的概述。

差异帧（每对连续帧之间的差异）以[N, W, H, C]格式存储为numpy数组。其中N是序列的长度，W是帧的宽度，H是帧的高度，C是通道的数量。在这种情况下，有六个通道，因为原始帧和差异帧各自占三个颜色通道。为了更快的数据加载，数据集中的所有视频通常被分解成几个不重叠的N个“块”（例如，180）帧序列。所有这些参数（N, W, H, C）都很容易更改和自定义。PPG波形标签以[N, 1]格式存储为numpy数组。整个预处理过程支持多线程处理，加快数据处理时间。

我们提供了UBFC-rPPG [22]、PURE [23]、SCAMPS [24]、MMPD [25]、BP 4D + [26]和UBFC-Phys [27]的预处理代码。这些数据集中的每一个都包含各种各样的真实世界条件，捕捉运动、照明、肤色/类型和背景等因素的变化，从而为任何信号处理和机器学习算法提出了强大的挑战。工具（python notebook）用于快速可视化预处理的数据集，并将在Wendix-J中进一步详细说明。我们还支持UBFC-rPPG [22]和PURE [23]数据集的增强版本的预处理和使用，我们将在第4.2节中进一步描述该功能。

UBFC-rPPG [22]: 该数据集采用Logitech C920 HD Pro网络摄像头以30 Hz录制的RGB视频。视频分辨率为640 x480，以未压缩的8位RGB格式存储。使用CMS 50 E透射式脉搏血氧仪获得参考PPG数据，从而提供金标准验证数据。在记录过程中，受试者被定位在距离摄像机大约一米的地方。这些视频是在自然阳光和人工照明相结合的室内条件下拍摄的。

PURE [23]: 该数据集由10名受试者的记录组成，包括8名男性和2名女性。视频片段是用SVS-Vistek GmbH的RGB eco 274 CVGE相机拍摄的，频率为30 Hz，分辨率为640 x480。受试者被定位在距离相机约1.1米处，并通过窗户过滤的环境自然光从前面照射。PPG和SpO 2的金标准真实值是在60 Hz下用固定在受试者手指上的CMS 50 E脉搏血氧仪获得的。每个参与者在不同的运动条件下完成了六次记录，从而有助于反映不同身体状态的一系列数据。

SCAMPS [24]: 该数据集包含2, 800个视频剪辑，包括168万帧，具有与心脏和呼吸信号对齐的合成化身。这些波形和视频是通过采用复杂的面部处理管道生成的，从而产生高保真、准真实感的渲染。为了提供稳健的测试条件，视频包含各种混杂因素，例如头部运动、面部表情和环境照明的变化。

MMPD [25]: 该数据集包括使用三星Galaxy S22 Ultra移动的手机以每秒30帧的速度录制的660个一分钟视频，分辨率为1280 x720像素，然后压缩到320 x240像素。使用HKG-07 C+血氧计在200 Hz下同时捕获真实PPG信号，然后下采样至30 Hz。它包含菲茨帕特里克皮肤类型3–6，四种不同的照明条件（LED低，LED高，白炽灯，自然），四种不同的照明条件，

activities (stationary, head rotation, talking, and walking), and exercise scenarios. With multiple labels provided, different subsets of this dataset can be easily used for research using our toolbox.

BP4D+ [26]: This dataset contains video footage captured at a rate of 25 frames per second, for 140 subjects, each participating in 10 emotion-inducing tasks, amounting to a total of 1400 trials and associated videos. In addition to the standard video footage, the dataset also includes 3D mesh models and thermal video, both captured at the same frame rate. Alongside these, the dataset offers supplementary data including blood pressure measurements (wave, systolic, diastolic, mean), heart rate in beats per minute, respiration (wave, rate bpm), electrodermal activity, and Facial Action Coding System (FACS) encodings for specified action units.

UBFC-Phys [27]: The UBFC-PHYS dataset, a multi-modal dataset, contains 168 RGB videos, with 56 subjects (46 women and 10 men) per a task. There are three tasks with significant amounts of unconstrained motion under static lighting conditions - a rest task, a speech task, and an arithmetic task. The dataset contains gold-standard blood volume pulse (BVP) and electrodermal activity (EDA) measurements that were collected via the Empatica E4 wristband. The videos were recorded at a resolution of 1024x1024 and 35Hz with a EO-23121C RGB digital camera. We utilized all three tasks and the same subject sub-selection list provided by the authors of the dataset in the second supplementary material of Sabour et al. [27] for evaluation. We reiterate this subject sub-selection list in Appendix-H.

3.2 Methods

3.2.1 Unsupervised Methods

The following methods all use linear algebra and traditional signal processing to recover the estimated PPG signal: 1) **Green** [28]: the green channel information is used as the proxy for the PPG after spatial averaging of RGB video; 2) **ICA** [29]: Independent Component Analysis (ICA) is applied to normalized, spatially averaged color signals to recover demixing matrices; 3) **CHROM** [30]: a linear combination of the chrominance signals obtained from the RGB video are used for estimation; 4) **POS** [31]: plane-orthogonal-to-the-skin (POS), is a method that calculates a projection plane orthogonal to the skin-tone based on physiological and optical principles. A fixed matrix projection is applied to the spatially normalized, averaged pixel values, which are used to recover the PPG waveform; 5) **PBV** [32]: a signature, that is determined by a given light spectrum and changes of the blood volume pulse, is used in order to derive the PPG waveform while offsetting motion and other noise in RGB videos; 6) **LGI** [33]: a feature representation method that is invariant to motion through differentiable local transformations.

3.2.2 Supervised Neural Methods

The following implementations of supervised learning algorithms are included in the toolbox. All implementations were done using PyTorch [37]. Common optimization algorithms, such as Adam [38] and AdamW [39], and criterion, such as mean squared error (MSE) loss, are utilized for training except for where noted. The learning rate scheduler typically follows the 1cycle policy [40], which anneals the learning rate from an initial learning rate to some maximum learning rate and then, from that maximum learning rate, to some learning rate much lower than the initial learning rate. The total steps in this policy are determined by the number of epochs multiplied by the number of training batches in an epoch. The 1cycle policy allows for convergence due to the learning rate being adjusted well below the initial, maximum learning rate throughout the cycle, and after numerous epochs in which the learning rate is much higher than the final learning rate. We found the 1cycle learning rate scheduler to provide stable results with convergence using a maximum learning rate of 0.009 and 30 epochs. We provide parameters in the toolbox that can enable the visualization of the losses and learning rate changes for both the training and validation phases. Further details on these key visualizations for supervised neural methods are provided in the GitHub repository.

DeepPhys [4]: A two-branch 2D convolutional attention network architecture. The two representations (appearance and difference frames) are processed by parallel branches with the appearance branch guiding the motion branch via a gated attention mechanism. The target signal is the first differential of the PPG waveform.

PhysNet [5]: A 3D convolutional network architecture. Yu et al. compared this 3D-CNN architecture with a 2D-CNN + RNN architecture, finding that a 3D-CNN version was able to achieve superior

四种不同的活动（静止、头部旋转、说话和行走）和锻炼场景。通过提供多个标签，该数据集的不同子集可以轻松地使用我们的工具箱进行研究。

BP4D+ [26]：该数据集包含140名受试者以每秒25帧的速率捕获的视频片段，每个受试者参与10项情绪诱导任务，总计1400次试验和相关视频。除了标准视频片段外，数据集还包括3D网格模型和热视频，两者都以相同的帧速率捕获。除此之外，该数据集还提供了补充数据，包括血压测量（波，收缩压，舒张压，平均值），心率（每分钟心跳次数），呼吸（波，速率bpm），皮肤电活动和面部动作编码系统（FACS）编码指定的动作单位。

UBFC-Phys [27]：UBFC-PHYS数据集是一个多模态数据集，包含168个RGB视频，每个任务有56个受试者（46名女性和10名男性）。在静态照明条件下，有三个任务具有大量的不受约束的运动-休息任务、语音任务和算术任务。该数据集包含通过Empatica E4腕带收集的金标准血容量脉搏（BVP）和皮肤电活动（EDA）测量值。使用EO-23121 C RGB数码相机以1024 x1024和35 Hz的分辨率记录视频。我们利用Sabour等人[27]的第二份补充材料中数据集作者提供的所有三项任务和相同的受试者子选择列表进行评价。我们重申了Escheridix-H中的主题子选择列表。

3.2 方法

3.2.1 无监督的方法

以下方法都使用线性代数和传统信号处理来恢复估计的PPG信号：1) 绿色[28]：在RGB视频的空间平均之后，将绿色通道信息用作PPG的代理；2) 伊卡[29]：将独立分量分析（ICA）应用于归一化的、空间平均的颜色信号以恢复去混合矩阵；3) CHROM [30]：从RGB视频获得的色度信号的线性组合用于估计；4) POS [31]：与皮肤正交的平面（POS），是基于生理和光学原理计算与肤色正交的投影平面的方法。将固定矩阵投影应用于空间归一化的平均像素值，其用于恢复PPG波形；5) PBV [32]：使用由给定光谱和血容量脉冲的变化确定的签名，以便导出PPG波形，同时抵消RGB视频中的运动和其他噪声；6) LGI [33]：一种通过可微局部变换对运动保持不变的特征表示方法。

3.2.2 监督神经方法

工具箱中包括监督学习算法的以下实现。所有实现都是使用PyTorch完成的[37]。常见的优化算法，如Adam [38]和AdamW [39]，以及标准，如均方误差（MSE）损失，用于训练，除非另有说明。学习率调度器通常遵循1cycle策略[40]，该策略将学习率从初始学习率退火到某个最大学习率，然后从该最大学习率退火到远低于初始学习率的某个学习率。该策略中的总步骤由epoch的数量乘以epoch中的训练批次的数量来确定。1cycle策略允许收敛，因为在整个周期内，学习率被调整到远低于初始的最大学习率，并且在学习率远高于最终学习率的许多时期之后。我们发现1cycle学习率调度器使用0.009的最大学习率和30个epoch提供稳定的收敛结果。我们在工具箱中提供了参数，可以在训练和验证阶段实现损失和学习率变化的可视化。GitHub存储库中提供了关于监督神经方法的这些关键可视化的更多细节。

DeepPhys [4]：一种双分支2D卷积注意力网络架构。这两个表示（外观和差异帧）由并行分支处理，外观分支通过门控注意机制引导运动分支。目标信号是PPG波形的一阶微分。

PhysNet [5]：3D卷积网络架构。Yu等人将这种3D-CNN架构与2D-CNN + RNN架构进行了比较，发现3D-CNN版本能够实现上级

Table 2: **Benchmark Results.** Performance on the UBFC-rPPG [22], PURE [23] UBFC-Phys [27] and MMPD [25] datasets generated using the rPPG toolbox. For the supervised methods we show cross-dataset training results using the UBFC-rPPG, PURE and SCAMPS datasets.

		Test Set								
		PURE [23]		UBFC-rPPG [22]		UBFC-Phys [27]		MMPD [25]		
	Method	Train Set	MAE \downarrow	MAPE \downarrow						
UNSUPERVISED	GREEN [28]	N/A	10.09	10.28	19.81	18.78	13.55	16.01	21.68	24.39
	ICA [29]	N/A	4.77	4.47	14.70	14.34	10.03	11.85	18.60	20.88
	CHROM [30]	N/A	5.77	11.52	3.98	3.78	4.49	6.00	13.66	15.99
	LGI [33]	N/A	4.61	4.96	15.80	14.70	6.27	7.83	17.08	18.98
	PBV [32]	N/A	3.91	4.82	15.90	15.17	12.34	14.63	17.95	20.18
	POS [31]	N/A	3.67	7.25	4.00	3.86	4.51	6.12	12.36	14.43
SUPERVISED			Test Set							
			PURE [23]		UBFC-rPPG [22]		UBFC-Phys [27]		MMPD [25]	
	TS-CAN [6]	UBFC-rPPG	3.69	3.38	N/A	N/A	5.13	6.53	14.00	15.47
	PURE		N/A		1.29		5.72		7.34	
	SCAMPS		4.66		5.83		3.53		5.55	
	PHYSNET [5]		UBFC-rPPG		8.06		N/A		6.91	
	PURE		N/A		0.98		5.79		19.05	
	SCAMPS		13.30		20.01		5.40		21.77	
	UBFC-rPPG		12.92		23.92		N/A		9.47	
	PHYSFORMER [34]		PURE		N/A		1.12		13.93	
DEEPPHYS [35]	SCAMPS		1.44		4.78		6.15		15.61	
	UBFC-rPPG		26.58		42.79		5.18		11.22	
	PURE		5.54		5.32		N/A		20.78	
EFF.PHYS-C [36]	SCAMPS		N/A		N/A		6.63		12.1	
	UBFC-rPPG		N/A		N/A		8.91		15.41	
EFF.PHYS-C [36]	PURE		N/A		2.07		6.04		14.57	
	SCAMPS		10.24		11.70		12.64		22.69	

MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%).

performance. Therefore, we included the 3D-CNN in this case. Instead of an MSE loss, a negative Pearson loss is utilized. It is worth noting that we used difference-normalized frames as input to PhysNet as the original paper does not specify a concrete input data format. Additional experiments involving raw frame inputs are included in Appendix-D.

PhysFormer [34]: PhysFormer is a video transformer-based architecture that adaptively aggregates both local and global spatio-temporal features toward rPPG representation enhancement. The architecture ultimately incorporated and emphasized long-term, global features and allowed for notable improvements in performance relative to various methods, including POS [31], DeepPhys [4], and PhysNet [5]. Instead of an MSE loss, a dynamic loss composed of numerous hyperparameters, a negative Pearson loss, a frequency cross-entropy loss, and a label distribution loss is used. Furthermore, the chosen learning rate scheduler differs from our toolbox’s default learning rate scheduler (1cycle policy [40]) in that we effectively opt to use a single, constant learning rate for training. We utilize difference-normalized frames as an input to PhysFormer.

TS-CAN [6]: A two-branch 2D convolutional attention network architecture that leverages temporal shift operation information across the time axis to perform efficient temporal and spatial modeling. This network is an on-device, real-time algorithm. The target signal is the first differential of the PPG waveform.

EfficientPhys-C [36]: A single-branch 2D convolutional neural network that aims to provide an end-to-end, super lightweight network for real-time on-device computation. The architecture has a normalization module that calculates frame differences and learnable normalization, as well as a self-attention module to help the network focus on skin pixels associated with PPG signal.

3.3 Pre-Processing, Training, Post-Processing and Evaluation

In the rPPG-Toolbox, we offer a configuration file system that enables users to modify all parameters used in pre-processing, training, post-processing, and evaluation. A YAML file is provided for every experiment and includes blocks for pre/post-processing, training, validation, testing, model

表2：基准结果。使用rPPG工具箱生成的UBFC-rPPG [22]、PURE [23] UBFC-Phys [27]和MMPD [25]数据集的性能。对于监督方法，我们使用UBFC-rPPG，PURE和SCAMPS数据集显示了交叉数据集训练结果。

		测试集 [25]第二十三话：我的世界，我的世界[26]															
		方法训练集MAE MAPE MAE MAPE MAE MAPE MAE MAPE															
G[28]	不适用	10.09	10.28	19.81	18.78	13.55	16.01	21.68	24.39	伊卡[29]	不适用	4.77	4.47	14.70	14.34	10.03	
		11.85	18.60	20.88	CHROM	[30]	不适用	阿5.77	11.52	3.98	3.78	4.49	6.00	13.66	15.99		
LGI	[33]	不适用	4.61	4.96	15.80	14.70	6.27	7.83	17.08	18.98	PBV	[32]	不适用	3.91	4.82	15.90	15.17
		12.34	14.63	17.95	20.18	POS	[31]	不适用	A 3.67	7.25	4.00	3.86	4.51	6.12	12.36	14.43	
UBFC-PPG	3.69	3.38	不适用	不适用	5.13	6.53	14.00	15.47	纯	不适用	不适用	1.29	1.50	5.72			
能 (S—CAN)	7.34	13.93	15.14	SCAMPS	4.66	5.83	3.62	3.53	5.55	6.91	19.05	21.77					
PN[5]	UBFC-PPG	8.06	13.67	不适用	不适用	5.79	7.69	9.47	11.11	纯	不适用	0.98	1.12	4.78	6.15		
		13.93	15.61	SCAMPS	13.30	20.01	5.40	5.43	8.53	11.22	20.78	24.43					
PF[34]	UBFC-PPG	12.92	23.92	不适用	不适用	6.63	8.91	12.1	15.41	纯	不适用	不适用	1.44	1.66	6.04		
		7.67	14.57	16.73	SCAMPS	26.58	42.79	4.56	5.18	11.91	15.57	22.69	27.06				
DP[35]	UBFC-PPG	5.54	5.32	不适用	不适用	6.62	8.21	17.49	19.26	纯	不适用	不适用	1.21	1.42	8.42		
		10.18	16.92	18.54	SCAMPS	3.95	4.25	3.10	3.08	4.75	5.89	15.22	16.56				
UBFC-PPG	5.47	5.39	N/A	N/A	A 4.93	6.25	13.78	15.15	PURE	N/A	N/A	2.07	2.10	5.31			
EfficientPhys-C[36]	5.31	SCAMPS	10.24	11.70	12.64	11.26	6.97	8.47	20.41	23.52	MAE = HR估计						
的平均绝对误差（心跳/分钟），MAPE = 平均百分比误差（%）。																	

性能因此，我们在这种情况下包括3D-CNN。代替MSE损失，使用负Pearson损失。值得注意的是，我们使用差分归一化帧作为PhysNet的输入，因为原始论文没有指定具体的输入数据格式。涉及原始帧输入的其他实验包括在EPDix-D中。

[34]第三十四话：PhysFormer是一种基于视频变换器的架构，自适应地聚合局部和全局时空特征，以增强rPPG表示。该架构最终纳入并强调了长期的全局特性，并允许相对于各种方法（包括POS [31]，DeepPhys [4]和PhysNet [5]）的性能显着改进。代替MSE损失，使用由许多超参数、负皮尔逊损失、频率交叉熵损失和标签分布损失组成的动态损失。此外，所选择的学习率调度程序与我们工具箱的默认学习率调度程序（1cycle policy [40]）不同，因为我们有效地选择使用单个恒定的学习率进行训练。我们利用差分归一化帧作为PhysFormer的输入。

TS-CAN [6]：一种双分支2D卷积注意力网络架构，其利用跨时间轴的时间移位操作信息来执行高效的时间和空间建模。该网络是一种设备上的实时算法。目标信号是PPG波形的一阶微分。

EfficientPhys-C [36]：一种单分支2D卷积神经网络，旨在为实时设备计算提供端到端的超轻量网络。该架构具有计算帧差和可学习归一化的归一化模块，以及帮助网络聚焦于与PPG信号相关的皮肤像素的自关注模块。

3.3 预处理、培训、后处理和评估

在rPPG-EXP中，我们提供了一个配置文件系统，使用户能够修改在预处理，训练，后处理和评估中使用的所有参数。每个实验都提供了一个YAML文件，包括预/后处理、训练、验证、测试、模型

hyperparameters, and computational resources. The pre/post-processing for neural and unsupervised methods share similar settings, such as the same input resolution and face cropping.

In terms of pre-processing, we provide three input data types: 1) "DiffNormalized", which calculates the difference of every two consecutive frames and labels, and normalizes them by their standard deviation; 2) "Standardized", which standardizes the raw frames and labels using z-score; 3) "Raw", which uses the original frames and labels without modification. Additionally, we provide parameters for face cropping, a vital aspect of our task. In the config file, users can use dynamic detection to perform face cropping every N frames and scale the face bounding box by a coefficient to maintain consistency of face cropping in motion videos. Users can also elect to use a median bounding box with dynamic detection in order to help filter out erroneous detections of the face.

With regard to the training of neural networks, our toolbox provides flexibility to parameterize which portion of the data is used for training, validation, or testing. For instance, we can use the first 80% of UBFC-rPPG for training, the last 20% of UBFC-rPPG for validation and then use the entire PURE dataset for testing. Moreover, the distinct parameters (e.g., dropout rate) of each neural network can be defined in the config file.

For post-processing and evaluation, there are several standard post-processing steps that are typically employed to improve model predictions. A 2nd-order Butterworth filter (cut-off frequencies of 0.75 and 2.5 Hz) is applied to filter the predicted PPG waveform. The choice of filtering parameters can have a significant impact on downstream results such as heart rate errors. A Fast Fourier Transform or a peak detection algorithm is then applied to the filtered signal to calculate the heart rate. In this toolbox, we support five metrics for video-level heart rate estimations: mean absolute error (MAE), root mean squared error (RMSE), mean absolute percentage error (MAPE), signal-noise ratio (SNR), and Pearson Correlation (ρ), along with a calculation of standard error for a better understanding of the accuracy of the aforementioned metrics. We also give users the option to visualize Bland-Altman plots as a part of evaluation. Finer details on the supported metrics (F), metric results not reported in the main paper (G), and Bland-Altman plots (J.3) appear in the respective appendices. For better reproducibility, we also provide pre-trained models in our GitHub repository to allow researchers to perform model inference. The detailed definition of each config parameter is also provided in the GitHub repository.

