SympCam: Remote Optical Measurement of Sympathetic Arousal

Björn Braun[1] Daniel McDuff[2] Tadas Baltrusaitis[3] Paul Streli[1] Max Moebus[1] Christian Holz[1]
[1]ETH Zurich [2]University of Washington [3]Microsoft
{bjoern.braun, paul.streli, max.moebus, christian.holz}@inf.ethz.ch,
dmcduff@uw.edu, tadas.baltrusaitis@microsoft.com

***Abstract*—Recent work has shown that a person's sympathetic arousal can be estimated from facial videos alone using basic signal processing. This opens up new possibilities in the field of telehealth and stress management, providing a non-invasive method to measure stress only using a regular RGB camera. In this paper, we present SympCam, a new 3D convolutional architecture tailored to the task of remote sympathetic arousal prediction. Our model incorporates a temporal attention module (TAM) to enhance the temporal coherence of our sequential data processing capabilities. The predictions from our method improve accuracy metrics of sympathetic arousal in prior work by 48% to a mean correlation of 0.77. We additionally compare our method with common remote photoplethysmography (rPPG) networks and show that they alone cannot accurately predict sympathetic arousal "out-of-the-box". Furthermore, we show that the sympathetic arousal predicted by our method allows detecting physical stress with a balanced accuracy of 90%—an improvement of 61% compared to the rPPG method commonly used in related work, demonstrating the limitations of using rPPG alone. Finally, we contribute a dataset designed explicitly for the task of remote sympathetic arousal prediction. Our dataset contains synchronized face and hand videos of 20 participants from two cameras synchronized with electrodermal activity (EDA) and photoplethysmography (PPG) measurements. We will make this dataset available to the community and use it to evaluate the methods in this paper. To the best of our knowledge, this is the first dataset available to other researchers designed for remote sympathetic arousal prediction.**

***Index Terms*—digital health, physiological computing**

I. Introduction

WEARABLE sensors, such as smart watches, continue to impact healthcare as they enable people to continuously and non-invasively measure physiological signals where they could not before. These sensors provide valuable data about an individual's health [1]–[4] and can serve to help detect cardiovascular disorders [5], stress [6] or pain [7]. While such wearable sensors have improved substantially, they have some limitations. They have to be worn on the body, are not easily scaled to an entire population, and usually only measure from one location on the body (e.g., the wrist) [8].

In contrast, cameras are versatile sensors that can unobtrusively capture spatial and temporal information at a distance. In addition, most of today's computers and mobile devices are equipped with user-facing cameras for the purpose of video telephony. These properties make cameras attractive as a means to measure physiological signals [9]. Examples of applications include remote patient monitoring [10] or stress measurement [11]. While, to date, some cardiac and pulmonary signals can be measured using a camera, including heart rate (HR) [12], [13], there are many signals for which there is little or no evidence that cameras alone are sufficient.

One important example is sympathetic arousal, which is a measure of the activation of the sympathetic nervous system. Following many different types of physical, mental, and/or emotional stressors, the sympathetic branch of the autonomic nervous system (ANS) is activated, leading to sympathetic arousal and "fight-or-flight" responses such as increased sweat responses [14]. Sympathetic arousal is, therefore, usually captured with the help of the electrodermal activity (EDA), which measures skin conductivity using electrodes placed on two different locations on the body. Multiple previous works have shown that changes in EDA are a reliable indicator of stress [6], [15] and pain [16], [17]. Traditionally, EDA has been measured at sites on the human body with a high density of sweat glands, such as the fingers or palms, using electrodes that are in steady contact with the person's skin [18]. Most recently, first works have shown that it is also possible to measure EDA and sympathetic arousal completely remotely from videos of the palm and the face [19], [20]. Bhamborea et al. [19] have directly measured EDA by counting the number of specular reflections on the palm. Braun et al. [20] were the first to infer sympathetic arousal from both the face and the palm by measuring blood perfusion, which they have shown correlates with contact EDA. Our goal was to advance the existing approaches and to provide the community with a dataset specifically designed for this task.

In this paper, we present a novel approach for measuring sympathetic arousal from facial videos leveraging neural networks for this task for the first time. Our method extends a 3D convolutional architecture with a temporal attention module (TAM) to learn spatial and temporal-domain features that lead to predictions that highly correlate with gold-standard contact EDA measurements. Our main contributions are:

- A 3D convolutional neural architecture that we tailored to the task of remote sympathetic arousal prediction by introducing a TAM and adapting the temporal dimension to the dynamics of the EDA signal. Using leave-one-subject-out (LOSO) cross-validation, we obtain a mean Spearman correlation of 0.77 between our predicted sympathetic arousal and the ground truth EDA signal, which is an

This work was partly funded by the Microsoft Swiss Joint Research Center.

improvement of 48% compared to previous work [20].

- A dataset with 20 participants on which we evaluate our model. It consists of videos of the face and hand synchronized with measurements of the EDA and PPG signals. We specifically designed the dataset for the task of remote sympathetic arousal assessment and now make it available on request to other researchers because we believe that it opens up new possibilities in the field of telehealth. To the best of our knowledge, this dataset is the first dataset available to other researchers designed to predict sympathetic arousal remotely.
- A classification model that leverages our predicted sympathetic arousal and a remote photoplethysmography (rPPG) signal for the task of detecting physical stress due to pain. We show on this task that our method outperforms rPPG-based approaches by 61%, achieving a balanced accuracy/F1 score of 0.9/0.83 in predicting whether a person is experiencing physical stress due to pain. We highlight the limitations of relying solely on the blood volume pulse (BVP) for detecting physical stress.

II. Related Work

Compared to self-reports, which are commonly used for stress detection, physiological sensing provides the opportunity for continuous temporal measurements. Traditionally, wearable sensors such as smartwatches were used for physiological measurement. Recently, non-contact (remote) methods, which only use a regular RGB camera, have gained popularity due to their potential for scalability and comfortability [9]. To date, a vast majority of the work in the field of remote physiological sensing has focused on measuring cardiopulmonary signals such as the BVP or the breathing rate. The BVP is inferred from the rPPG signal, which is calculated by measuring the peripheral blood flow via light reflected from the skin [21]–[24]. However, recent work has shown that the HR and other extracted