3.4 Benchmarking

To show that the implementations of the baseline methods are functioning as expected and to provide benchmark results for consumers of the toolbox to reference and reproduce, we performed a set of baseline experiments using three commonly used video rPPG datasets for training: SCAMPS [24], UBFC-rPPG [22] and PURE [23] and tested on four datasets including UBFC-rPPG [22], PURE [23], UBFC-Phys [27], and MMPD [25]. Except where noted in the GitHub repository, neural models utilized a training batch size of 4, 30 epochs, and an inference batch size of 4 for all experiments. Due to the multi-institution team behind this toolbox, different kinds of GPUs were utilized to produce benchmark results. Tables 2 and 4 were produced on a machine using a single NVIDIA RTX A4500 GPU. Tables 3 and 5 were produced on a machine using a single NVIDIA GeForce 2080 Ti GPU. As illustrated in Table 2, we show MAE and MAPE computed between the video-level heart rate estimations and gold standard measurements. Additional cross-dataset experiment results and metric results can be found in Appendix-G and F.

4 Additional Features

4.1 Weakly Supervised Training

Supervised rPPG training requires high fidelity synchronous PPG waveform labels. However not all datasets contain such high quality labels. In these cases we offer the option to train on synchronous PPG "pseudo" labels derived through a signal processing methodology as described by [41]. These labels are produced through POS-generated [31] PPG waveforms, which are then bandpass filtered around the normal heart-rate frequencies (cut-off frequencies of 0.70 and 3.0 Hz), and finally amplitude normalized using a Hilbert-signal envelope. The tight filtering and envelope normalization results in a strong periodic proxy signal, but at the cost of limited signal morphology.

超参数和计算资源。神经和无监督方法的预处理/后处理共享类似的设置，例如相同的输入分辨率和面部裁剪。

在预处理方面，我们提供了三种输入数据类型：1) “DiffNormalized”，它计算每两个连续帧和标签的差异，并通过它们的标准差对其进行归一化；2) “Standardized”，它使用z得分对原始帧和标签进行归一化；3) “Raw”，它使用原始帧和标签而不进行修改。此外，我们提供了面部裁剪的参数，这是我们任务的一个重要方面。在配置文件中，用户可以使用动态检测来每N帧执行面部裁剪，并通过系数缩放面部边界框，以保持运动视频中面部裁剪的一致性。用户还可以选择使用具有动态检测的中值边界框，以帮助过滤掉对面部的错误检测。

关于神经网络的训练，我们的工具箱提供了灵活性，可以参数化数据的哪一部分用于训练，验证或测试。例如，我们可以使用UBFC-rPPG的前80%进行训练，使用UBFC-rPPG的后20%进行验证，然后使用整个PURE数据集进行测试。此外，不同的参数（例如，dropout rate）可以在配置文件中定义。

对于后处理和评估，有几个标准的后处理步骤通常用于改进模型预测。应用二阶巴特沃思滤波器（截止频率为0.75和2.5 Hz）对预测的PPG波形进行滤波。滤波参数的选择可能对心率误差等下游结果产生重大影响。然后将快速傅立叶变换或峰值检测算法应用于滤波后的信号以计算心率。在这个工具箱中，我们支持视频级心率估计的五个指标：平均绝对误差 (MAE)，均方根误差 (RMSE)，平均绝对百分比误差 (MAPE)，信噪比 (SNR) 和皮尔逊相关性 (ρ)，沿着标准误差的计算，以便更好地理解上述指标的准确性。我们还为给予用户提供了可视化Bland–Altman图的选项，作为评估的一部分。支持指标 (F)、主论文中未报告的指标结果 (G) 和Bland–Altman图 (J.3) 的更详细信息见相应附录。为了更好的再现性，我们还在GitHub存储库中提供了预训练的模型，以允许研究人员进行模型推理。每个配置参数的详细定义也在GitHub存储库中提供。

3.4 标杆

为了表明基线方法的实现按预期运行，并为工具箱的消费者提供基准结果以供参考和复制，我们使用三个常用的视频rPPG数据集进行了一组基线实验：SCAMPS [24]、UBFC–rPPG [22]和PURE [23]，并在包括UBFC–rPPG [22]、PURE [23]、UBFC–Phys [27]和MMPD [25]在内的四个数据集上进行测试。除了在GitHub存储库中注明的地方，神经模型在所有实验中使用的训练批量大小为4, 30 epochs，推理批量大小为4。由于该工具箱背后的多机构团队，使用了不同类型的GPU来生成基准测试结果。表2和表4是在使用单个NVIDIA RTX A4500 GPU的机器上生成的。表3和表5在使用单个NVIDIA GeForce 2080 Ti GPU的机器上产生。如表2所示，我们显示了在视频级心率估计和黄金标准测量之间计算的MAE和MAPE。其他的跨数据集实验结果和度量结果可以在Eudix–G和F中找到。

4 附加特征

4.1 弱监督训练

监督rPPG训练需要高保真同步PPG波形标签。然而，并非所有数据集都包含如此高质量的标签。在这些情况下，我们提供了在通过[41]所述的信号处理方法导出的同步PPG“伪”标签上训练的选项。这些标签是通过POS生成的[31] PPG波形产生的，然后在正常心率频率周围进行带通滤波（截止频率为0.70和3.0 Hz），最后使用希尔伯特信号包络进行幅度归一化。紧密滤波和包络归一化导致强周期性代理信号，但以有限的信号形态为代价。

For instance, in the BP4D+ dataset [26], the cardiac ground truth is represented by a blood pressure waveform. Although this waveform exhibits the same periodicity as the PPG signal, it has a phase shift that adversely affects model training. A figure that illustrates sample pseudo labels derived for BP4D+ [26] videos plotted against the ground truth blood pressure waveform as shown in Figure 3. Table 3 presents results for supervised methods, trained on BP4D+ [26] pseudo labels. We extend this feature to all of the supported datasets.

Table 3: **Training with Pseudo Labels.** For the supervised methods we show results training with the (entire) BP4D+ [26] dataset, using POS [31] derived pseudo training labels.

Training Set Testing Set	BP4D+[26] with POS Pseudo Labels					
	UBFC-rPPG [22]			PURE [23]		
	MAE↓	MAPE↓	$\rho \uparrow$	MAE↓	MAPE↓	$\rho \uparrow$
Supervised						
TS-CAN [6]	4.69	4.51	0.78	1.29	1.60	0.97
PhysNet(Normalized) [5]	1.78	1.92	0.96	3.69	7.35	0.88
DeepPhys [35]	2.74	2.81	0.93	2.47	2.49	0.89
EfficientPhys-C [36]	2.43	2.52	0.90	3.59	3.27	0.80

MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%), ρ = Pearson Correlation in HR estimation.

4.2 Motion Augmented Training

The usage of synthetic data in the training of machine learning models for medical applications is becoming a key tool that warrants further research [42]. In addition to providing support for the fully synthetic dataset SCAMPS [24], we provide support for synthetic, motion-augmented versions of the UBFC-rPPG [22], PURE [23], SCAMPS [24], and UBFC-PHYS [27] datasets for further exploration toward the use of synthetic data for training rPPG models. The synthetic, motion-augmented datasets are generated using an open-source motion augmentation pipeline targeted for increasing motion diversity in rPPG videos [43]. We present cross-dataset results using a motion-augmented version of the UBFC-rPPG [22] dataset in Table 4. We also provide tools that leverage OpenFace [44] for extracting, visualizing, and analyzing motion in rPPG video datasets. Further details regarding these tools are shared in our GitHub repository.

Table 4: **Training with Motion-Augmented Data.** We demonstrate results training on a motion-augmented (MA) version of the UBFC-rPPG [22] dataset generated using an open-source motion augmentation pipeline [43] and testing on the unaugmented version of the PURE [23] dataset.

Training Set Testing Set	MAUBFC-rPPG [22]								
	PURE [23]			UBFC-Phys [27]			MMPD [25]		
	MAE↓	MAPE↓	$\rho \uparrow$	MAE↓	MAPE↓	$\rho \uparrow$	MAE↓	MAPE↓	$\rho \uparrow$
TS-CAN [6]	1.07	1.20	0.97	5.03	6.36	0.75	12.59	13.77	0.23
PhysNet (Normalized) [5]	17.03	32.37	0.38	5.51	7.50	0.68	10.67	13.99	0.33
DeepPhys [35]	1.15	1.40	0.97	4.95	6.26	0.75	12.71	13.70	0.21
EfficientPhys-C [36]	2.59	2.67	0.88	4.80	6.10	0.79	13.39	14.50	0.14

MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%), ρ = Pearson Correlation in HR estimation.

4.3 Extending the rPPG-Toolbox for Physiological Multitasking

While this toolbox is primarily targeted towards rPPG model training and evaluation, it can be easily extended to support multi-tasking of physiological signals. As an example, we implement BigSmall [41], an architecture that multi-tasks PPG, respiration, and facial action. Similar to [41] we present 3-fold cross-validation results across the action unit (AU) subset of BP4D+ [26] (the portion of the dataset with AU labels), and use the same subject folds and hyperparameters as implemented in the original publication. These results can be found in Table 5. Note, that like [41], facial action metrics are calculated across 12 common AUs (AU #s 1, 2, 4, 6, 7, 10, 12, 14, 15, 17, 23, 24). Additional details of training and evaluation could be found in Appendix-I.

例如，在BP 4D+数据集[26]中，心脏基础事实由血压波形表示。尽管该波形表现出与PPG信号相同的周期性，但它具有对模型训练产生不利影响的相移。该图显示了针对BP 4D + [26]视频导出的样本伪标签，并根据图3所示的地面真实血压波形绘制。表3显示了在BP 4D + [26]伪标签上训练的监督方法的结果。我们将此功能扩展到所有支持的数据集。

表3：使用伪标签进行训练。对于监督方法，我们使用（整个）BP 4D + [26]数据集，使用POS [31]衍生的伪训练标签显示训练结果。

	训练集BP 4D +[26]与POS伪标签测试集UBFC-rPPG [22] PURE [23]					
	MAE ↓	MAPE ↓	ρ ↑	MAE ↓	MAPE ↓	ρ ↑
监督						
TS-CAN [6]	4.69	4.51	0.78	1.29	1.60	0.97
PhysNet (归一化)						
[5]	1.78	1.92	0.96	3.69	7.35	0.88
DeepPhys [35]	2.74	2.81	0.93	2.47	2.49	0.89
EffcientPhys-C [36]						
	2.43	2.52	0.90	3.59	3.27	0.80
MAE = HR 估计的平均绝对误差（心跳/分钟）， MAPE = 平均百分比误差（%）， ρ = HR估计的皮尔逊相关性。						

4.2 运动增强训练

在用于医疗应用的机器学习模型的训练中使用合成数据正在成为值得进一步研究的关键工具[42]。除了为全合成数据集SCAMPS [24]提供支持外，我们还为UBFC-rPPG [22]、PURE [23]、SCAMPS [24]和UBFC-PHYS [27]数据集的合成、运动增强版本提供支持，以进一步探索使用合成数据训练rPPG模型。使用开源运动增强管道生成合成的运动增强数据集，旨在增加rPPG视频中的运动多样性[43]。我们使用表4中的UBFC-rPPG [22]数据集的运动增强版本提供了交叉数据集结果。我们还提供了利用OpenFace [44]提取、可视化和分析rPPG视频数据集中运动的工具。关于这些工具的更多细节在我们的GitHub存储库中共享。

表4：使用运动增强数据进行训练。我们展示了使用开源运动增强管道[43]生成的UBFC-rPPG [22]数据集的运动增强（MA）版本的训练结果[43]，以及对PURE [23]数据集的未增强版本的测试结果。

	训练集MAUBFC-rPPG [22]测试集PURE [23]					
	UBFC-Phys [27]	MMPD [25]				
MAE ↓	MAPE ↓	ρ ↑	MAE ↓	MAPE ↓	ρ ↑	MAE ↓
TS-CAN [6]	1.07					
PhysNet (归一化)	1.20	0.97	5.03	6.36	0.75	12.59
[5]	13.77	0.23				
			17.03	32.37	0.38	
5.51	7.50	0.68	10.67	13.99	0.33	
DeepPhys [35]	1.15	1.40	0.97	4.95	6.26	0.75
EffcientPhys-C [36]						
	12.71	13.70	0.21			
0.79	13.39	14.50	0.14	MAE = HR 估计的平均绝对误差（心跳/分钟）， MAPE = 平均百分比误差（%）， ρ = HR 估计中的皮尔逊相关性。		

4.3 将rPPG–神经网络扩展为生理多任务处理

虽然该工具箱主要针对rPPG模型训练和评估，但它可以轻松扩展以支持生理信号的多任务处理。例如，我们实现了BigSmall [41]，这是一种多任务PPG，呼吸和面部动作的架构。与[41]类似，我们在BP 4D + [26]的作用单位（Au）子集（具有Au标签的数据集部分）中呈现3折交叉验证结果，并使用与原始出版物中实现的相同的受试者折叠和超参数。这些结果见表5。请注意，与[41]类似，面部动作度量是在12个公共AU（AU #1, 2, 4, 6, 7, 10, 12, 14, 15, 17, 23, 24）上计算的。

关于培训和评价的更多细节，见《第一次报告》。

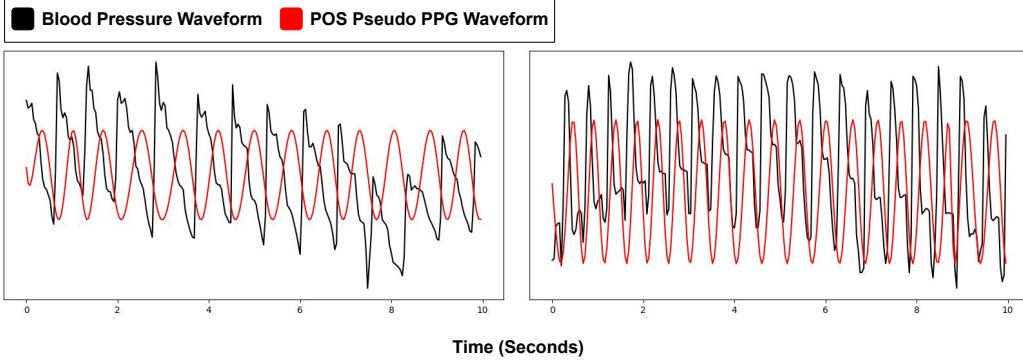


Figure 3: **Generated Pseudo Labels.** Samples of POS [31] generated PPG pseudo labels plotted against ground truth blood pressure waveforms from BP4D+ [26].

Table 5: **Multitasking Results.** For the BigSmall [41] method we show results for multi-tasking PPG, respiration, and action unit classification; training with the BP4D+ [26] (AU subset) dataset, using POS [31] derived pseudo training PPG labels.

Task	Training Set			BP4D+ [26]			Testing Set			BP4D+ [26]		
	Testing Set			Respiration			Testing Set			Facial Action		
	rPPG	MAE↓	MAPE↓	$\rho \uparrow$	Respiration	MAE↓	MAPE↓	$\rho \uparrow$	F1↑	Prec. ↑	Acc. ↑	
BigSmall [41]		3.23	3.51	0.83		5.19	26.28	0.14	42.82	39.85	65.73	

MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%), ρ = Pearson Correlation in HR estimation, F1 = average F1 across 12 action units, Prec. = average precision across 12 action units, Acc. = average accuracy across 12 action units.

4.4 Training, Evaluation and Analysis Features

In this toolbox, we have incorporated a diverse set of training and evaluation functionalities. These include: 1) data pre-processing visualization tools, enabling users to inspect and understand their data before feeding it into the model; 2) comprehensive tracking and visualization of key training metrics such as training loss, validation loss, and learning rate, facilitating a thorough monitoring of the model’s learning progress; 3) the implementation of Bland-Altman plots, providing a robust method for assessing agreement between two ground-truth HR and predicted HR; 4) advanced motion analysis capabilities. For an in-depth exploration of these features, we direct the reader to Appendix-J and our Github page, where detailed descriptions and usage examples are provided.

5 Limitations

We acknowledge there are many limitations in our current toolbox and plan to continue improve it in the future. In the ensuing phases of our research, we envision a collaborative approach, working in concert with the wider scientific community, to enhance the efficacy and capabilities of the rPPG-Toolbox. The current limitations include 1) this toolbox does not support all of the latest neural architectures and diverse datasets; 2) it does not support unsupervised and self-supervised learning paradigms; 3) it does not support applications beyond of heart rate calculation such as heart rate variability, blood pressure, SpO₂, and other importasnt physiological measures.

6 Broader Impacts

Camera sensing has advantages and benefits with the potential to make important cardiac measurement more accessible and comfortable. One of the motivating use-cases for rPPG is turning everyday devices equipped with cameras into scalable health sensors. However, pervasive measurement can also feel intrusive. We are releasing the rPPG toolbox with a Responsible AI License [45] that

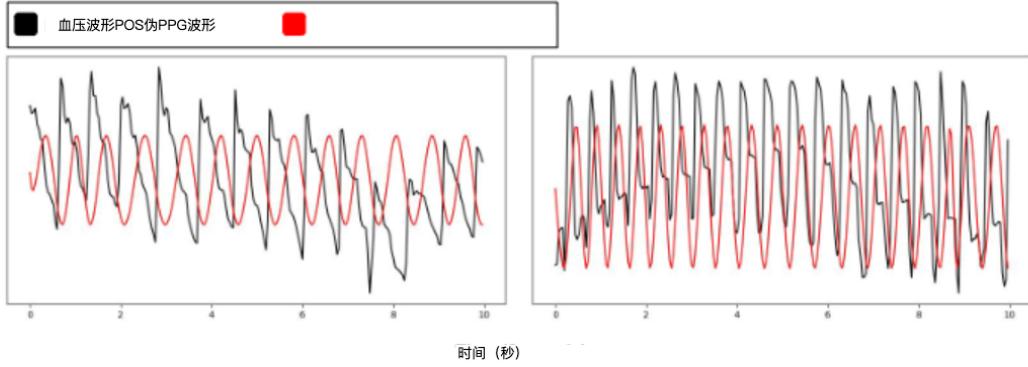


图3：生成的伪标签。POS [31]的样本生成了PPG伪标签，并根据BP 4D + [26]的真实血压波形绘制。

表5：多任务处理结果。对于BigSmall [41]方法，我们显示了多任务PPG、呼吸和动作单元分类的结果；使用POS [31]衍生的伪训练PPG标签，使用BP 4D + [26]（Au子集）数据集进行训练。

[26]第二十六话
测试套件BP 4D + [26]
任务rPPG呼吸面部动作
MAE ↓ MAPE ↓ ρ ↑ MAE ↓ MAPE ↓ ρ ↑ F1↑ Prec. ↑ Acc. ↑ BigSmall [41] 3.23 3.51 0.83 5.19 26.28 0.14
42.82 39.85 65.73 MAE = HR估计的平均绝对误差(心跳/分钟)，MAPE = 平均百分比误差(%)，ρ = HR估计中的Pearson相关性，F1 = 12个动作单位的平均F1，Prec. = 12个动作单位的平均精密度，Acc. = 12个行动单元的平均准确度。

4.4 培训、评估和分析功能

在这个工具箱中，我们整合了各种培训和评估功能。其中包括：1) 数据预处理可视化工具，使用户能够在将其输入模型之前检查和理解他们的数据；2) 全面跟踪和可视化关键训练指标，如训练损失，验证损失和学习率，促进对模型学习进度的全面监控；3) Bland–Altman图的实现，提供了一种用于评估两个地面真实HR和预测HR之间一致性的稳健方法；4) 先进的运动分析功能。为了深入探索这些功能，我们将读者引导到GitHub页面和我们的GitHub页面，其中提供了详细的描述和使用示例。

5 限制

我们承认目前的工具箱存在许多局限性，并计划在未来继续改进。在我们研究的随后阶段，我们设想了一种协作方法，与更广泛的科学界合作，以提高rPPGdim的功效和能力。目前的局限性包括：1) 该工具箱不支持所有最新的神经架构和不同的数据集；2) 它不支持无监督和自监督学习范例；3) 它不支持心率计算以外的应用，如心率变异性，血压，SpO和其他重要的生理指标。

6 更广泛的影响

摄像机传感具有优势和益处，有可能使重要的心脏测量更容易获得和舒适。rPPG的一个激励用例是将配备摄像头的日常设备转变为可扩展的健康传感器。然而，普遍的测量也会让人感到侵入性。我们正在发布带有负责任AI许可证的rPPG工具箱[45]，

restricts negative and unintended uses of the toolbox. We also acknowledge the presence of several potential negative concerns and impacts, which are described as follows.

Privacy Concerns: Camera-based physiological sensing offers a revolutionary way to extract physiological signals from video recordings, enabling a myriad of applications, from remote patient monitoring to daily fitness tracking. However, these advancements come with significant privacy concerns. First and foremost, the very nature of remote sensing allows for the collection of personal data without direct physical interaction or, in some cases, knowledge of the individual being monitored. This can potentially enable unauthorized entities to capture sensitive physiological data covertly. Furthermore, as these systems become more widespread, there is a risk that everyday places such as shopping malls, public transport, and even workplaces might employ rPPG systems, leading to widespread passive data collection. Such extensive monitoring can lead to privacy invasions, where individuals are constantly under physiological surveillance without their explicit consent.

Potential Negative Impact: Beyond individual privacy, there are broader societal implications of widespread contactless camera-based physiological sensing. There's the potential for the creation of a pervasive surveillance state where citizens are continuously monitored, not just for their actions but also for their physiological responses. Such monitoring can lead to "physiological profiling," where individuals are judged, categorized, or even discriminated against based on their bodily responses. For instance, elevated heart rates or other physiological markers might be misinterpreted as signs of nervousness, guilt, or deceit, potentially affecting decision-making in areas such as law enforcement, job interviews, or public services. Moreover, a continuous emphasis on physiological metrics might foster an environment of physiological conformism, where people feel pressured to exhibit 'normal' physiological signs even if they aren't feeling well or are under duress.

Potential Ethical Concerns: The ethical implications of rPPG are multi-faceted. Firstly, there is the concern of informed consent. As technology becomes more integrated into our environments, it becomes challenging to ensure that individuals are aware of, and have agreed to, the collection of their physiological data. Moreover, the accuracy and reliability of rPPG systems can vary depending on factors like skin tone, lighting conditions, and other external factors. This introduces the risk of systematic biases, where certain groups might be inaccurately assessed or marginalized due to technological limitations or inherent biases in the algorithms. Ethical concerns also arise from potential misuse. For example, businesses might use rPPG data to gauge consumer reactions to products or advertisements, leading to manipulative strategies that target individual vulnerabilities. In extreme cases, authoritarian regimes might use such technologies to monitor citizens for signs of dissent or unrest. As with any potent tool, the ethical application of rPPG requires careful consideration of its potential for both benevolent and malevolent use.

7 Conclusion

Research relies on the sharing of ideas, this not only allows methods to be verified, saving time and resources, but also allows researchers to more effectively build upon existing work. Without these resources and open-sourced code bases, fair evaluation and comparison of methods is difficult, creates needless repetitions, and wastes resources. We present an end-to-end and comprehensive toolbox, called rPPG-Toolbox, containing code for pre-processing multiple public datasets, implementations of supervised machine learning (including training pipeline) and unsupervised methods, and post-processing and evaluation tools.

8 Acknowledgement

This project is supported by a Google PhD Fellowship for Xin Liu and a research grant from Cisco for the University of Washington as well as a career start-up funding grant from the Department of Computer Science at UNC Chapel Hill. This research is also supported by Tsinghua University Initiative Scientific Research Program, Beijing Natural Science Foundation, and the Natural Science Foundation of China (NSFC).

限制工具箱的负面和非预期用途。我们还认识到存在若干潜在的负面关切和影响，具体如下。

隐私问题：基于摄像头的生理传感提供了一种从视频记录中提取生理信号的革命性方法，从而实现了从远程患者监护到日常健身跟踪的各种应用。然而，这些进步带来了严重的隐私问题。首先，遥感的性质本身允许在没有直接物理互动或在某些情况下不了解被监测个人的情况下收集个人数据。这可能使未经授权的实体能够秘密地捕获敏感的生理数据。此外，随着这些系统变得越来越普遍，存在这样的风险，即诸如购物中心、公共交通甚至工作场所等日常场所可能会使用rPPG系统，从而导致广泛的被动数据收集。这种广泛的监控可能导致隐私侵犯，个人在未经其明确同意的情况下不断受到生理监视。

潜在负面影响：除了个人隐私之外，广泛使用的基于非接触式摄像头的生理传感还有更广泛的社会影响。有可能创造一个无处不在的监视状态，公民被持续监视，不仅是他们的行为，而且是他们的生理反应。这种监测可能会导致“生理分析”，其中个人被判断，分类，甚至歧视基于他们的身体反应。例如，心率升高或其他生理指标可能会被误解为紧张、内疚或欺骗的迹象，可能会影响执法、求职面试或公共服务等领域的决策。此外，持续强调生理指标可能会培养一种生理顺从的环境，在这种环境中，人们感到有压力表现出“正常”的生理迹象，即使他们感觉不舒服或受到胁迫。

潜在伦理问题：rPPG的伦理影响是多方面的。首先，是知情同意的问题。随着技术越来越多地融入我们的环境，确保个人意识到并同意收集他们的生理数据变得具有挑战性。此外，rPPG系统的准确性和可靠性可能会因肤色、照明条件和其他外部因素等因素而异。这引入了系统性偏差的风险，由于技术限制或算法中的固有偏差，某些群体可能会被不准确地评估或边缘化。潜在的滥用也会引起伦理问题。例如，企业可能会使用rPPG数据来衡量消费者对产品或广告的反应，从而导致针对个别漏洞的操作策略。在极端情况下，独裁政权可能会使用这些技术来监视公民的异议或动乱迹象。与任何有效的工具一样，rPPG的伦理应用需要仔细考虑其善意和恶意使用的潜力。

7 结论

研究依赖于思想的共享，这不仅可以验证方法，节省时间和资源，还可以让研究人员更有效地建立在现有的工作。如果没有这些资源和开源代码库，就很难对方方法进行公平的评估和比较，这会造成不必要的重复，并浪费资源。我们提出了一个端到端的综合工具箱，称为rPPG-PLAN，包含用于预处理多个公共数据集的代码，监督机器学习（包括训练管道）和无监督方法的实现，以及后处理和评估工具。

8 确认

该项目得到了Google为Xin Liu提供的博士奖学金、思科为华盛顿大学提供的研究资助以及来自查佩尔山计算机科学系的职业启动资助。本研究得到了清华大学重大科学研究计划、北京市自然科学基金、国家自然科学基金的资助。

References

- [1] Lonneke AM Aarts, Vincent Jeanne, John P Cleary, C Lieber, J Stuart Nelson, Sidarto Bambang Oetomo, and Wim Verkruyse. Non-contact heart rate monitoring utilizing camera photoplethysmography in the neonatal intensive care unit—a pilot study. *Early human development*, 89(12):943–948, 2013.
- [2] L Tarassenko, M Villarroel, A Guazzi, J Jorge, DA Clifton, and C Pugh. Non-contact video-based vital sign monitoring using ambient light and auto-regressive models. *Physiological measurement*, 35(5):807, 2014.
- [3] Bryan P Yan, William HS Lai, Christy KY Chan, Alex CK Au, Ben Freedman, Yukkee C Poh, and Ming-Zher Poh. High-throughput, contact-free detection of atrial fibrillation from video with deep learning. *JAMA cardiology*, 5(1):105–107, 2020.
- [4] Weixuan Chen and Daniel McDuff. DeepPhys: Video-Based Physiological Measurement Using Convolutional Attention Networks. *arXiv:1805.07888 [cs]*, August 2018. arXiv: 1805.07888.
- [5] Zitong Yu, Wei Peng, Xiaobai Li, Xiaopeng Hong, and Guoying Zhao. Remote heart rate measurement from highly compressed facial videos: an end-to-end deep learning solution with video enhancement. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 151–160, 2019.
- [6] Xin Liu, Josh Fromm, Shwetak Patel, and Daniel McDuff. Multi-task temporal shift attention networks for on-device contactless vitals measurement. *Advances in Neural Information Processing Systems*, 33:19400–19411, 2020.
- [7] Zitong Yu, Xiaobai Li, Pichao Wang, and Guoying Zhao. Transrppg: Remote photoplethysmography transformer for 3d mask face presentation attack detection. *IEEE Signal Processing Letters*, 2021.
- [8] Daniel McDuff, Javier Hernandez, Erroll Wood, Xin Liu, and Tadas Baltrusaitis. Advancing non-contact vital sign measurement using synthetic avatars. *arXiv preprint arXiv:2010.12949*, 2020.
- [9] Zhen Wang, Yunhao Ba, Pradyumna Chari, Oyku Deniz Bozkurt, Gianna Brown, Parth Patwa, Niranjan Vaddi, Laleh Jalilian, and Achuta Kadambi. Synthetic generation of face videos with plethysmograph physiology. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 20587–20596, 2022.
- [10] Xin Liu, Yuntao Wang, Sinan Xie, Xiaoyu Zhang, Zixian Ma, Daniel McDuff, and Shwetak Patel. Mobilephys: Personalized mobile camera-based contactless physiological sensing. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.*, 6(1), mar 2022.
- [11] Xin Liu, Ziheng Jiang, Josh Fromm, Xuhai Xu, Shwetak Patel, and Daniel McDuff. Metaphys: few-shot adaptation for non-contact physiological measurement. In *Proceedings of the Conference on Health, Inference, and Learning*, pages 154–163, 2021.
- [12] Xin Liu, Mingchuan Zhang, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. Federated remote physiological measurement with imperfect data. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2155–2164, 2022.
- [13] Zhaodong Sun and Xiaobai Li. Contrast-phys: Unsupervised video-based remote physiological measurement via spatiotemporal contrast. In *European Conference on Computer Vision*, pages 492–510. Springer, 2022.
- [14] John Gideon and Simon Stent. The way to my heart is through contrastive learning: Remote photoplethysmography from unlabelled video. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 3995–4004, 2021.
- [15] Jeremy Speth, Nathan Vance, Patrick Flynn, and Adam Czajka. Non-contrastive unsupervised learning of physiological signals from video. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 14464–14474, 2023.

参考文献

- [1]Lonneke AM Aarts、Vincent Jeanne、John P Cleary、C Lieber、J Stuart纳尔逊、Sidarto Bambang Oetomo和Wim Verkruyse。在新生儿重症监护病房使用照相机光电容积描记法进行非接触式心率监测的初步研究。Early Human Development, 89 (12) : 943–948, 2013.
- [2]L Tarassenko, M Villarroel, A Guazzi, J Jorge, DA克利夫顿和C Pugh。使用环境光和自回归模型的基于视频的非接触式生命体征监测。生理测量, 35 (5) : 807, 2014。
- [3]Bryan P Yan、William HS Lai、Christy KY Chan、Alex CK Au、Ben Freedman、Yukkee C Poh和Ming-Zher Poh。通过深度学习从视频中高通量、无接触地检测房颤。JAMA心脏病学, 5 (1) : 105–107, 2020。
- [4]Weixuan Chen和丹尼尔麦克达夫。DeepPhys: 使用卷积注意力网络进行基于视频的生理测量。arXiv: 1805.07888 [cs], 2018年8月。arXiv: 1805.07888。
- [5]于梓潼, 彭伟, 李小白, 洪小鹏, 赵国英。从高度压缩的面部视频中进行远程心率测量: 端到端深度学习解决方案,

视频增强在IEEE/CVF计算机视觉国际会议论文集, 第151–160页, 2019年。

- [6]刘欣、乔什·弗洛姆、舒威塔克·帕特尔和丹尼尔·麦克达夫。用于设备上非接触式生命体征测量的多任务时间转移注意力网络。神经信息处理系统的进展, 33: 19400–19411, 2020。
- [7]于梓潼, 李小白, 王不超, 赵国英。Transrppg: 用于3D面具面部呈现攻击检测的远程光电体积描记Transformer。IEEE信号处理快报, 2021年。
- [8]丹尼尔麦克达夫, 哈维尔埃尔南德斯, 埃罗尔伍德, 刘欣, 和塔达斯Baltrusaitis。使用合成化身推进非接触式生命体征测量。arXiv预印本arXiv: 2010.12949, 2020。
- [9]Zhen Wang, Yunhao Ba, Pradyumna Chari, Oyku Deniz博兹库尔特, Gianna Brown, Parth Patwa, Niranjan Vaddi, Laleh Jalilian, and Achuta Kadambi.人脸视频的合成生成

体积描记生理学在IEEE/CVF计算机视觉和模式识别会议论文集, 第20587–20596页, 2022年。

- [10]Xin Liu, Yuntao Wang, 希南谢, Xiaoyu Zhang, Zixian Ma, 丹尼尔麦克达夫, 和Shwetak 帕特尔Mobilephys: 基于移动摄像头的个性化非接触式生理传感。Proc. ACM Interact.暴徒可穿戴无处不在技术, 6 (1), 2022年3月。
- [11]Xin Liu, Xin Jiang, Josh Fromm, Xuhai Xu, Shwetak Patel, and丹尼尔麦克达夫. Metaphys: 用于非接触式生理测量的少量调整。在健康, 推理和学习会议论文集, 第154–163页, 2021年。
- [12]Xin Liu, Mingchuan Zhang, Shwetak Patel, and丹尼尔麦克达夫.联邦远程生理测量数据不完善。在IEEE/CVF计算机视觉和模式识别会议论文集, 第2155–2164页, 2022年。
- [13]孙照东和李小白。造影: 基于视频的无监督远程生理检查
通过时空对比测量。欧洲计算机视觉会议, 第492–510页。斯普林格, 2022年。
- [14]约翰·吉迪恩和西蒙·斯坦特通往我心灵的道路是通过对比学习: 远程
来自未标记视频的光电体积描记术。IEEE/CVF计算机视觉国际会议论文集, 第3995–4004页, 2021年。
- [15]杰里米·斯佩思, 内森·万斯, 帕特里克·弗林, 亚当·查卡.非对比无监督
从视频中学习生理信号。在IEEE/CVF计算机视觉和模式识别会议论文集, 第14464–14474页, 2023年。

- [16] Yuzhe Yang, Xin Liu, Jiang Wu, Silviu Borac, Dina Katabi, Ming-Zher Poh, and Daniel McDuff. Simper: Simple self-supervised learning of periodic targets. *arXiv preprint arXiv:2210.03115*, 2022.
- [17] Daniel McDuff and Ethan Blackford. iphys: An open non-contact imaging-based physiological measurement toolbox. In *2019 41st Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*, pages 6521–6524. IEEE, 2019.
- [18] Giuseppe Boccignone, Donatello Conte, Vittorio Cuculo, Alessandro d’Amelio, Giuliano Grossi, and Raffaella Lanzarotti. An open framework for remote-ppg methods and their assessment. *IEEE Access*, 8:216083–216103, 2020.
- [19] Christian Pilz. On the vector space in photoplethysmography imaging. In *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pages 0–0, 2019.
- [20] Daniel McDuff, Sarah Gontarek, and Rosalind W. Picard. Remote Detection of Photoplethysmographic Systolic and Diastolic Peaks Using a Digital Camera. *IEEE Transactions on Biomedical Engineering*, 61(12):2948–2954, December 2014.
- [21] Giuseppe Boccignone, Donatello Conte, Vittorio Cuculo, Alessandro D’Amelio, Giuliano Grossi, Raffaella Lanzarotti, and Edoardo Mortara. pyvhr: a python framework for remote photoplethysmography. *PeerJ Computer Science*, 8:e929, 2022.
- [22] Serge Bobbia, Richard Macwan, Yannick Beneszeth, Alamin Mansouri, and Julien Dubois. Unsupervised skin tissue segmentation for remote photoplethysmography. *Pattern Recognition Letters*, 124:82–90, 2019.
- [23] Ronny Stricker, Steffen Müller, and Horst-Michael Gross. Non-contact video-based pulse rate measurement on a mobile service robot. In *The 23rd IEEE International Symposium on Robot and Human Interactive Communication*, pages 1056–1062. IEEE, 2014.
- [24] Daniel McDuff, Miah Wander, Xin Liu, Brian L Hill, Javier Hernandez, Jonathan Lester, and Tadas Baltrusaitis. Scamps: Synthetics for camera measurement of physiological signals. *arXiv preprint arXiv:2206.04197*, 2022.
- [25] Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, Daniel McDuff, and Xin Liu. Mmpd: Multi-domain mobile video physiology dataset, 2023.
- [26] Zheng Zhang, Jeff M Girard, Yue Wu, Xing Zhang, Peng Liu, Umar Ciftci, Shaun Canavan, Michael Reale, Andy Horowitz, Huiyuan Yang, et al. Multimodal spontaneous emotion corpus for human behavior analysis. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, pages 3438–3446, 2016.
- [27] Rita Meziati Sabour, Yannick Beneszeth, Pierre De Oliveira, Julien Chappe, and Fan Yang. Ubfc-phys: A multimodal database for psychophysiological studies of social stress. *IEEE Transactions on Affective Computing*, 2021.
- [28] Wim Verkruyse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Optics express*, 16(26):21434–21445, 2008.
- [29] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Advancements in noncontact, multiparameter physiological measurements using a webcam. *IEEE transactions on biomedical engineering*, 58(1):7–11, 2010.
- [30] Gerard De Haan and Vincent Jeanne. Robust pulse rate from chrominance-based rppg. *IEEE Transactions on Biomedical Engineering*, 60(10):2878–2886, 2013.
- [31] Wenjin Wang, Albertus C den Brinker, Sander Stuijk, and Gerard de Haan. Algorithmic principles of remote ppg. *IEEE Transactions on Biomedical Engineering*, 64(7):1479–1491, 2016.
- [32] Gerard De Haan and Arno Van Leest. Improved motion robustness of remote-ppg by using the blood volume pulse signature. *Physiological measurement*, 35(9):1913, 2014.

- [16]Yuzhe Yang, Xin Liu, Jiang Wu, Silviu Borac, Dina Katabi, Ming-Zher Poh, and 丹尼尔·麦克达夫。Simper: 周期性目标的简单自监督学习。arXiv预印本arXiv: 2210.03115, 2022。
- [17]丹尼尔·麦克达夫和伊桑·布莱克福德。iphys: 一种开放的基于非接触成像的生理测量工具箱在2019年第41届IEEE医学和生物学工程学会(EMBC)年度国际会议上, 第6521–6524页。IEEE, 2019年。
- [18]Giuseppe Boccignone, Donatello Conte, Vittorio Cucola, Alessandro d'Amelio, Giuliano Grossi, 意大利和拉斐尔·兰扎罗蒂远程ppg方法及其评估的开放框架。IEEE Access, 8: 216083—216103, 2020。
- [19]克里斯蒂安·皮尔兹。光电容积脉搏波成像中的向量空间。在IEEE/CVF计算机视觉研讨会国际会议论文集, 第0–0页, 2019年。
- [20]丹尼尔·麦克达夫, 莎拉Gontarek和罗莎琳德W.皮卡德光电容积描记器的远程检测使用数码相机的收缩和舒张峰图形。IEEE Transactions on Biomedical Engineering, 61 (12) : 2948–2954, 2014年12月。
- [21]Giuseppe Boccignone, Donatello Conte, Vittorio Cuacolo, Alessandro D'Amelio, Giuliano Grossi, Raffaella Lanzarotti, and Edoardo Mortara. Pyvhr: 一个用于远程光电体积描记的Python框架。PeerJ Computer Science, 8: e929, 2022。
- [22]Serge Bobbia, Richard Macwan, Yannick Benerezeth, Alamin Mandarin, and Julien Dubois. 用于远程光电体积描记术的无监督皮肤组织分割。Pattern Recognition Letters, 124: 82–90, 2019。
- [23]罗尼·史翠克, 斯蒂芬·穆勒, 霍斯特-迈克尔·格罗斯。非接触式视频脉率测量上的移动的服务机器人。第23届IEEE机器人与人类交互通信国际研讨会, 第1056–1062页。IEEE, 2014年。
- [24]丹尼尔·麦克达夫, 米亚·万德, 刘欣, 布赖恩·L·希尔, 哈维尔·埃尔南德斯, 乔纳森·莱斯特, 和塔达斯·巴尔特鲁赛提斯Spamps: 用于生理信号相机测量的合成器。arXiv预印本arXiv: 2206.04197, 2022。
- [25]Jiankai Tang, Kequan Chen, Yuntao Wang, Yuanchun Shi, Shwetak Patel, 丹尼尔·麦克达夫, Xin Liu. MMPD: 多域移动的视频生理数据集, 2023年。
- [26]Zheng Zhang, Jeff M吉拉德, Yue Wu, Xing Zhang, Peng Liu, Umur Ciftci, Shaun Canaway, 杨慧媛, 杨文, 等。多模态自发情绪语料库的研究。在IEEE计算机视觉和模式识别会议论文集, 第3438–3446页, 2016年。
- [27]Rita Meziati Sabour、Yannick Benerezeth、Pierre De Oliveira、Julien Chappe和Fan Yang。Uxfc-phys: 社会压力心理生理学研究的多模式数据库。IEEE Transactions on Affective Computing, 2021。
- [28]Wim Verkruyse, Lars O Svaasand和J Stuart纳尔逊。使用环境光的远程体积描记成像。Optics express, 16 (26) : 21434–21445, 2008。
- [29]Ming-Zher Poh, 丹尼尔J McDuff, Rosalind W Picard.非接触技术的进步, 使用网络摄像头进行多参数生理测量。IEEE生物医学工程学报, 58 (1) : 7–11, 2010年。
- [30]杰拉德·德哈安和文森特·珍妮。来自基于色度的rppg的稳健脉冲率。IEEE Transactions on Biomedical Engineering, 60 (10) : 2878–2886, 2013。
- [31]Wenjin Wang, Albertus C den Brinker, Sander Stuijk, 和Gerard de Haan。Algorithmic算法远程PPG的原理IEEE Transactions on Biomedical Engineering, 64 (7) : 1479–1491, 2016。
- [32]杰拉德·德哈安和阿诺·货车·莱斯特。通过使用血容量脉搏特征来提高远程ppg的运动鲁棒性。生理测量, 35 (9) : 1913, 2014。

- [33] Christian S Pilz, Sebastian Zaunseder, Jarek Krajewski, and Vladimir Blazek. Local group invariance for heart rate estimation from face videos in the wild. In *Proceedings of the IEEE conference on computer vision and pattern recognition workshops*, pages 1254–1262, 2018.
- [34] Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip HS Torr, and Guoying Zhao. Physformer: Facial video-based physiological measurement with temporal difference transformer. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pages 4186–4196, 2022.
- [35] Weixuan Chen and Daniel McDuff. Deepphys: Video-based physiological measurement using convolutional attention networks. In *Proceedings of the European Conference on Computer Vision (ECCV)*, pages 349–365, 2018.
- [36] Xin Liu, Brian Hill, Ziheng Jiang, Shwetak Patel, and Daniel McDuff. Efficientphys: Enabling simple, fast and accurate camera-based cardiac measurement. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pages 5008–5017, 2023.
- [37] Adam Paszke, Sam Gross, Francisco Massa, Adam Lerer, James Bradbury, Gregory Chanan, Trevor Killeen, Zeming Lin, Natalia Gimelshein, Luca Antiga, et al. Pytorch: An imperative style, high-performance deep learning library. In *Advances in neural information processing systems*, pages 8026–8037, 2019.
- [38] Diederik P Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.
- [39] Ilya Loshchilov and Frank Hutter. Decoupled weight decay regularization. *arXiv preprint arXiv:1711.05101*, 2017.
- [40] Leslie N Smith and Nicholay Topin. Super-convergence: Very fast training of neural networks using large learning rates. In *Artificial intelligence and machine learning for multi-domain operations applications*, volume 11006, pages 369–386. SPIE, 2019.
- [41] Girish Narayanswamy, Yujia Liu, Yuzhe Yang, Chengqian Ma, Xin Liu, Daniel McDuff, and Shwetak Patel. Bigsmall: Efficient multi-task learning for disparate spatial and temporal physiological measurements. *arXiv preprint arXiv:2303.11573*, 2023.
- [42] Daniel McDuff, Theodore Curran, and Achuta Kadambi. Synthetic data in healthcare. *arXiv preprint arXiv:2304.03243*, 2023.
- [43] Akshay Paruchuri, Xin Liu, Yulu Pan, Shwetak Patel, Daniel McDuff, and Soumyadip Sengupta. Motion matters: Neural motion transfer for better camera physiological sensing. *arXiv preprint arXiv:2303.12059*, 2023.
- [44] Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Morency. Openface 2.0: Facial behavior analysis toolkit. In *2018 13th IEEE international conference on automatic face & gesture recognition (FG 2018)*, pages 59–66. IEEE, 2018.
- [45] Danish Contractor, Daniel McDuff, Julia Katherine Haines, Jenny Lee, Christopher Hines, Brent Hecht, Nicholas Vincent, and Hanlin Li. Behavioral use licensing for responsible ai. In *2022 ACM Conference on Fairness, Accountability, and Transparency*, pages 778–788, 2022.
- [46] Vladimir Blazek, Ting Wu, and Dominik Hoelscher. Near-infrared ccd imaging: Possibilities for noninvasive and contactless 2d mapping of dermal venous hemodynamics. In *Optical Diagnostics of Biological Fluids V*, volume 3923, pages 2–9. International Society for Optics and Photonics, 2000.
- [47] Chihiro Takano and Yuji Ohta. Heart rate measurement based on a time-lapse image. *Medical engineering & physics*, 29(8):853–857, 2007.
- [48] Wim Verkruyse, Lars O Svaasand, and J Stuart Nelson. Remote plethysmographic imaging using ambient light. *Opt. Express*, 16(26):21434–21445, Dec 2008.
- [49] Ming-Zher Poh, Daniel J McDuff, and Rosalind W Picard. Non-contact, automated cardiac pulse measurements using video imaging and blind source separation. *Optics express*, 18(10):10762–10774, 2010.

- [33]Christian S Pilz、塞巴斯蒂安Zaunseder、Jarek Kraubernski和弗拉基米尔Blazek。本地组不变性的心率估计从人脸视频在野外。在IEEE计算机视觉和模式识别研讨会会议论文集, 第1254–1262页, 2018年。
- [34]Zitong Yu, Yuming Shen, Jingang Shi, Hengshuang Zhao, Philip HS Torr, and Guoying Zhao. Physformer: 基于面部视频的生理测量, 具有时间差转换
前者在IEEE/CVF计算机视觉和模式识别会议论文集, 第4186–4196页, 2022年。
- [35]Weixuan Chen和丹尼尔麦克达夫。Deepphys: 基于视频的生理测量,
卷积注意力网络在欧洲计算机视觉会议 (ECCV) 的会议记录中, 第349–365页, 2018年。
- [36]Xin Liu, Brian Hill, Jiang, Shwetak Patel, and丹尼尔麦克达夫. Efficientphys: 启用简单、快速和准确的基于相机的心脏测量。在IEEE/CVF计算机视觉应用冬季会议论文集, 第5008–5017页, 2023年。
- [37]亚当-帕斯克, 萨姆-格罗斯, 弗朗西斯科马萨, 亚当-莱勒, 詹姆斯-布拉德伯里, 格雷戈里-查南, 特雷弗基林, 林泽明, 娜塔莉亚Gimelshein, 卢卡Antiga, 等. Pytorch: An imperative style, high-performance deep learning library.神经信息处理系统的进展, 第8026–8037页, 2019年。
- [38]Diederik P Kingma和Jimmy Ba。Adam: 一种随机优化方法。arXiv预印本arXiv: 1412.6980, 2014年。
- [39]伊利亚·洛希洛夫和法兰克·哈特。解耦权重衰减正则化。arXiv预印本arXiv: 1711.05101, 2017。
- [40]Leslie N Smith和Nicholay Topin。超收敛: 神经网络的快速训练
使用大的学习率。在人工智能和机器学习的多域操作应用程序, 卷11006, 第369–386页。
SPIE, 2019年。
- [41]Girish Narayanswamy, Yujia Liu, Yuzhe Yang, Chengqian Ma, Xin Liu, 丹尼尔麦克达夫, 和瑞塔克·帕特尔Bigsrnal: 针对不同空间和时间生理测量的高效多任务学习。arXiv预印本arXiv: 2303.11573, 2023。
- [42]Daniel McDuff, Theodore Curran, and Achuta Kadambi. Synthetic data in healthcare. arXiv preprint arXiv:2304.03243, 2023.
- [43]Akshay Paruchuri, Xin Liu, Yulu Pan, Shwetak Patel, Daniel McDuff, and Soumyadip Sengupta. 运动至关重要: 神经运动传输可实现更好的相机生理感知。arXiv预印本arXiv: 2303.12059, 2023.
- [44]Tadas Baltrusaitis, Amir Zadeh, Yao Chong Lim, and Louis-Philippe Mod. Openface 2.0: 面部行为分析工具包。在2018年第13届IEEE自动人脸和手势识别国际会议 (FG 2018) 上, 第59–66页。IEEE, 2018年。
- [45]丹麦承包商, 丹尼尔麦克达夫, 朱莉娅凯瑟琳海恩斯, 珍妮李, 克里斯托弗海恩斯, 布伦特赫克特, 尼古拉斯文森特和李翰林。负责任人工智能的行为使用许可。2022年ACM公平, 问责制和透明度会议, 第778–788页, 2022年。
- [46]弗拉基米尔布拉泽克, 吴婷和多米尼克霍尔舍。近红外ccd成像: 可能用于皮肤静脉血流动力学的非侵入性和非接触式2D映射。在Optical Diagnostics of Biological Fluids V, 第3923卷, 第2–9页中。国际光学与光子学学会, 2000年。
- [47]高野千寻和太田裕二。基于延时图像的心率测量。医学工程与物理, 29 (8) : 853–857, 2007。
- [48]Wim Verkruyse, Lars O Svaasand和J Stuart纳尔逊。使用环境光的远程体积描记成像。Opt. Express, 16 (26) : 21434–21445, 2008年12月。
- [49]Ming-Zher Poh, 丹尼尔J McDuff, Rosalind W Picard.非接触式自动心脏搏动
测量使用视频成像和盲源分离。Optics express, 18 (10) : 10762–10774, 2010.

- [50] Antony Lam and Yoshinori Kuno. Robust heart rate measurement from video using select random patches. In *Proceedings of the IEEE International Conference on Computer Vision*, pages 3640–3648, 2015.
- [51] Shuchang Xu, Lingyun Sun, and Gustavo Kunde Rohde. Robust efficient estimation of heart rate pulse from video. *Biomedical optics express*, 5(4):1124–1135, 2014.
- [52] Qi Zhan, Wenjin Wang, and Gerard de Haan. Analysis of cnn-based remote-ppg to understand limitations and sensitivities. *Biomedical Optics Express*, 11(3):1268–1283, 2020.
- [53] Eugene Lee, Evan Chen, and Chen-Yi Lee. Meta-rppg: Remote heart rate estimation using a transductive meta-learner. In *European Conference on Computer Vision*, pages 392–409. Springer, 2020.
- [54] Xuesong Niu, Shiguang Shan, Hu Han, and Xilin Chen. Rhythmnet: End-to-end heart rate estimation from face via spatial-temporal representation. *IEEE Transactions on Image Processing*, 29:2409–2423, 2019.
- [55] Xuesong Niu, Zitong Yu, Hu Han, Xiaobai Li, Shiguang Shan, and Guo Ying Zhao. Video-based remote physiological measurement via cross-verified feature disentangling. In *European Conference on Computer Vision*, pages 295–310. Springer, 2020.
- [56] Rencheng Song, Huan Chen, Juan Cheng, Chang Li, Yu Liu, and Xun Chen. Pulsegan: Learning to generate realistic pulse waveforms in remote photoplethysmography. *IEEE Journal of Biomedical and Health Informatics*, 25(5):1373–1384, 2021.
- [57] Hao Lu, Hu Han, and S Kevin Zhou. Dual-gan: Joint bvp and noise modeling for remote physiological measurement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 12404–12413, 2021.
- [58] Yassine Ouzar, Djamaleddine Djeldjli, Frédéric Bousefsaf, and Choubeila Maaoui. X-ippgnet: A novel one stage deep learning architecture based on depthwise separable convolutions for video-based pulse rate estimation. *Computers in Biology and Medicine*, 154:106592, 2023.
- [59] Xudong Tan, Menghan Hu, Guangtao Zhai, Yan Zhu, Wenfang Li, and Xiao-Ping Zhang. Lightweight video-based respiration rate detection algorithm: An application case on intensive care. *IEEE Transactions on Multimedia*, 2023.
- [60] Hailan Kuang, Can Ao, Xiaolin Ma, and Xinhua Liu. Shuffle-rppgnet: Efficient network with global context for remote heart rate variability measurement. *IEEE Sensors Journal*, 2023.
- [61] Joaquim Comas, Adria Ruiz, and Federico Sukno. Efficient remote photoplethysmography with temporal derivative modules and time-shift invariant loss. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2182–2191, 2022.
- [62] Hao Wang, Euijoon Ahn, and Jinman Kim. Self-supervised learning framework for remote heart rate estimation using spatiotemporal augmentation. *arXiv preprint arXiv:2107.07695*, 2021.
- [63] Wei-Hao Chung, Cheng-Ju Hsieh, Sheng-Hung Liu, and Chiou-Ting Hsu. Domain generalized rppg network: Disentangled feature learning with domain permutation and domain augmentation. In *Proceedings of the Asian Conference on Computer Vision*, pages 807–823, 2022.
- [64] Hao Lu, Zitong Yu, Xuesong Niu, and Ying-Cong Chen. Neuron structure modeling for generalizable remote physiological measurement. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 18589–18599, 2023.
- [65] Fabian Schrumpf, Patrick Frenzel, Christoph Aust, Georg Osterhoff, and Mirco Fuchs. Assessment of deep learning based blood pressure prediction from ppg and rppg signals. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3820–3830, 2021.

- [50]安东尼林和久野良纪。使用select从视频中进行稳健的心率测量
随机补丁。在IEEE计算机视觉国际会议论文集, 第3640–3648页, 2015年。
- [51]Shuchang Xu, Lingyun Sun, and Gustavo Kunde Rohde.从视频中鲁棒有效地估计心率脉冲。
Biomedical Optics Express, 5 (4) : 1124–1135, 2014.
- [52]Qi Zhan, Wenjin Wang, and Gerard de哈安.分析基于cnn的远程ppg以了解其局限性和敏感性。
Biomedical Optics Express, 11 (3) : 1268–1283, 2020.
- [53]尤金李, 埃文陈, 和陈毅李。Meta-rppg: 使用远程心率估计
一个转换的元学习者。欧洲计算机视觉会议, 第392–409页。
斯普林格, 2020年。
- [54]牛雪松, Shiguang Shan, Hu Han, 和Xilin Chen。Rhythmnet: 端到端心率估计—
通过时空表示从面部获取信息。IEEE Transactions on Image Processing, 29: 2409–2423,
2019。
- [55]牛雪松, 余梓潼, 韩虎, 李小白, Shiguang Shan, 赵国英。视频—
通过交叉验证的特征解缠进行基于远程生理测量。欧洲计算机视觉会议, 第295–310页。斯普林
格, 2020年。
- [56]宋仁成、陈欢、程隽、常莉、刘玉、陈勋。Pulsegan:
学习在远程光电容积描记中生成逼真的脉搏波形。IEEE Journal of Biomedical and Health
Informatics, 25 (5) : 1373–1384, 2021。
- [57]Hao Lu, Hu Han, and S Kevin Zhou. Dual-Gan: 联合bvp和噪声建模用于远程
生理测量在IEEE/CVF计算机视觉和模式识别会议论文集, 第12404–12413页, 2021年。
- [58]Yassine Ouzar、Djamaeddine Djeldjli、Frédéric Bousefsaf和Choubeila Maaoui。X-ippgnet:
基于深度可分离卷积的一级深度学习架构, 用于基于视频的脉冲率估计。生物学和医学中的计算
机, 154: 106592, 2023。
- [59]Xudong Tan, Menghan Hu, Guangtao Zhai, Yan Zhu, Wenfang Li, and Xiao-Ping Zhang.
轻量级视频呼吸率检测算法: 重症监护应用案例。IEEE Transactions on Multimedia, 2023。
- [60]匡海兰, 敖灿, 马小林, 刘新华。Shuffle-rppgnet: 具有全局上下文的高效网络, 用于远程心
率变异性测量。IEEE传感器杂志, 2023年。
- [61]Joaquim Comas, Adria Ruiz和Federico Sukno。有效的远程光电体积描记术,
时间导数模块和时移不变损失。在IEEE/CVF计算机视觉和模式识别会议论文集, 第2182–2191
页, 2022年。
- [62]Hao Wang, Euijoon Ahn, and Jinman Kim.远程自监督学习框架
使用时空增强的心率估计。arXiv预印本arXiv: 2107.07695, 2021。
- [63]Wei-Hao Chung, Cheng-Ju Hsieh, Sheng-Hung Liu, Chiou-Ting Hsu.域广义化
rppg网络: 使用域置换和域增强的分解特征学习。
在亚洲计算机视觉会议论文集, 第807–823页, 2022年。
- [64]Hao Lu, Zitong Yu, Xuesong Niu, and Ying-Cong Chen.神经元结构建模
可推广的远程生理测量。在IEEE/CVF计算机视觉和模式识别会议论文集, 第18589–18599页,
2023年。
- [65]Fabian Schrumpf, Patrick Frenzel, Christoph Aust, Georg Osterhoff和Mirco Fuchs。AS—
从PPG和RPPG信号评估基于深度学习的血压预测。在IEEE/CVF计算机视觉和模式识别会议论文
集, 第3820–3830页, 2021年。

- [66] Theodore Curran, Xin Liu, Daniel McDuff, Shwetak Patel, and Eugene Yang. Camera-based remote photoplethysmography to measure heart rate and blood pressure in ambulatory patients with cardiovascular disease: Preliminary analysis. *Journal of the American College of Cardiology*, 81(8_Supplement):2301–2301, 2023.
- [67] Ying-Ying Zheng, Yi-Tong Ma, Jin-Ying Zhang, and Xiang Xie. Covid-19 and the cardiovascular system. *Nature Reviews Cardiology*, 17(5):259–260, 2020.
- [68] Mauricio Villarroel, Sitthichok Chaichulee, João Jorge, Sara Davis, Gabrielle Green, Carlos Arteta, Andrew Zisserman, Kenny McCormick, Peter Watkinson, and Lionel Tarassenko. Non-contact physiological monitoring of preterm infants in the neonatal intensive care unit. *npj Digital Medicine*, 2(1):1–18, 2019.

[66]西奥多柯伦、刘欣、丹尼尔麦克达夫、施韦塔克帕特尔和尤金杨。基于相机远程光电容积描记术测量非卧床心血管疾病患者的心率和血压：初步分析。美国心脏病学会杂志, 81 (8_增刊) : 2301–2301, 2023。

[67]郑莹莹, 马一彤, 张金英, 谢翔。2019冠状病毒病和心血管系统。Nature Reviews Cardiology, 17 (5) : 259–260, 2020。

[68]Mauricio Villarroel, Sitthichok Chaichulee, João Jorge, Sara Davis, Gabrielle绿色, 卡洛斯阿尔特塔, 安德鲁·齐瑟曼, 肯尼·麦考密克, 彼得·沃金森, 和莱昂内尔·塔拉森科。新生儿重症监护室中早产儿的非接触式生理监测。npj数字医学, 2 (1) : 1–18, 2019。

A Overview of Appendices

Our appendices contain the following additional details and results:

- In Section B, we provide details regarding the background of rPPG and an overview of existing methods.
- In Section C, we provide an overview of potential applications out of rPPG technologies.
- In Section D, we provide additional results about PhysNet.
- In Section E, we provide an overview of rPPG network recommendations.
- In Section F we provide details toward metrics supported by our toolbox. We also provide additional metric results in Section G that were not included in the main paper due to space constraints.
- Section H briefly details which subjects we utilized for exclusion, or conversely sub-selection, in each task when dealing with the UBFC-Phys [27] dataset. We also briefly describe video filtering criteria available via the toolbox and useful for subject sub-selection.
- Additional details related to training and evaluation for physiological multitasking is shared in Section I.
- Section J briefly describes additional features included in the toolbox. These features, including pre-processed data visualization, loss and learning visualization, Bland-Altman plots, and motion analysis, are further detailed with exemplar usage in the rPPG-Toolbox’s GitHub repo: <https://github.com/ubicomplab/rPPG-Toolbox>

B Background and Existing Research in rPPG

B.1 Background

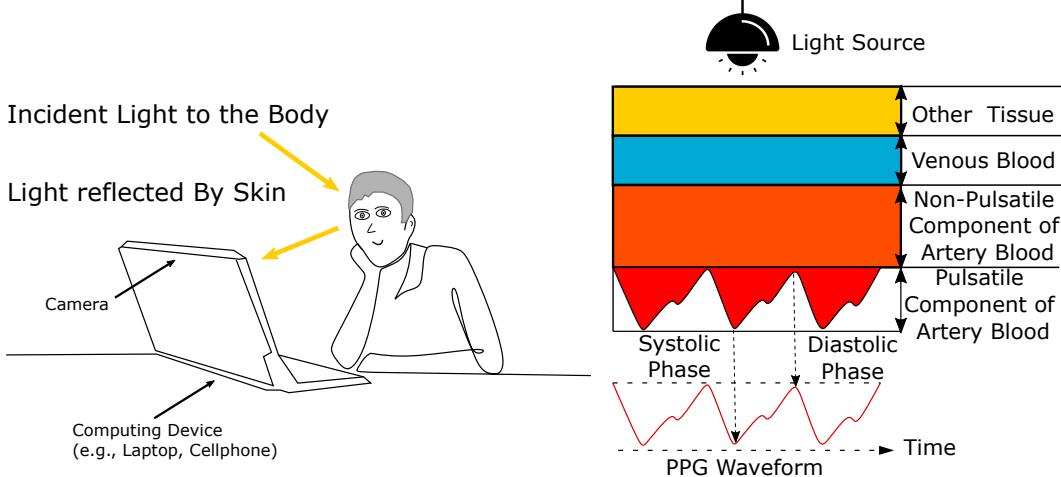


Figure 4: The principles behind camera-based physiological sensing. The volumetric changes of blood under the surface of the skin cause changes in light absorption and reflection, which is the source of PPG signal.

Remote PPG (rPPG) measurement involves the development of computational methods for extracting physiological parameters (e.g., pulse rate, respiration rate, blood oxygenation, blood pressure) based on light reflected from, or transmitted through, the human body. Essentially these methods use pixel information to quantify changes in visible light, or other electromagnetic radiation (e.g., infrared or thermal), that are modulated by blood flow in the periphery of the skin (see Fig.4). This reflected radiation is also affected by body motions and absorption characteristics of the skin [46, 47, 48, 49, 50, 51, 31]. In this article, we will focus primarily on the use of visible light, due to the ubiquitous nature of RGB cameras.

As visible light penetrates between 4 to 5 mm below the skin’s surface, it is modulated by the volume of oxygenated and deoxygenated hemoglobin enabling the measurement of the peripheral blood

附录概述

我们的附录包含以下附加细节和结果：

- 在第B节中，我们提供了有关rPPG背景的详细信息和现有方法的概述。
- 在C部分中，我们概述了rPPG技术的潜在应用。
- 在D部分，我们提供了关于PhysNet的其他结果。
- 在E节中，我们提供了rPPG网络建议的概述。
- 在第F节中，我们提供了工具箱支持的指标的详细信息。我们还在G节中提供了由于篇幅限制而未包含在主论文中的其他度量结果。
- H节简要介绍了在处理UBFC-Phys [27]数据集时，我们在每个任务中用于排除或相反地进行子选择的受试者。我们还简要介绍了视频过滤标准，可通过工具箱和有用的主题子选择。
- 有关生理多任务处理的培训和评估的更多详细信息，请参见第I节。
- J节简要介绍了工具箱中包含的其他功能。这些功能，包括预处理数据可视化，损失和学习可视化，Bland-Altman图和运动分析，在rPPG-100的GitHub存储库中有更多的详细说明
<https://github.com/ubicomplab/rPPG-Toolbox>

rPPG的B背景和现有研究

B.1 背景

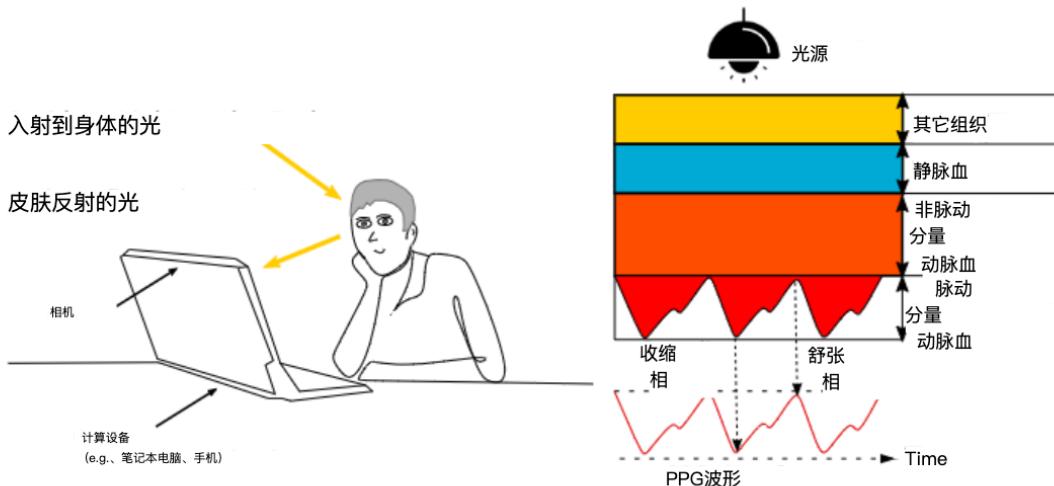


图4：基于摄像头的生理传感原理。皮肤表面下血液的体积变化引起光吸收和反射的变化，这是PPG信号的来源。

远程PPG (rPPG) 测量涉及用于提取生理参数（例如，脉搏率、呼吸率、血氧、血压）。基本上这些方法使用像素信息来量化可见光或其他电磁辐射（例如，红外线或热），其由皮肤周围的血流调节（见图4）。这种反射的辐射也受到身体运动和皮肤吸收特性的影响[46, 47, 48, 49, 50, 51, 31]。在本文中，我们将主要关注可见光的使用，因为RGB相机无处不在。

当可见光穿透皮肤表面以下4至5 mm时，它受到氧合血红蛋白和脱氧血红蛋白体积的调制，从而能够测量外周血

volume pulse (BVP). The frequency channels offered by multiband (e.g., RGB) cameras enable the composition of blood, including the oxygen saturation to be measured. In addition, these pixels are affected by the motion as a person breathes in and out and by the mechanical effects of the heart beating, enabling the measurement of breathing signals and the ballistocardiogram (BCG). Analyzing the morphology of these signals, and combining them together, offers the possibility of measuring correlates of blood pressure.

B.2 Optical Principles

The Lambert-Beer law (LBL) [50, 51, 35] and Shafer’s dichromatic reflection model (DRM) [31] are two models which provide a framework for capturing the effects of an imager, lighting, body motions and physiological processes on recorded pixel intensities. Given the optical characteristics of oxygenated and deoxygenated blood, we also have priors on the wavelengths of light that contain the strongest or weakest pulsatile information. This prior knowledge is important for measuring physiological parameters accurately. Most computational methods are built upon this grounding.

B.3 Algorithms

Many computational approaches for recovering physiological signals from videos have similar steps. The first step typically involves localizing a region of interest within each video frame. In a large majority of cases the face or head are the region of interest and therefore facial detection and/or landmark detection are used. However, in other cases skin segmentation might be preferred. Aggregating pixels spatially is a subsequent step that has been used to help to reduce noise from camera quantization errors. The operation can be performed by downsampling an image [35] or simply averaging all pixel values with a region of interest [49]. Many cameras capture frames from more than one frequency band (e.g., RGB) that provide complementary measurements to capture different properties of the light reflected from the body. This information can be used in two ways: 1) for understanding the composition of the blood (e.g., oxygen saturation), 2) improving the signal-to-noise ratio of the recovered blood volume pulse. Typically, computational methods leverage multiple bands and learn a linear or non-linear signal decomposition to estimate the pulse waveform. This manipulation of the color channel signals can be grounded in the optical properties of the skin [31, 30] or learned in a data-driven manner given a specialized learning criteria or loss function.

More recently supervised machine learning has become the most popular approach. Specifically, deep learning and convolutional neural networks provide the current state-of-the-art results. These methods present the opportunity for more “end-to-end” learning and researchers have gradually tried to replace handcrafted processing steps with learnable components. Since the relationship between underlying physiological signal and skin pixels in a video is complicated, deep neural networks have shown superior performance on modeling such non-linear relationship compared to traditional source separation methods [35, 5, 52, 53, 6, 54, 55, 56, 57, 34]. Moreover, due to the flexibility of neural network, researchers have also explored neural based methods for real-time on-device inference [6, 36, 58, 59, 60, 61], self-supervised learning [14, 53, 62, 15, 13, 16], domain generalization [11, 63, 64] and estimating new vital measurement such as blood pressure [65, 66].

C Potential Applications of rPPG

The SARS-CoV-2 (COVID-19) pandemic has accelerated the pace of change in healthcare services. In particular, how healthcare services are delivered around the world has needed to be rethought in the presence of new risks to patients and providers and restrictions on travel. The virus has been linked to the increased risk of cardiopulmonary (heart and lung related) illness with symptoms such as respiratory distress syndrome, myocarditis, and the associated chronic damage to the cardiovascular system. Experts suggest that particular attention should be given to cardiovascular protection during treatment of COVID-19 [67]. While measurement is not the sole solution to these problems, they have acutely highlighted the need for scalable and accurate physiological monitoring. Ubiquitous or pervasive health sensing technology could help patients conduct daily screenings, monitor the effects of medication on their symptoms, and help clinicians make more informed and accurate decisions.

The potential advantages that video-based contactless measurement offers have helped to draw a significant amount of attention to the field in recent years. Contact biomedical sensors (e.g.,

容积脉搏 (BVP)。由多频带提供的频率信道（例如，RGB）摄像机能够测量血液的成分，包括要测量的氧饱和度。此外，这些像素受到人吸气和呼气时的运动以及心脏跳动的机械效应的影响，从而能够测量呼吸信号和心冲击图 (BCG)。分析这些信号的形态，并将它们组合在一起，提供了测量血压相关性的可能性。

B.2 光学原理

Lambert–Beer定律 (LBL) [50, 51, 35]和Shafer的二色反射模型 (DRM) [31]是两个模型，它们提供了一个框架，用于捕获成像仪，照明，身体运动和生理过程对记录的像素强度的影响。考虑到含氧和脱氧血液的光学特性，我们也有关于包含最强或最弱脉动信息的光的波长的先验知识。这种先验知识对于准确测量生理参数是重要的。大多数计算方法都建立在这个基础上。

B.3 算法

用于从视频恢复生理信号的许多计算方法具有类似的步骤。第一步通常涉及定位每个视频帧内的感兴趣区域。在大多数情况下，面部或头部是感兴趣区域，因此使用面部检测和/或界标检测。然而，在其他情况下，皮肤分割可能是优选的。在空间上聚合像素是已经用于帮助减少来自相机量化误差的噪声的后续步骤。该操作可以通过对图像进行下采样[35]或简单地对感兴趣区域的所有像素值进行平均[49]来执行。许多相机从多于一个频带（例如，RGB），其提供互补测量以捕获从身体反射的光的不同性质。该信息可以以两种方式使用：1) 用于了解血液的成分（例如，氧饱和度），2) 改善恢复的血容量脉搏的信噪比。通常，计算方法利用多个频带并学习线性或非线性信号分解以估计脉搏波形。颜色通道信号的这种操纵可以基于皮肤的光学特性[31, 30]，或者在给定专门的学习标准或损失函数的情况下以数据驱动的方式学习。

最近，监督机器学习已经成为最流行的方法。具体来说，深度学习和卷积神经网络提供了当前最先进的结果。这些方法为更多的“端到端”学习提供了机会，研究人员逐渐尝试用可学习的组件取代手工制作的处理步骤。由于视频中潜在的生理信号和皮肤像素之间的关系是复杂的，与传统的源分离方法相比，深度神经网络在建模这种非线性关系方面表现出上级性能[35, 5, 52, 53, 6, 54, 55, 56, 57, 34]。此外，由于神经网络的灵活性，研究人员还探索了基于神经的方法，用于实时设备上的推理[6, 36, 58, 59, 60, 61]，自监督学习[14, 53, 62, 15, 13, 16]，域泛化[11, 63, 64]和估计新的生命测量，如血压[65, 66]。

C rPPG的潜在应用

SARS-CoV-2 (COVID-19) 疫情加快了医疗服务的变革步伐。特别是，在患者和提供者面临新风险以及旅行限制的情况下，需要重新考虑如何在世界各地提供医疗保健服务。该病毒与心肺（心脏和肺相关）疾病的风险增加有关，症状包括呼吸窘迫综合征、心肌炎和相关的心血管系统慢性损伤。专家建议，在治疗COVID-19期间应特别注意心血管保护[67]。虽然测量不是解决这些问题的唯一方法，但它们突出了对可扩展和准确的生理监测的需求。无处不在或无处不在的健康传感技术可以帮助患者进行日常筛查，监测药物对其症状的影响，并帮助临床医生做出更明智和准确的决策。

近年来，基于视频的非接触式测量提供的潜在优势帮助吸引了该领域的大量关注。接触式生物医学传感器（例如，

electrocardiograms, pulse oximeters) are the standard used for clinical screening and at-home measurement. However, these devices are usually bulky and are still not ubiquitously available, especially in low-resource settings. On the other hand, non-contact camera-based physiological sensing presents a new opportunity for highly scalable and low-cost physiological monitoring through ordinary cameras (e.g., webcams or smartphone cameras) [29]. Besides the convenience and potential scalability, this technology could also reduce the risk of infection for vulnerable patients and discomfort caused by obtrusive leads and electrodes [68]. Finally, we believe there are two specifically compelling advantages of cameras over contact sensors. The first, is that they can capture multi-modal signals, including but not limited to, the activity of the subject, their appearance, facial expressions and gestures, motor control and context. One reason this helps is that the physiological measurements can be interpreted in context. For example, if someone appears in pain, an elevated heart rate can be interpreted differently than without in pain. Secondly, cameras are spatial sensors allowing for the measurement of signals from multiple parts of the body to be measured concomitantly, presenting greater opportunities for characterizing vascular parameters such as pulse transit time.

We would also argue that camera-based physiological sensing could be an influential technology in telehealth. Current telehealth procedures are mainly telephone or video-based communication services where patients see their physician or healthcare provider via Cisco Webex, Zoom or Microsoft Teams. Performing clinical visits at home increases the efficiency of clinical visits and helps people who live in remote locations. There is still a debate over whether high-quality care can be delivered over telehealth platforms. One of notable issues with current telehealth systems is that there is no way for physicians to assess patient’s physiological states. The development of accurate and efficient non-contact camera-based physiological sensing technology would provide remote physicians access to the physiological data to make more informed clinical decisions.

D Investigation of PhysNet

Table 6: **PhysNet Ablation Results.** Investigation of variants of PhysNet models on the PURE [23] and UBFC-rPPG [22] datasets obtained using the rPPG toolbox.

Method	Train Set	Test Set			
		PURE [23]		UBFC-rPPG [22]	
		MAE \downarrow	MAPE \downarrow	MAE \downarrow	MAPE \downarrow
PHYSNET-Raw [5]	UBFC-RPPG	19.25	33.75	N/A	N/A
	PURE	N/A	N/A	11.17	11.64
	SCAMPS	18.40	31.74	10.57	11.04
PHYSNET-DIFFNORM [6]	UBFC-RPPG	8.06	13.67	N/A	N/A
	PURE	N/A	N/A	0.98	1.12
	SCAMPS	13.30	20.10	5.40	5.42
PHYSNET-RAW-TUNED [35]	UBFC-RPPG	4.80	8.46	N/A	N/A
	PURE	N/A	N/A	1.99	1.86
	SCAMPS	14.35	23.00	10.33	10.71

Metrics explained: MAE = Mean Absolute Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%).

Based on the rPPG community’s help and suggestions, we found that there are two important implementation details missed in the original PhysNet paper: 1) the raw input range has to be set to 0-1; 2) adding normalization to the model output. However, with these changes and the raw input format, PhysNet is not able to achieve results that are close to SOTA in UBFC and PURE. As shown in Table 6’s PhysNet-Raw-Tuned, we further fine-tuned the training parameters (changed learning rate from 0.009 to 0.09) and post-processing bandpass filter parameters (set bandpass filtering parameters to [0.5,2.5], detrend value to 200). However, we don’t recommend using these network specified parameters as it is not a fair comparison across different network architectures. This toolbox aims to provide a standardized training regime across all the networks and datasets. All in all, we recommend using DiffNorm frames as the input to PhysNet to make training easier to converge.

We also notice that results while training on SCAMPS and testing on PURE are not good. It is worth noting that SCAMPS is a synthetic dataset and can easily cause overfitting with a complex network architecture. Unlike other 2D-CNN based networks, PhysNet is a 3D-CNN based network which is easier to overfit simple datasets.

心电图、脉搏血氧计)是用于临床筛查和家庭测量的标准。然而,这些设备通常体积庞大,并且仍然不是普遍可用的,特别是在低资源环境中。另一方面,基于非接触式相机的生理感测为通过普通相机(例如,网络摄像头或智能手机摄像头)[29]。除了方便性和潜在的可扩展性外,该技术还可以降低脆弱患者的感染风险以及由突出的电极导线和电极引起的不适[68]。最后,我们认为相机相对于接触式传感器有两个特别引人注目的优势。首先,它们可以捕获多模态信号,包括但不限于受试者的活动,他们的外观,面部表情和手势,运动控制和上下文。这有帮助的一个原因是生理测量可以在上下文中解释。例如,如果有人出现疼痛,心率升高的解释可能与没有疼痛不同。其次,摄像机是空间传感器,允许同时测量来自身体多个部位的信号,为表征血管参数(如脉搏传导时间)提供了更大的机会。

我们还认为,基于摄像头的生理传感可能是远程医疗中一项有影响力的技术。目前的远程医疗程序主要是基于电话或视频的通信服务,患者可以通过Cisco Webex、Zoom或Microsoft Teams与医生或医疗保健提供者见面。在家中进行临床访问可以提高临床访问的效率,并帮助居住在偏远地区的人们。关于是否可以通过远程保健平台提供高质量的护理,仍存在争议。当前远程医疗系统的一个显著问题是,医生无法评估患者的生理状态。准确和有效的非接触式基于相机的生理传感技术的发展将提供远程医生访问的生理数据,以作出更明智的临床决策。

PhysNet的三维研究

表6: PhysNet消融结果。研究使用rPPG工具箱获得的PURE [23]和UBFC-rPPG [22]数据集上的PhysNet模型变体。

		测试集 [22]第二十二话					
	Method	train	set	mae	mamae	maape	
PN-R[5]	UBFC-PPG	19.25	33.75	N/A	N/A		
	纯N/A	N/A	11.17	11.64	SCAMPS	18.40	31.74
PN-DN[6]	11.04						10.57
	UBFC-PPG	8.06	13.67	N/A	N/A		
PURE	纯N/A	N/A	0.98	1.12			
	SCAMPS	13.30	20.10	5.40	5.42		
UBFC-PPG	4.80	8.46	N/A	N/A			
	N/A	N/A	[35]第918页	CAMPS	14.35	23.00	10.33
解释: MAE = HR估计的平均绝对误差(心跳/分钟), MAPE = 平均百分比误差(%)。							

基于rPPG社区的帮助和建议,我们发现原始PhysNet论文中遗漏了两个重要的实现细节:1)原始输入范围必须设置为0-1;2)将归一化添加到模型输出。然而,由于这些变化和原始输入格式,PhysNet无法在UBFC和PURE中实现接近SOTA的结果。如表6的PhysNet-Raw-Tuned所示,我们进一步微调了训练参数(将学习率从0.009更改为0.09)和后处理带通滤波器参数(将带通滤波参数设置为[0.5, 2.5],去趋势值为200)。但是,我们不建议使用这些网络指定的参数,因为它不是不同网络架构之间的公平比较。该工具箱旨在为所有网络和数据集提供标准化的培训制度。总而言之,我们建议使用DiffNorm帧作为PhysNet的输入,以使训练更容易收敛。

我们还注意到,在SCAMPS上训练和在PURE上测试的结果并不好。值得注意的是,SCAMPS是一个合成数据集,很容易导致复杂网络架构的过拟合。与其他基于2D-CNN的网络不同,PhysNet是一个基于3D-CNN的网络,更容易过拟合简单的数据集。

E Network Recommendation

In the dynamic realm of rPPG research, architectural choices critically influence the balance between efficiency and accuracy. This section elucidates our findings and offers recommendations on architectures, factoring both computational aspects and performance nuances.

- **Mobile and Computational Efficiency:** For scenarios demanding computational frugality, such as mobile or edge deployments, 2D-CNN based architectures are recommended. They gracefully balance performance and computational overhead, ensuring expeditious responses in resource-limited environments.
- **High-Performance Scenarios:** 3D-CNN or Transformer-based networks, while resource-intensive, offer superior performance. These architectures are apt for applications with lenient resource constraints where the pinnacle of accuracy is sought.
- **Performance Saturation:** Notably, a saturation trend was observed across several network architectures. This insinuates diminishing performance returns with increased complexity or depth, emphasizing the need to pragmatically select architectures based on actual task necessities.

A salient lesson from our research emphasizes the indispensability of diverse dataset incorporation. The rPPG community grapples with challenges such as: 1) motion-infused videos, introducing substantial rPPG signal noise; 2) videos under different lighting conditions (e.g., natural light, LED, incandescent), affecting visual cues critical for models; 3) featuring darker-skinned individuals, historically underrepresented, leading to potential model biases. To bolster robustness and cater to diverse real-world scenarios, it's imperative to amalgamate videos encapsulating the aforementioned conditions in training sets. Such holistic training paradigms ensure the model's aptitude in handling a vast spectrum of real-world challenges.

F Metric Details

F.1 rPPG Metrics

We present explanations of metrics supported by our toolbox below.

Mean Absolute Error (MAE): For predicted signal rate R_p , ground truth signal rate R_g , and for N instances:

$$MAE = \frac{1}{N} \sum_{n=1}^N |R_g - R_p|$$

Root Mean Square Error (RMSE): For predicted signal rate R_p , ground truth signal rate R_g , and for N instances:

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (R_g - R_p)^2}$$

Mean Absolute Percentage Error (MAPE): For predicted signal rate R_p , ground truth signal rate R_g , and for N instances:

$$MAPE = \frac{1}{N} \sum_{n=1}^N \left| \frac{R_g - R_p}{R_g} \right|$$

Pearson Correlation (ρ): For predicted signal rate R_p , ground truth signal rate R_g , and for N instances, and \bar{R} the average of R for N samples:

$$\rho = \frac{\sum_{n=1}^N (R_{g,n} - \bar{R}_g)(R_{p,n} - \bar{R}_p)}{\sqrt{\left(\sum_{n=1}^N R_{g,n} - \bar{R}_g \right)^2 \left(\sum_{n=1}^N R_{p,n} - \bar{R}_p \right)^2}}$$

E网推荐

在rPPG研究的动态领域中，架构选择对效率和准确性之间的平衡有着至关重要的影响。本节阐述了我们的研究结果，并提供了有关架构的建议，同时考虑了计算方面和性能方面的细微差别。

- 移动的和计算效率：对于要求计算节俭的场景，例如移动的或边缘部署，建议使用基于2D-CNN的架构。它们优雅地平衡了性能和计算开销，确保在资源有限的环境中快速响应。
- 高性能场景：3D-CNN或基于Transformer的网络虽然资源密集，但可提供上级性能。这些架构适用于具有宽松资源约束的应用程序，其中寻求精确度的顶峰。
- 性能饱和：值得注意的是，在几种网络架构中观察到饱和趋势。这意味着随着复杂性或深度的增加，性能回报会减少，强调需要根据实际任务的需要务实地选择架构。

从我们的研究中得到的一个突出教训强调了多样化数据集整合的不可或缺性。rPPG社区努力应对挑战，例如：1) 运动注入视频，引入大量rPPG信号噪声；2) 不同照明条件下的视频（例如，自然光、LED、白炽灯），影响对模型至关重要的视觉线索；3) 以肤色较深的个体为特征，历史上代表性不足，导致潜在的模型偏见。为了增强鲁棒性并满足不同的现实场景，必须将视频合并到训练集中，将上述条件封装起来。这种全面的培训模式确保了模型在处理大量现实世界挑战方面的能力。

F指标详细信息

F.1 rPPG ®

我们在下面介绍了我们的工具箱支持的指标的解释。

平均绝对误差 (MAE)：对于预测信号速率R_t、地面实况信号速率R_t，以及对于N个实例：

$$AE = AE \text{ 的 } \frac{1}{N} \sum_{n=1}^N |R_t - R_{gt}|$$

均方根误差 (RMSE)：对于预测信号速率R_t、地面实况信号速率R_t，并且对于N个实例：

$$RMSE = \sqrt{\frac{1}{N} \sum_{n=1}^N (R_t - R_{gt})^2}$$

平均绝对百分比误差 (MAPE)：对于预测信号速率R_t、地面实况信号速率R_t，以及对于N个实例：

$$MAPE = \frac{1}{N} \sum_{n=1}^N \frac{|R_t - R_{gt}|}{R_{gt}} \times 100\%$$

Pearson Correlation (ρ)：对于预测信号速率R_t、地面真实信号速率R_t，以及对于N个实例，以及R是N个样本的R的平均值：

$$\rho = \frac{\frac{P}{n=1} \frac{R_t - \bar{R}}{R - \bar{R}} \frac{R_{gt} - \bar{R}}{R - \bar{R}}}{\sqrt{\frac{P}{n=1} \frac{(R_t - \bar{R})^2}{R - \bar{R}} \frac{(R_{gt} - \bar{R})^2}{R - \bar{R}}}} = \frac{P}{\sqrt{S}}$$

Signal-to-Noise Ratio (SNR): As in [30], we calculate the Signal-to-Noise Ratio (SNR) for a predicted signal as the ratio between the area under the curve of the power spectrum around the first and second harmonic of the ground truth heart rate frequency and the area under the curve of the rest of the power spectrum. This is mathematically represented as follows:

$$SNR = 10 \log_{10} \left(\frac{\sum_{45}^{150} (U_t(f)S(f))^2}{\sum_{45}^{150} ((1 - U_t(f))S(f))^2} \right)$$

Where S is the power spectrum of the estimated rPPG signal. $U_t(f)$ is equal to 1 around the first and second harmonics of the ground truth rPPG signal, while being 0 elsewhere in the power spectrum. In the context of the rPPG-Toolbox, only the power spectrum between 0.75 Hz and 2.5 Hz, or 45 beats/min and 150 beats/min, is considered. We report the mean of the SNR values calculated per video or test sample, such that:

$$MSNR = \frac{1}{N} \sum_{n=1}^N SNR$$

Standard Error ($\pm SE$): The standard error is a measure of the statistical accuracy of an estimate, such as the mean, and is equal to the standard deviation of the theoretical distribution of a large population of such estimates. The standard error takes into account the number of samples utilized in measurement, which is especially useful in the case of remote PPG datasets where the number of test samples can vary significantly from dataset to dataset. For all metrics except for the Pearson correlation (ρ), we calculate the standard error as:

$$SE = \frac{\sigma}{\sqrt{n}}$$

Where σ is the standard deviation and n is the number of samples. For the Pearson correlation (ρ), the standard error is calculated as:

$$SE_\rho = \sqrt{\frac{1 - r^2}{n - 2}}$$

Where r is the correlation coefficient and n is the number of samples. Similar to how a standard deviation is reported, we report standard error as $\pm SE$.

F.2 Additional Multitask Metrics

We present explanations of additional metrics added to evaluate the BigSmall [41] model in order to exemplify how this toolbox can be extended to support physiological multitasking.

Evaluated Action Units (AU): Similar to [41], and other AU literature, facial action metrics are calculated for the following 12 commonly used AUs: AU01, AU02, AU04, AU06, AU07, AU10, AU12, AU14, AU15, AU17, AU23, AU24.

F1: The harmonic mean of recall and precision. For true positive count TP , false positive count FP , and false negative count FN .

$$F1 = 100 * \frac{2TP}{2TP + FP + FN}$$

Precision (Prec.): For true positive count TP , and false positive count FP :

$$Precision = 100 * \frac{TP}{TP + FP}$$

Accuracy (Acc.) For true positive count TP , true negative count TN , false positive count FP , and false negative count FN :

$$Accuracy = 100 * \frac{TP + TN}{TP + TN + FP + FN}$$

信噪比 (SNR) : 与[30]中一样, 我们计算预测信号的信噪比 (SNR), 作为地面真实心率频率的一次和二次谐波周围的功率谱曲线下的面积与功率谱其余部分曲线下的面积之间的比率。这在数学上表示如下:

$$SNR = 10 \log \frac{P}{P} \frac{\int_{f_1}^{f_2} (U(f) S(f))}{\int_{f_1}^{f_2} ((1 - U(f)) S(f))}$$

其中S是估计的rPPG信号的功率谱。U(f) 在真实rPPG信号的第一和第二谐波附近等于1, 而在功率谱中的其他地方为0。在rPPG-EEG的上下文中, 仅考虑0.75 Hz和2.5 Hz之间或45次/min和150次/min之间的功率谱。我们报告每个视频或测试样本计算的SNR值的平均值, 使得:

$$M SNR = \frac{1}{N} \sum_{n=1}^N SNR_n$$

标准误差 ($\pm SE$) : 标准误是一个估计的统计准确性的度量, 如平均值, 等于大量此类估计的理论分布的标准差。标准误差考虑了测量中使用的样本数量, 这在远程PPG数据集的情况下特别有用, 其中测试样本的数量可以在数据集之间显著变化。对于除Pearson相关性 (ρ) 之外的所有指标, 我们计算标准误为:

$$SE = \sqrt{\frac{\sigma^2}{n}}$$

其中 σ 是标准差, n 是样本数量。对于皮尔逊相关系数 (ρ), 标准误差计算如下:

$$SE = \sqrt{\frac{r}{n-2}}$$

其中 r 是相关系数, n 是样本数。与报告标准差的方式类似, 我们将标准误差报告为 $\pm SE$ 。

F.2 额外的多任务处理

我们解释了为评估BigSmall [41]模型而添加的其他指标, 以说明如何扩展此工具箱以支持生理多任务处理。

评价的动作单位 (Au) : 与[41]和其他Au文献相似, 计算以下12个常用AU的面部动作指标: AU01、AU02、AU04、AU06、AU07、AU10、AU12、AU14、AU15、AU17、AU23、AU24。

F1: 查全率和查准率的调和平均值。对于真阳性计数TP、假阳性计数FP和假阴性计数FN。

$$F1 = \frac{2TP}{2TP + FP + FN}$$

精密度 (精密度) : 对于真阳性计数TP和假阳性计数FP:

$$P_{精度} = \frac{TP}{TP + FP}$$

准确度 (Acc.) 对于真阳性计数TP、真阴性计数TN、假阳性计数FP和假阴性计数FN:

$$精度 = \frac{TP + TN}{TP + TN + FP}$$

G Additional Results

We reiterate results provided in the main paper and present additional results including the RMSE, SNR, Pearson correlation, and the corresponding standard errors. Note that there may be minor differences between results in the following tables and the main paper, as they were generated on a different machine using the latest version of the rPPG-Toolbox.

Table 7: Benchmark Results. Performance on the PURE [23] and UBFC-rPPG [22] datasets obtained using the rPPG toolbox. For the supervised methods we show cross-dataset training results using the UBFC-rPPG, PURE, and SCAMPS datasets.

		Test Set PURE [23]						
	Method	Train Set	MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow	
UNSUPERVISED	GREEN [28]	N/A	10.09 \pm 2.81	23.85 \pm 217.81	10.28 \pm 2.33	0.34 \pm 0.12	-2.66 \pm 1.43	
	ICA [29]	N/A	4.77 \pm 2.08	16.07 \pm 153.84	4.47 \pm 1.65	0.72 \pm 0.09	5.24 \pm 1.77	
	CHROM [30]	N/A	5.77 \pm 1.79	14.93 \pm 81.53	11.52 \pm 3.75	0.81 \pm 0.08	4.58 \pm 0.85	
	LGI [33]	N/A	4.61 \pm 1.91	15.38 \pm 134.14	4.96 \pm 1.72	0.77 \pm 0.08	4.50 \pm 1.21	
	PBV [32]	N/A	3.92 \pm 1.61	12.99 \pm 123.60	4.84 \pm 1.49	0.84 \pm 0.07	2.30 \pm 1.31	
	POS [31]	N/A	3.67 \pm 1.46	11.82 \pm 66.87	7.25 \pm 3.03	0.88 \pm 0.06	6.87 \pm 0.95	
SUPERVISED	TS-CAN [6]	UBFC-RPPG	3.69 \pm 1.74	13.8 \pm 113.84	3.39 \pm 1.44	0.82 \pm 0.08	5.26 \pm 1.11	
		SCAMPS	4.66 \pm 1.68	13.69 \pm 92.53	5.83 \pm 2.03	0.82 \pm 0.08	0.95 \pm 1.04	
	PHYSNET [5]	UBFC-RPPG	8.06 \pm 2.34	19.71 \pm 129.34	13.67 \pm 4.04	0.61 \pm 0.11	6.68 \pm 1.16	
		SCAMPS	13.30 \pm 2.00	20.30 \pm 94.85	20.01 \pm 2.97	0.51 \pm 0.11	-8.73 \pm 0.78	
	PHYSFORMER [34]	UBFC-RPPG	12.92 \pm 2.69	24.36 \pm 132.24	23.92 \pm 5.22	0.47 \pm 0.12	2.16 \pm 1.05	
		SCAMPS	26.58 \pm 2.14	31.24 \pm 133.33	42.79 \pm 4.06	0.12 \pm 0.13	-12.56 \pm 0.53	
	DEEPPHYS [35]	UBFC-RPPG	5.54 \pm 2.30	18.51 \pm 173.09	5.32 \pm 1.90	0.66 \pm 0.10	4.40 \pm 1.32	
		SCAMPS	3.96 \pm 1.67	13.44 \pm 98.86	4.25 \pm 1.60	0.83 \pm 0.07	5.07 \pm 1.15	
	EFF.PHYS-C [36]	UBFC-RPPG	5.47 \pm 2.10	17.04 \pm 143.80	5.40 \pm 1.76	0.71 \pm 0.09	4.09 \pm 1.16	
		SCAMPS	10.24 \pm 2.48	21.65 \pm 173.96	11.70 \pm 2.28	0.46 \pm 0.12	-5.49 \pm 1.05	
			Test Set UBFC-rPPG [22]					
	Method	Train Set	MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow	
GREEN [28]	N/A	19.73 \pm 3.75	31.00 \pm 235.38	18.72 \pm 3.33	0.37 \pm 0.15	-11.18 \pm 1.63		
ICA [29]	N/A	16.00 \pm 3.09	25.65 \pm 163.58	15.35 \pm 2.77	0.44 \pm 0.14	-9.91 \pm 1.78		
CHROM [30]	N/A	4.06 \pm 1.21	8.83 \pm 33.93	3.84 \pm 1.10	0.89 \pm 0.07	-2.96 \pm 1.18		
LGI [33]	N/A	15.80 \pm 3.67	28.55 \pm 236.17	14.70 \pm 3.20	0.36 \pm 0.15	-8.15 \pm 1.41		
SUPERVISED	PBV [32]	N/A	15.90 \pm 3.25	26.40 \pm 199.71	15.17 \pm 2.91	0.48 \pm 0.14	-9.16 \pm 1.35	
	POS [31]	N/A	4.08 \pm 1.01	7.72 \pm 21.87	3.93 \pm 0.91	0.92 \pm 0.06	-2.39 \pm 1.14	
	TS-CAN [6]	PURE	1.30 \pm 0.40	2.87 \pm 3.05	1.50 \pm 0.47	0.99 \pm 0.02	1.49 \pm 1.13	
		SCAMPS	3.62 \pm 0.91	6.92 \pm 18.30	3.53 \pm 0.84	0.93 \pm 0.06	-3.91 \pm 0.98	
	PHYSNET [5]	PURE	0.98 \pm 0.35	2.48 \pm 2.55	1.12 \pm 0.42	0.99 \pm 0.02	1.09 \pm 1.15	
		SCAMPS	5.40 \pm 1.46	10.89 \pm 48.53	5.43 \pm 1.38	0.82 \pm 0.09	-4.97 \pm 1.03	
	PHYSFORMER [34]	PURE	1.44 \pm 0.54	3.77 \pm 7.93	1.66 \pm 0.62	0.98 \pm 0.03	0.18 \pm 1.12	
		SCAMPS	4.56 \pm 1.46	10.48 \pm 68.96	5.18 \pm 1.93	0.81 \pm 0.09	-6.34 \pm 0.80	
	DEEPPHYS [35]	PURE	1.21 \pm 0.41	2.90 \pm 3.75	1.42 \pm 0.49	0.99 \pm 0.02	1.74 \pm 1.16	
		SCAMPS	3.10 \pm 1.44	9.81 \pm 74.70	3.08 \pm 1.32	0.87 \pm 0.08	-0.79 \pm 1.22	
	EFF.PHYS-C [36]	PURE	2.07 \pm 0.92	6.32 \pm 32.01	2.10 \pm 0.87	0.94 \pm 0.05	-0.12 \pm 1.20	
		SCAMPS	12.64 \pm 3.15	23.99 \pm 182.44	11.26 \pm 2.67	0.34 \pm 0.15	-9.36 \pm 1.05	

MAE = Mean Absolute Error in HR estimation (Beats/Min), RMSE = Root Mean Square Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%), ρ = Pearson Correlation in HR estimation, SNR = Signal-to-Noise Ratio (dB) when comparing predicted spectrum to ground truth spectrum.

G其他结果

我们重申在主要文件中提供的结果，并提出额外的结果，包括均方根误差，信噪比，皮尔逊相关性，以及相应的标准误差。请注意，下表中的结果与主论文中的结果之间可能存在微小差异，因为它们是使用最新版本的rPPG-1000在不同的机器上生成的。

表7：基准结果。使用rPPG工具箱获得的PURE [23]和UBFC-rPPG [22]数据集的性能。对于监督方法，我们使用UBFC-rPPG，PURE和SCAMPS数据集显示了交叉数据集训练结果。

		测试集 PURE [23]						
		方法训练集MAE ↓ RMSE ↓ MAPE ↓ ρ ↑ SNR ↑						
U非监督 S上覆	G[28]	不适用	10.09 ± 2.81	23.85 ± 217.81	10.28 ± 2.33	0.34 ± 0.12	-2.66 ± 1.43	伊卡[29]不适用
			16.07 ± 153.84	4.47 ± 1.65	0.72 ± 0.09	5.24 ± 1.77	CHROM [30]不适用	A 5.77 ± 1.79
PBV [32]	LGI [33]	不适用	4.61 ± 1.91	15.38 ± 134.14	4.96 ± 1.72	0.77 ± 0.08	4.50 ± 1.21	14.93 ± 81.53
			A 3.92 ± 1.61	12.99 ± 123.60	4.84 ± 1.49	0.84 ± 0.07	2.30 ± 1.31	11.52 ± 3.75
			1.46	11.82 ± 66.87	7.25 ± 3.03	0.88 ± 0.06	6.87 ± 0.95	POS [31]不适用
		TS-CAN[6]	UBFC-PPG	3.69 ± 1.74	13.8 ± 113.84	3.39 ± 1.44	0.82 ± 0.08	5.26 ± 1.11
		抽搐	4.66 ± 1.68	13.69 ± 92.53	5.83 ± 2.03	0.82 ± 0.08	0.95 ± 1.04	
		PN[5]	UBFC-PPG	8.06 ± 2.34	19.71 ± 129.34	13.67 ± 4.04	0.61 ± 0.11	6.68 ± 1.16
		抽搐	13.30 ± 2.00	20.30 ± 94.85	20.01 ± 2.97	0.51 ± 0.11	8.73 ± 0.78	
		PF[34]	UBFC-PPG	12.92 ± 2.69	24.36 ± 132.24	23.92 ± 5.22	0.47 ± 0.12	2.16 ± 1.05
		抽搐	26.58 ± 2.14	31.24 ± 133.33	42.79 ± 4.06	0.12 ± 0.13	12.56 ± 0.53	
		DP[35]	UBFC-PPG	5.54 ± 2.30	18.51 ± 173.09	5.32 ± 1.90	0.66 ± 0.10	4.40 ± 1.32
		抽搐	3.96 ± 1.67	13.44 ± 98.86	4.25 ± 1.60	0.83 ± 0.07	5.07 ± 1.15	
		E.P-C [36]	UBFC-PPG	5.47 ± 2.10	17.04 ± 143.80	5.40 ± 1.76	0.71 ± 0.09	4.09 ± 1.16
		SCAMPS	10.24 ± 2.48	21.65 ± 173.96	11.70 ± 2.28	0.46 ± 0.12	-5.49 ± 1.05	?
		测试集 UBFC-rPPG [22]						
		方法训练集MAE ↓ RMSE ↓ MAPE ↓ ρ ↑						
U非监督 S上覆	G[28]	不适用	19.73 ± 3.75	31.00 ± 235.38	18.72 ± 3.33	0.37 ± 0.15	-11.18 ± 1.63	伊卡[29]不适用
			3.09	25.65 ± 163.58	15.35 ± 2.77	0.44 ± 0.14	-9.91 ± 1.78	CHROM [30]不适用
			33.93	3.84 ± 1.10	0.89 ± 0.07	-2.96 ± 1.18		A 4.06 ± 1.21
		LGI [33]	N/A	15.80 ± 3.67	28.55 ± 236.17	14.70 ± 3.20	0.36 ± 0.15	-8.15 ± 1.41
		抽搐	15.90 ± 3.25	26.40 ± 199.71	15.17 ± 2.91	0.48 ± 0.48	0.14 ± 1.95	1.31 [35] N/A A 4.08 ± 1.01
			7.72 ± 21.87	3.93 ± 0.91	0.92 ± 0.06	-2.39 ± 1.14		
		TS-CAN [1]	PURE	1.30 ± 0.40	2.87 ± 3.05	1.50 ± 0.47	0.99 ± 0.02	1.49 ± 1.13
			SCAMPS	3.62 ± 0.91	6.92 ± 18.30	3.53 ± 0.84	± 0.93 ± 0.06	-3.91 ± 0.98
		PN [4]	PURE	0.98 ± 0.35	2.48 ± 2.55	1.12 ± 0.42	0.99 ± 0.02	1.09 ± 1.15
			SCAMPS	5.40 ± 1.46	10.89 ± 48.53	5.43 ± 1.38	0.82 ± 0.09	-4.97 ± 1.03
		PF [34]	PURE	1.44 ± 0.54	3.77 ± 7.93	1.66 ± 0.62	0.98 ± 0.03	0.18 ± 1.12
			SCAMPS	4.56 ± 1.46	10.48 ± 68.96	5.18 ± 1.93	0.81 ± 0.09	-6.34 ± 0.80
		DP [34]	PURE	1.21 ± 0.41	2.90 ± 3.75	1.42 ± 0.49	0.99 ± 0.02	1.74 ± 1.16
			SCAMPS	3.10 ± 1.44	9.81 ± 74.70	3.08 ± 1.32	0.87 ± 0.08	-0.79 ± 1.22
		E. P-C [36]	PURE	2.07 ± 0.92	6.32 ± 32.01	2.10 ± 0.87	0.94 ± 0.05	-0.12 ± 1.20
			SCAMPS	12.64 ± 3.15	23.99 ± 182.44	11.26 ± 2.67	0.34 ± 0.15	-9.36 ± 1.05

MAE = HR估计中的平均绝对误差（心跳/分钟），RMSE = HR估计中的均方根误差（心跳/分钟），MAPE = 平均百分比误差（%），ρ = HR估计中的皮尔逊相关性，SNR = 将预测频谱与真实频谱进行比较时的信噪比（dB）。

Table 8: **Benchmark Results.** Performance on the UBFC-Phys [27] and MMPD [25] datasets generated using the rPPG toolbox. For the supervised methods we show cross-dataset training results using the UBFC-rPPG, PURE and SCAMPS datasets.

		Test Set					
		UBFC-Phys [27]					
	Method	Train Set	MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow
UNSUPERVISED	GREEN [28]	N/A	13.55 \pm 1.30	18.80 \pm 48.87	16.01 \pm 1.42	0.29 \pm 0.10	-10.34 \pm 0.65
	ICA [29]	N/A	10.04 \pm 1.20	15.73 \pm 43.63	11.85 \pm 1.35	0.36 \pm 0.09	-5.28 \pm 0.98
	CHROM [30]	N/A	4.49 \pm 0.60	7.56 \pm 13.84	6.00 \pm 0.88	0.80 \pm 0.06	-1.92 \pm 0.85
	LGI [33]	N/A	6.27 \pm 0.83	10.41 \pm 22.76	7.83 \pm 0.99	0.70 \pm 0.07	-3.30 \pm 0.91
	PBV [32]	N/A	12.34 \pm 1.22	17.43 \pm 47.24	14.63 \pm 1.33	0.33 \pm 0.09	-9.33 \pm 0.71
	POS [31]	N/A	4.51 \pm 0.68	8.16 \pm 17.36	6.12 \pm 0.99	0.77 \pm 0.06	-1.28 \pm 0.90
SUPERVISED			Test Set				
			UBFC-Phys [27]				
	TS-CAN [6]	UBFC-RPPG	5.13 \pm 0.63	8.12 \pm 18.47	6.53 \pm 0.85	0.76 \pm 0.07	-1.95 \pm 0.81
	PURE		5.72 \pm 0.66	8.78 \pm 16.94	7.34 \pm 0.90	0.72 \pm 0.07	-3.72 \pm 0.78
	SCAMPS		5.55 \pm 0.67	8.71 \pm 16.96	6.91 \pm 0.85	0.72 \pm 0.07	-4.40 \pm 0.66
	PHYSNET [5]		5.79 \pm 0.76	9.60 \pm 17.64	7.69 \pm 1.07	0.70 \pm 0.07	-1.63 \pm 0.99
	PURE		4.78 \pm 0.72	8.68 \pm 18.99	6.15 \pm 0.98	0.73 \pm 0.07	-0.71 \pm 1.00
	SCAMPS		8.53 \pm 0.98	13.02 \pm 33.08	11.22 \pm 1.35	0.43 \pm 0.10	-7.15 \pm 0.60
	PHYSFORMER [34]		6.63 \pm 0.77	10.22 \pm 18.12	8.91 \pm 1.12	0.69 \pm 0.07	-3.58 \pm 0.93
	PURE		6.04 \pm 0.76	9.77 \pm 18.38	7.67 \pm 0.99	0.65 \pm 0.08	-2.16 \pm 0.95
	SCAMPS		11.91 \pm 1.13	16.42 \pm 46.43	15.57 \pm 1.64	0.27 \pm 0.097	-10.38 \pm 0.39
SUPERVISED			Test Set				
			MMPD [25]				
	DEEPPHYS [35]	UBFC-RPPG	6.62 \pm 0.84	10.69 \pm 25.90	8.21 \pm 1.04	0.66 \pm 0.08	-2.98 \pm 0.82
	PURE		8.42 \pm 1.09	13.80 \pm 38.06	10.18 \pm 1.29	0.44 \pm 0.09	-4.41 \pm 0.84
	SCAMPS		4.75 \pm 0.58	7.50 \pm 14.47	5.89 \pm 0.72	0.82 \pm 0.06	-2.04 \pm 0.76
	EFF.PHYS-C [36]		4.93 \pm 0.58	7.65 \pm 14.44	6.25 \pm 0.79	0.79 \pm 0.06	-2.09 \pm 0.82
	PURE		5.31 \pm 0.78	9.44 \pm 27.67	6.61 \pm 0.96	0.70 \pm 0.07	-2.22 \pm 0.81
	SCAMPS		6.97 \pm 0.79	10.58 \pm 22.70	8.47 \pm 0.91	0.64 \pm 0.08	-7.38 \pm 0.47
			Test Set				
UNSUPERVISED			MMPD [25]				
	GREEN [28]	N/A	21.68 \pm 0.67	27.69 \pm 42.21	24.39 \pm 0.64	-0.01 \pm 0.04	-14.34 \pm 0.26
	ICA [29]	N/A	18.60 \pm 0.61	24.30 \pm 33.80	20.88 \pm 0.58	0.01 \pm 0.04	-13.84 \pm 0.27
	CHROM [30]	N/A	13.66 \pm 0.50	18.76 \pm 23.82	16.00 \pm 0.57	0.08 \pm 0.04	-11.74 \pm 0.21
	LGI [33]	N/A	17.08 \pm 0.62	23.32 \pm 34.46	18.98 \pm 0.60	0.04 \pm 0.04	-13.15 \pm 0.25
	PBV [32]	N/A	17.95 \pm 0.60	23.58 \pm 32.45	20.18 \pm 0.58	0.09 \pm 0.04	-13.88 \pm 0.24
SUPERVISED	POS [31]	N/A	12.36 \pm 0.49	17.71 \pm 23.65	14.43 \pm 0.55	0.18 \pm 0.04	-11.53 \pm 0.22
	TS-CAN [6]	UBFC-RPPG	14.01 \pm 0.61	21.04 \pm 30.02	15.48 \pm 0.61	0.24 \pm 0.04	-10.18 \pm 0.28
	PURE		13.94 \pm 0.64	21.61 \pm 33.02	15.15 \pm 0.63	0.20 \pm 0.04	-9.94 \pm 0.27
	SCAMPS		19.05 \pm 0.58	24.20 \pm 31.90	21.77 \pm 0.60	0.14 \pm 0.04	-13.24 \pm 0.25
	PHYSNET [5]	UBFC-RPPG	9.47 \pm 0.50	16.01 \pm 22.74	11.11 \pm 0.58	0.31 \pm 0.04	-8.15 \pm 0.26
	PURE		13.93 \pm 0.57	20.29 \pm 27.57	15.61 \pm 0.59	0.17 \pm 0.04	-10.59 \pm 0.27
SUPERVISED	SCAMPS		20.78 \pm 0.55	25.09 \pm 31.92	24.43 \pm 0.62	0.17 \pm 0.04	-15.86 \pm 0.20
	PHYSFORMER [34]	UBFC-RPPG	12.1 \pm 0.51	17.79 \pm 23.77	15.41 \pm 0.74	0.17 \pm 0.04	-10.53 \pm 0.20
	PURE		14.57 \pm 0.57	20.71 \pm 29.1	16.73 \pm 0.63	0.15 \pm 0.039	-12.15 \pm 0.22
	SCAMPS		22.69 \pm 0.57	26.94 \pm 34.41	27.06 \pm 0.70	0.15 \pm 0.04	-18.83 \pm 0.32
	DEEPPHYS [35]	UBFC-RPPG	17.50 \pm 0.70	25.00 \pm 38.62	19.27 \pm 0.68	0.06 \pm 0.04	-11.72 \pm 0.33
	PURE		16.92 \pm 0.70	24.61 \pm 38.03	18.54 \pm 0.68	0.05 \pm 0.04	-11.53 \pm 0.31
SUPERVISED	SCAMPS		15.22 \pm 0.68	23.17 \pm 38.46	16.56 \pm 0.66	0.09 \pm 0.04	-10.23 \pm 0.31
	EFF.PHYS-C [36]	UBFC-RPPG	13.78 \pm 0.68	22.25 \pm 37.94	15.15 \pm 0.70	0.09 \pm 0.04	-9.13 \pm 0.31
	PURE		14.03 \pm 0.64	21.62 \pm 32.95	15.32 \pm 0.63	0.17 \pm 0.04	-9.95 \pm 0.29
	SCAMPS		20.41 \pm 0.57	25.06 \pm 31.72	23.52 \pm 0.61	0.11 \pm 0.04	-14.28 \pm 0.24

MAE = Mean Absolute Error in HR estimation (Beats/Min), RMSE = Root Mean Square Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%), ρ = Pearson Correlation in HR estimation, SNR = Signal-to-Noise Ratio (dB) when comparing predicted spectrum to ground truth spectrum.

表8：基准测试结果。性能在UBFC—Phys [27]和MMPD [25]数据集生成，使用rPPG工具箱。对于监督方法，我们使用UBFC—rPPG、PURE和SCAMPS数据集显示跨数据集培训结果。

		测试集 UBFC-Phys [27]									
		方法训练集MAE ↓ RMSE ↓ MAPE ↓ ρ ↑ SNR ↑									
U非监督		G[28]不适用 13.55 ± 1.30 18.80 ± 48.87 16.01 ± 1.42 0.29 ± 0.10 -10.34 ± 0.65 伊卡[29]不适用 10.04 ± 1.20 15.73 ± 43.63 11.85 ± 1.35 0.36 ± 0.09 -5.28 ± 0.98 CHROM [30]不适用 A 4.49 ± 0.60 7.56 ± 13.84 6.00 ± 0.88 0.80 ± 0.06 -1.92 ± 0.85 LGI [33] N/A 6.27 ± 0.83 10.41 ± 22.76 7.83 ± 0.99 0.70 ± 0.07 -3.30 ± 0.91 PBV [32] N/A 12.34 ± 1.22 17.43 ± 47.24 14.63 ± 1.33 0.33 ± 0.09 -9.33 ± 0.71 POS [31]不适用 4.51 ± 0.68 8.16 ± 17.36 6.12 ± 0.99 0.77 ± 0.06 -1.28 ± 0.90									
U上覆		UBFC-PPG 5.13 ± 0.63 8.12 ± 18.47 6.53 ± 0.85 0.76 ± 0.07 -1.95 ± 0.81 PURE 5.72 ± 0.66 8.78 ± 16.94 7.34 ± 0.90 0.72 ± 0.07 -3.72 ± 0.78 SCAMPS 5.55 ± 0.67 8.71 ± 16.96 6.91 ± 0.85 0.72 ± 0.07 -4.40 ± 0.66  									
S上覆		TS-CAN [6] UBFC-PPG 5.79 ± 0.76 9.60 ± 17.64 7.69 ± 1.07 0.70 ± 0.07 -1.63 ± 0.99 PURE 4.78 ± 0.72 8.68 ± 18.99 6.15 ± 0.98 0.73 ± 0.07 -0.71 ± 1.00 SCAMPS 8.53 ± 0.98 13.02 ± 33.08 11.22 ± 1.35 0.43 ± 0.10 -7.15 ± 0.60  									
S上覆		PN[5] PF[34] UBFC-PPG 6.63 ± 0.77 10.22 ± 18.12 8.91 ± 1.12 0.69 ± 0.07 -3.58 ± 0.93   PURE 6.04 ± 0.76 9.77 ± 18.38 7.67 ± 0.99 0.65 ± 0.08 -2.16 ± 0.95 SCAMPS 11.91 ± 1.13 16.42 ± 46.43 15.57 ± 1.64 0.27 ± 0.097 -10.38 ± 0.39  									
S上覆		DP[35] UBFC-PPG 6.62 ± 0.84 10.69 ± 25.90 8.21 ± 1.04 0.66 ± 0.08 -2.98 ± 0.82 PURE 8.42 ± 1.09 13.80 ± 38.06 10.18 ± 1.29 0.44 ± 0.09 -4.41 ± 0.84 SCAMPS 4.75 ± 0.58 7.50 ± 14.47 5.89 ± 0.72 0.82 ± 0.06 -2.04 ± 0.76  									
E.P-C		E.P-C [36] UBFC-PPG 4.93 ± 0.58 7.65 ± 14.44 6.25 ± 0.79 0.79 ± 0.06 -2.09 ± 0.82 PURE 5.31 ± 0.78 9.44 ± 27.67 6.61 ± 0.96 0.70 ± 0.07 -2.22 ± 0.81 SCAMPS 6.97 ± 0.79 10.58 ± 22.70 8.47 ± 0.91 0.64 ± 0.08 -7.38 ± 0.47  									
		测试集 MMPD [25]									
		方法训练集MAE ↓ RMSE ↓ MAPE ↓ ρ ↑ SNR ↑									
U非监督		G[28] N/A 21.68 ± 0.67 27.69 ± 42.21 24.39 ± 0.64 -0.01 ± 0.04 -14.34 ± 0.26 伊卡[29] N/A 18.60 ± 0.61 24.30 ± 33.80 20.88 ± 0.58 0.01 ± 0.04 -13.84 ± 0.27 CHROM [30]不适用 13.66 ± 0.50 18.76 ± 23.82 16.00 ± 0.57 0.08 ± 0.04 -11.74 ± 0.21 LGI [33]不适用 17.08 ± 0.62 23.32 ± 34.46 18.98 ± 0.60 0.04 ± 0.04 -13.15 ± 0.25 PBV [32]不适用 A 17.95 ± 0.60 23.58 ± 32.45 20.18 ± 0.58 0.09 ± 0.04 -13.88 ± 0.24 POS [31]不适用 12.36 ± 0.49 17.71 ± 23.65 14.43 ± 0.55 0.18 ± 0.04 -11.53 ± 0.22									
能 (S-CAN)		UBFC-PPG 14.01 ± 0.61 21.04 ± 30.02 15.48 ± 0.61 0.24 ± 0.04 -10.18 ± 0.28 PURE 13.94 ± 0.64 21.61 ± 33.02 15.15 ± 0.63 0.20 ± 0.04 9.94 ± 0.27 SCAMPS 19.05 ± 0.58 24.20 ± 31.90 21.77 ± 0.60 0.14 ± 0.04 -13.24 ± 0.25									
能 (S-CAN)		PN[5] UBFC-PPG 9.47 ± 0.50 16.01 ± 22.74 11.11 ± 0.58 0.31 ± 0.04 -8.15 ± 0.26 PURE 13.93 ± 0.57 20.29 ± 27.57 15.61 ± 0.59 0.17 ± 0.04 -10.59 ± 0.27 SCAMPS 20.78 ± 0.55 25.09 ± 31.92 24.43 ± 0.62 0.17 ± 0.04 -15.86 ± 0.20  									
能 (S-CAN)		PF[34] UBFC-PPG 12.1 ± 0.51 17.79 ± 23.77 15.41 ± 0.74 0.17 ± 0.04 -10.53 ± 0.20   PURE 14.57 ± 0.57 20.71 ± 29.71 16.73 ± 0.63 0.15 ± 0.039 -12.15 ± 0.22 SCAMPS 22.69 ± 0.57 26.94 ± 34.41 27.06 ± 0.70 0.15 ± 0.04 -18.83 ± 0.32  									
DP[35]		UBFC-PPG 17.50 ± 0.70 25.00 ± 38.62 19.27 ± 0.68 0.06 ± 0.04 -11.72 ± 0.33 PURE 16.92 ± 0.70 24.61 ± 38.03 18.54 ± 0.68 0.05 ± 0.04 -11.53 ± 0.31 SCAMPS 15.22 ± 0.68 23.17 ± 38.46 16.56 ± 0.66 0.09 ± 0.04 -10.23 ± 0.31  									
E.P-C		E.P-C [36] UBFC-PPG 13.78 ± 0.68 22.25 ± 37.94 15.15 ± 0.70 0.09 ± 0.04 -9.13 ± 0.31 PURE 14.03 ± 0.64 21.62 ± 32.95 15.32 ± 0.63 0.17 ± 0.04 -9.95 ± 0.29 SCAMPS 20.41 ± 0.57 25.06 ± 31.72 23.52 ± 0.61 0.11 ± 0.04 -14.28 ± 0.24									

MAE = HR estimation中的平均绝对误差 (Beats/Min) , RMSE = HR estimation中的根平均平方误差 (Beats/Min) , MAPE = 平均百分比误差 (%) , ρ = HR estimation中的皮尔逊相关性, SNR = 信号到噪声比 (dB) 当与真实频谱进行比较时。

Table 9: **Training with Motion-Augmented Data.** We demonstrate results training on a motion-augmented (MA) version of the UBFC-rPPG [22] dataset generated using an open-source motion augmentation pipeline [43] and testing on the unaugmented versions of the PURE [23] dataset, UBFC-Phys [27], and MMPD [25] datasets.

Metric (\pm Std. Err.)	Training Set Testing Set	MAUBFC-rPPG [22] PURE [23]				
		MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow
Supervised						
TS-CAN [6]	1.07 \pm 0.75	5.89 \pm 33.75	1.20 \pm 0.83	0.97 \pm 0.03	8.86 \pm 0.95	
PhysNet (Normalized) [5]	17.03 \pm 2.97	28.50 \pm 149.16	32.37 \pm 5.82	0.38 \pm 0.12	7.27 \pm 0.88	
DeepPhys [35]	1.15 \pm 0.76	5.95 \pm 33.75	1.40 \pm 0.85	0.97 \pm 0.03	9.94 \pm 1.00	
EfficientPhys-C [36]	2.59 \pm 1.43	11.29 \pm 96.01	2.67 \pm 1.27	0.88 \pm 0.06	6.75 \pm 1.12	
Metric (\pm Std. Err.)	Training Set Testing Set	MAUBFC-rPPG [22] UBFC-Phys [27]				
		MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow
Supervised						
TS-CAN [6]	5.03 \pm 0.67	8.39 \pm 18.26	6.36 \pm 0.90	0.75 \pm 0.07	-1.15 \pm 0.81	
PhysNet (Normalized) [5]	5.51 \pm 0.88	0.44 \pm 37.65	7.50 \pm 1.32	0.68 \pm 0.07	-0.57 \pm 1.08	
DeepPhys [35]	4.95 \pm 0.67	8.37 \pm 21.53	6.26 \pm 0.90	0.75 \pm 0.07	-0.78 \pm 0.85	
EfficientPhys-C [36]	4.80 \pm 0.58	7.52 \pm 15.02	6.10 \pm 0.79	0.79 \pm 0.06	-0.87 \pm 0.86	
Metric (\pm Std. Err.)	Training Set Testing Set	MAUBFC-rPPG [22] MMPD [25]				
		MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow
Supervised						
TS-CAN [6]	12.59 \pm 0.62	20.23 \pm 31.27	13.77 \pm 0.62	0.23 \pm 0.04	-9.19 \pm 0.29	
PhysNet (Normalized) [5]	10.68 \pm 0.49	16.56 \pm 19.72	14.01 \pm 0.72	0.32 \pm 0.04	-9.28 \pm 0.21	
DeepPhys [35]	12.71 \pm 0.65	21.04 \pm 35.40	13.70 \pm 0.64	0.21 \pm 0.04	-8.85 \pm 0.31	
EfficientPhys-C [36]	13.42 \pm 0.66	21.64 \pm 35.46	14.52 \pm 0.65	0.14 \pm 0.04	-9.20 \pm 0.31	

MAE = Mean Absolute Error in HR estimation (Beats/Min), RMSE = Root Mean Square Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%), ρ = Pearson Correlation in HR estimation, SNR = Signal-to-Noise Ratio (dB) when comparing predicted spectrum to ground truth spectrum.

Table 10: **Training with Pseudo Labels.** For the supervised methods we show results training with the (entire) BP4D+ [26] dataset, using POS [31] derived pseudo training labels.

Metric (\pm Std. Err.)	Training Set Testing Set	BP4D+[26] with POS Pseudo Labels UBFC-rPPG [22]				
		MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow
Supervised						
TS-CAN [6]	4.69 \pm 1.88	13.04 \pm 100.15	4.51 \pm 1.65	0.78 \pm 0.10	0.01 \pm 1.27	
PhysNet(Normalized) [5]	1.78 \pm 0.67	4.68 \pm 11.94	1.92 \pm 0.72	0.96 \pm 0.04	1.24 \pm 1.08	
DeepPhys [35]	2.74 \pm 0.96	6.78 \pm 27.43	2.81 \pm 0.91	0.93 \pm 0.06	-0.22 \pm 1.33	
EfficientPhys-C [36]	2.43 \pm 1.29	8.68 \pm 67.51	2.52 \pm 1.20	0.90 \pm 0.07	0.39 \pm 1.27	
Metric (\pm Std. Err.)	Training Set Testing Set	BP4D+[26] with POS Pseudo Labels PURE [23]				
		MAE \downarrow	RMSE \downarrow	MAPE \downarrow	$\rho \uparrow$	SNR \uparrow
Supervised						
TS-CAN [6]	1.29 \pm 0.76	6.00 \pm 33.74	1.60 \pm 0.86	0.97 \pm 0.03	8.61 \pm 1.02	
PhysNet(Normalized) [5]	3.69 \pm 1.46	11.79 \pm 64.42	7.35 \pm 3.01	0.88 \pm 0.06	8.33 \pm 0.06	
DeepPhys [35]	2.47 \pm 1.41	11.11 \pm 93.02	2.49 \pm 1.21	0.89 \pm 0.061	7.32 \pm 1.09	
EfficientPhys-C [36]	3.59 \pm 1.84	14.55 \pm 135.51	3.27 \pm 1.50	0.80 \pm 0.08	7.48 \pm 1.15	

MAE = Mean Absolute Error in HR estimation (Beats/Min), RMSE = Root Mean Square Error in HR estimation (Beats/Min), MAPE = Mean Percentage Error (%), ρ = Pearson Correlation in HR estimation, SNR = Signal-to-Noise Ratio (dB) when comparing predicted spectrum to ground truth spectrum.

表9：使用运动增强数据进行训练。我们展示了使用开源运动增强管道[43]生成的UBFC-rPPG [22]数据集的运动增强 (MA) 版本的训练结果[43]，以及PURE [23]数据集，UBFC-Phys [27]和MMPD [25]数据集的未增强版本的测试结果。

训练集MAUBFC-rPPG [22]测试集PURE [23]度量 (±标准)						
	(Err.) MAE ↓	RMSE ↓	MAPE ↓	$\rho \uparrow$	SNR ↑	
监督						
TS-CAN [6]	1.07 ± 0.75	5.89 ± 33.75	1.20 ± 0.83	0.97 ± 0.03	8.86 ± 0.95	PhysNet
(归一化) [5]	17.03 ± 2.97	28.50 ± 149.16	32.37 ± 5.82	0.38 ± 0.12	7.27 ± 0.88	
DeepPhys [35]	1.15 ± 0.76	5.95 ± 33.75	1.40 ± 0.85	0.97 ± 0.03	9.94 ± 1.00	
EfficientPhys-C [36]	2.59 ± 1.43	11.29 ± 96.01	2.67 ± 1.27	0.88 ± 0.06	6.75 ± 1.12	

训练集MAUBFC-rPPG [22]测试集UBFC-Phys [27]度量 (±标准)						
	(Err.) MAE ↓ RMSE ↓	MAPE ↓	$\rho \uparrow$	SNR ↑		
监督						
TS-CAN [6]	5.03 ± 0.67	8.39 ± 18.26	6.36 ± 0.90	0.75 ± 0.07	-1.15 ± 0.81	
PhysNet (归一化) [5]	5.51 ± 0.88	0.44 ± 37.65	7.50 ± 1.32	0.68 ± 0.07	-0.57 ± 1.08	
DeepPhys [35]	4.95 ± 0.67	8.37 ± 21.53	6.26 ± 0.90	0.75 ± 0.07	-0.78 ± 0.85	
EfficientPhys-C [36]	4.80 ± 0.58	7.52 ± 15.02	6.10 ± 0.79	0.79 ± 0.06	-0.87 ± 0.86	

训练集MAUBFC-rPPG [22]测试集MMPD [25]度量 (±标准)						
	(Err.) MAE ↓	RMSE ↓	MAPE ↓	$\rho \uparrow$	SNR ↑	
监督						
TS-CAN [6]	12.59 ± 0.62	20.23 ± 31.27	13.77 ± 0.62	0.23 ± 0.04	-9.19 ± 0.29	
PhysNet (归一化) [5]	10.68 ± 0.49	16.56 ± 19.72	14.01 ± 0.72	0.32 ± 0.04	-9.28 ± 0.29	
DeepPhys [35]	12.71 ± 0.65	21.04 ± 35.40	13.70 ± 0.64	0.21 ± 0.04	-8.85 ± 0.31	EfficientPhys-C [36]
13.42 ± 0.66	21.64 ± 35.46	14.52 ± 0.65	0.14 ± 0.04	-9.20 ± 0.31	MAE = HR估计中的平均绝对误差 (心跳/分钟)， RMSE = HR估计中的均方根误差 (心跳/分钟)， MAPE = 平均百分比误差 (%)， ρ = HR估计中的皮尔逊相关性， SNR = 比较预测频谱与地面真实频谱时的信噪比 (dB)。	

表10：使用伪标签的训练。对于监督方法，我们使用（整个）BP4D+ [26]数据集，使用POS [31]衍生的伪训练标签显示训练结果。

具有POS伪标签的训练集BP 4D+[26]测试集UBFC-rPPG [22]度量 (±标准)						
	(±标准) (Err.) MAE ↓	RMSE ↓	MAPE ↓	$\rho \uparrow$	SNR ↑	
监督						
TS-CAN [6]	4.69 ± 1.88	13.04 ± 100.15	4.51 ± 1.65	0.78 ± 0.10	0.01 ± 1.27	PhysNet
(归一化) [5]	1.78 ± 0.67	4.68 ± 11.94	1.92 ± 0.72	0.96 ± 0.04	1.24 ± 1.08	
DeepPhys [35]	2.74 ± 0.96	6.78 ± 27.43	2.81 ± 0.91	0.93 ± 0.06	-0.22 ± 1.33	
EfficientPhys-C [36]	2.43 ± 1.29	8.68 ± 67.51	2.52 ± 1.20	0.90 ± 0.07	0.39 ± 1.27	

训练集BP4D+[26]与POS伪标签测试集PURE [23]度量 (±标准) (Err.) MAE ↓						
	RMSE ↓	MAPE ↓	$\rho \uparrow$	SNR ↑		
监督						
TS-CAN [6]	1.29 ± 0.76	6.00 ± 33.74	1.60 ± 0.86	0.97 ± 0.03	8.61 ± 1.02	PhysNet
(归一化) [5]	3.69 ± 1.46	11.79 ± 64.42	7.35 ± 3.01	0.88 ± 0.06	8.33 ± 0.06	
DeepPhys [35]	2.47 ± 1.41	11.11 ± 93.02	2.49 ± 1.21	0.89 ± 0.061	7.32 ± 1.09	EfficientPhys-C [36]
3.59 ± 1.84	14.55 ± 135.51	3.27 ± 1.50	0.80 ± 0.08	7.48 ± 1.15	MAE = HR估计中的平均绝对误差 (心跳/分钟)， RMSE = HR估计中的均方根误差 (心跳/分钟)， MAPE = 平均百分比误差 (%)， ρ = HR估计中的皮尔逊相关性， SNR = 比较预测频谱与地面真实频谱时的信噪比 (dB)。	

Table 11: **Full 3-Fold Multitasking Results.** For the BigSmall [41] method we show the full 3-fold results for multi-tasking PPG, respiration, and action unit classification; training and evaluating on the BP4D+ [26] (AU subset) dataset, using POS [31] derived pseudo training PPG labels.

Training Set Testing Set Fold	BP4D+[26]				
	BP4D+[26]		BP4D+[26]		Fold 3
	Fold 1	Fold 2	Fold 3		
Heart (Metric \pm Std. Err. \downarrow)	MAE \downarrow	4.24 \pm 0.73	2.91 \pm 0.49	2.54 \pm 0.48	
	RMSE \downarrow	10.76 \pm 33.20	7.26 \pm 13.90	7.06 \pm 16.67	
	MAPE \downarrow	4.55 \pm 0.74	3.22 \pm 0.53	2.75 \pm 0.49	
	$\rho\uparrow$	0.68 \pm 0.05	0.90 \pm 0.03	0.91 \pm 0.03	
	SNR \uparrow	3.85 \pm 0.69	6.27 \pm 0.67	6.53 \pm 0.63	
Respiration (Metric \pm Std. Err. \downarrow)	MAE \downarrow	5.28 \pm 0.31	4.96 \pm 0.33	5.34 \pm 0.35	
	RMSE \downarrow	6.74 \pm 4.38	6.67 \pm 4.96	7.18 \pm 5.12	
	MAPE \downarrow	24.41 \pm 1.55	25.30 \pm 2.08	29.14 \pm 2.72	
	$\rho\uparrow$	0.15 \pm 0.07	0.16 \pm 0.72	0.12 \pm 0.07	
	SNR \uparrow	7.69 \pm 0.64	10.53 \pm 0.75	9.34 \pm 0.64	
Facial Action (AU) (F1\uparrow, Prec.\uparrow)	AU01	18.62	11.04	18.88	11.34
	AU02	20.76	12.73	18.28	10.89
	AU04	12.57	8.08	11.48	7.85
	AU06	66.73	66.58	64.71	61.09
	AU07	74.86	78.68	70.08	75.10
	AU10	74.92	77.32	70.09	74.48
	AU12	72.69	70.79	67.75	68.54
	AU14	67.21	72.84	70.18	69.11
	AU15	22.56	13.91	22.33	13.38
	AU17	25.77	18.01	20.95	12.45
Facial Action (AU) (Metric Mean)	F1 \uparrow	41.53		39.97	46.96
	Prec. \uparrow	36.42		36.22	44.89
	Acc. (%) \uparrow	61.91		62.42	72.83

For HR estimation, MAE = Mean Absolute Error, RMSE = Root Mean Square Error, MAPE = Mean Percentage Error (%), ρ = Pearson Correlation, SNR = Signal-to-Noise Ratio (dB) when comparing predicted spectrum to ground truth spectrum. For AU classification F1 = harmonic mean of precision and recall, Prec. = precision, Acc. = accuracy.

表11：完整的3重多任务处理结果。对于BigSmall [41]方法，我们显示了多任务PPG、呼吸和动作单元分类的完整3倍结果;使用POS [31]衍生的伪训练PPG标签，在BP 4D + [26] (Au子集) 数据集上进行训练和评估。

	训练组BP 4D +[26]			测试设置BP 4D +[26]			
	Fold 1的			Fold 2		Fold 3	
心脏MAE ↓	4.24 ± 0.73	2.91 ± 0.49	2.54 ± 0.48	(Metric ± Std.)	()	(1)	(7) (0)
(16)							
MAPE ↓	4.55 ± 0.74	3.22 ± 0.53	2.75 ± 0.49				
ρ ↑	0.68 ± 0.05	0.90 ± 0.03	0.91 ± 0.03	SNR ↑			
	3.85 ± 0.69	6.27 ± 0.67	6.53 ± 0.63				
Respiration MAE ↓	5.28 ± 0.31	4.96 ± 0.33	5.34 ± 0.35	(Metric ± Std.)	(2)	RMSE	
6.74 ± 4.38	6.67 ± 4.96	7.18 ± 5.12					
MAPE ↓	24.41 ± 1.55	25.30 ± 2.08	29.14 ± 2.72				
ρ ↑	0.15 ± 0.07	0.16 ± 0.72	0.12 ± 0.07	SNR ↑	7.69		
	± 0.64	10.53 ± 0.75	9.34 ± 0.64				
面部动作 (Au)	AU 01	18.62	11.04	18.88	11.34	24.32	16.43
(F1^ (前) ^	AU02	20.76	12.73	18.28	10.89	15.46	9.07
	11.48	7.85	14.43	8.63	AU06	66.73	66.58
	76.44	79.20	74.86	78.68	788	74.86	788
	86.34	AU10	74.92	77.32	70.09	74.48	82.09
	72.69	67.75	68.54	80.96	88.02	AU14	672.84
	69.11	66.73	70.93	AU15	2525.26	88.02	88.02
	13.38	29.64	22.13	AU17	25.77	18.01	20.95
	28.06	AU23	34.64	27.41	34.21	24.19	40.68
	7.00	3.71	10.70	6.20	19.03	28.76	AU24
							10.81
面部动作 (Au)	F1↑	41.53			39.97		46.96
(公制平均值)	前数↑	36.42	36.22	44.89			
	应计 (%) ↑	61.91			62.42		72.83

对于HR估计，MAE =平均绝对误差，RMSE =均方根误差，MAPE =平均百分比误差(%)， ρ =皮尔逊相关性，SNR =将预测频谱与真实频谱进行比较时的信噪比(dB)。对于Au分类，F1 =精确度和召回率的调和平均值，Prec. =精确度，Acc. =准确度。

H UBFC-Phys Video Exclusion

For evaluation of the UBFC-Phys [27] dataset in our main paper and by default in our toolbox, we utilized all three tasks and the same subject exclusion, or conversely sub-selection, list provided by the authors of the dataset in the second supplementary material of their paper [27]. Based on the aforementioned supplementary material, we eliminated 14 subjects (s3, s8, s9, s26, s28, s30, s31, s32, s33, s40, s52, s53, s54, s56) for the rest task (T1), 30 subjects (s1, s4, s6, s8, s9, s11, s12, s13, s14, s19, s21, s22, s25, s26, s27, s28, s31, s32, s33, s35, s38, s39, s41, s42, s45, s47, s48, s52, s53, s55) for the speech task (T2), and 23 subjects (s5, s8, s9, s10, s13, s14, s17, s22, s25, s26, s28, s30, s32, s33, s35, s37, s40, s47, s48, s49, s50, s52, s53) for the arithmetic task (T3).

In our toolbox, video exclusion is achieved using dataset filtering criteria specified in the config file. Specifically, an exclusion list or a task selection list can be provided to respectively exclude videos from being included or to select specific tasks as a part of a dataset.

I Multitasking Training and Evaluation Details

To show how this toolbox may be extended for physiological multitasking, we implement BigSmall [41] a model that multitasks PPG, respiration, and facial action. Here we reiterate information from [41], with slight modifications, for clarification.

I.1 Cross Validation Subject Folds

Fold 1: F003, F004, F005, F006, F009, F017, F022, F028, F029, F031, F032, F033, F038, F044, F047, F048, F052, F053, F055, F061, F063, F067, F068, F074, F075, F076, F081, M003, M005, M006, M009, M012, M019, M025, M026, M028, M031, M036, M037, M040, M046, M047, M049, M051, M054, M056.

Fold 2: F001, F002, F008, F018, F021, F025, F026, F035, F036, F037, F039, F040, F041, F042, F046, F049, F057, F058, F060, F062, F064, F066, F070, F071, F072, F073, F077, F082, M001, M002, M007, M013, M014, M016, M022, M023, M024, M027, M029, M030, M034, M035, M041, M042, M043, M048, M055.

Fold 3: F007, F010, F011, F012, F013, F014, F015, F016, F019, F020, F023, F024, F027, F030, F034, F043, F045, F050, F051, F054, F056, F059, F065, F069, F078, F079, F080, M004, M008, M010, M011, M015, M017, M018, M020, M021, M032, M033, M038, M039, M044, M045, M050, M052, M053, M057, M058.

I.2 AU Subset

The AU subset used for model training and evaluation (in this toolbox) is made up of dataset subset which contains action unit labels. This consists of approximately 20 seconds worth of data from the following tasks for each subject: T1, T6, T7, T8.

I.3 Subject Fold Splits

[41] is evaluated using 3 fold cross validation, where the folds are comprised of trials from mutually exclusive subjects in the dataset. These subject-wise folds are outlined below.

J Additional Features

J.1 Pre-processed Data Visualization

Pre-processing is an important aspect of the rPPG task that we hope to help standardize using our toolbox. It is advantageous to be able to quickly visualize and visually evaluate pre-processed image data and ground truth signals. Image data in particular can be especially useful to observe in order to inspect the effectiveness of out-of-the-box face detection and cropping techniques used in our toolbox, and to ultimately get an idea as to how much of the face region is visible in a given video. We provide simple Jupyter Notebooks for quickly visualizing image data and ground truth signals pre-processed by our toolbox. Further details regarding these notebooks can be found in our GitHub repo and the associated README.

H UBFC-Phys视频排除

为了在我们的主要论文中评估UBFC-Phys [27]数据集，并且默认情况下在我们的工具箱中，我们利用了所有三个任务和相同主题的排除，或者相反的子选择，数据集作者在他们的论文[27]的第二个补充材料中提供的列表。根据上述补充材料，我们排除了14名受试者 (s3, s8, s9, s26, s28, s30, s31, s32, s33, s40, s52, s53, s54, s56) 对于休息任务 (T1)，30名受试者 (s1, s4, s6, s8, s9, s11, s12, s13, s14, s19, s21, s22, s25, s26, s27, s28, s31, s32, s33, s35, s38, s39, s41, s42, s45, s47, s48, s52, s53, s55)，23名受试者 (s5, s8, s9, s10, s13, s14, s17, s22, s25, s26, s28, s30, s32, s33, s35, s37, s40, s47, s48, s49, s50, s52, s53) 进行算术任务 (T3)。

在我们的工具箱中，视频排除是使用配置文件中指定的数据集过滤条件来实现的。具体地，可以提供排除列表或任务选择列表以分别排除视频不被包括或选择特定任务作为数据集的一部分。

一、多任务培训和评估详情

为了展示这个工具箱如何扩展到生理多任务处理，我们实现了BigSmall [41]—一个多任务处理PPG，呼吸和面部动作的模型。在这里，我们重申[41]的信息，略有修改，以澄清。

I.1 交叉确认受试者折叠

Fold 1: F003, F004, F005, F006, F009, F017, F022, F028, F029, F031, F032, F033, F038, F044, F047, F048, F052, F053, F055, F061, F063, F067, F068, F074, F075, F076, F081, M003, M005, M006, M009, M012, M019, M025, M026, M028, M031, M036, M037, M040, M046, M047, M049, M051, M054, M056.  

Fold 2: F001, F002, F008, F018, F021, F025, F026, F035, F036, F037, F039, F040, F041, F042, F046, F049, F057, F058, F060, F062, F064, F066, F070, F071, F072, F073, F077, F082, M001, M002, M007, M013, M014, M016, M022, M023, M024, M027, M029, M030, M034, M035, M041, M042, M043, M048, M055.  

Fold 3: F007, F010, F011, F012, F013, F014, F015, F016, F019, F020, F023, F024, F027, F030, F034, F043, F045, F050, F051, F054, F056, F059, F065, F069, F078, F079, F080, M004, M008, M010, M011, M015, M017, M018, M020, M021, M032, M033, M038, M039, M044, M045, M050, M052, M053, M057, M058.  

用于模型训练和评估的Au子集（在此工具箱中）由包含动作单元标签的数据集子集组成。这包括来自每个受试者的以下任务的大约20秒的数据：T1, T6, T7, T8。

I.3 受试者折叠拆分

[41]使用3折交叉验证进行评估，其中折由数据集中互斥受试者的试验组成。这些主题折叠概述如下。

J附加功能

J.1 预处理数据可视化

预处理是rPPG任务的一个重要方面，我们希望使用我们的工具箱帮助标准化。有利的是能够快速地可视化和视觉地评估预处理的图像数据和地面实况信号。图像数据特别有用，可以用来观察我们工具箱中使用的开箱即用的人脸检测和裁剪技术的有效性，并最终了解给定视频中有多少人脸区域可见。我们提供简单的嵌入式笔记本，用于快速可视化图像数据和经过工具箱预处理的地面实况信号。关于这些笔记本的更多细节可以在我们的GitHub repo和相关的README中找到。

J.2 Training Loss, Validation Loss, and Learning Rate Visualization

The rPPG-Toolbox assumes certain defaults across most config files for supervised methods, including a default learning rate of 0.009 used alongside the Adam [38] or AdamW [39] optimizers, a criterion such as mean squared error (MSE) loss or Negative Pearson Correlation Loss, and the 1cycle learning rate scheduler [40] are utilized for training. An exception is with BigSmall [41], which uses a default learning rate of 0.001 that remains constant throughout training. It can be valuable to visualize losses such as those involved in training or validation phases. Furthermore, it may be useful to simultaneously visualize the learning rate, especially when users stray from the defaults in order to target an optimal set of training, validation, and testing parameters for their research efforts. The toolbox’s configs contain parameters that enable the visualization of the training loss, validation loss, and the learning rate for any given supervised method.

J.3 Bland-Altman Plots

We provide Bland-Altman plots as an additional metric in the rPPG-Toolbox. Users can enable the plots via an evaluation parameter in the config file, and will be given further options to configure the plots as the toolbox is refined and expanded. For more details, please refer to the GitHub repo and the associated README.

J.4 Motion Analysis

We also provide scripts that leverage OpenFace [44] for extracting, visualizing, and analyzing motion in rPPG video datasets. Specifically, we include a Python script to convert datasets into the .mp4 format for subsequent analysis by OpenFace, a shell script that leverages OpenFace to perform both rigid and non-rigid head motion analysis, and a separate Python script that plots exemplar plots that showcase comparisons of motion between different datasets. Further details can be found in our GitHub repo and the associated README.

J.2 训练损失、验证损失和学习率可视化

rPPG-Risk在大多数配置文件中假设了监督方法的某些默认值，包括与Adam [38]或AdamW [39]优化器一起使用的默认学习率0.009，标准如均方误差（MSE）损失或负皮尔逊相关损失，以及用于训练的1cycle学习率调度器[40]。一个例外是BigSmall [41]，它使用默认的学习率0.001，在整个训练过程中保持不变。可视化损失可能是有价值的，例如那些涉及训练或验证阶段的损失。此外，同时可视化学习率可能是有用的，特别是当用户偏离默认值时，以便为他们的研究工作瞄准最佳的训练，验证和测试参数集。该工具箱的参数包含参数，可以可视化任何给定监督方法的训练损失、验证损失和学习率。

J.3 Bland–Altman图

我们提供Bland–Altman图作为rPPG–Risk中的附加指标。用户可以通过配置文件中的评估参数启用绘图，并且随着工具箱的改进和扩展，将为用户提供更多选项来配置绘图。有关更多详细信息，请参阅GitHub repo和相关的README。

J.4 运动分析

我们还提供了利用OpenFace [44]提取、可视化和分析rPPG视频数据集中运动的脚本。具体来说，我们包括一个Python脚本，用于将数据集转换为.mp4格式，以便OpenFace进行后续分析，一个shell脚本，利用OpenFace执行刚性和非刚性头部运动分析，以及一个单独的Python脚本，用于绘制示例图，展示不同数据集之间的运动比较。更多细节可以在我们的GitHub repo和相关的README中找到。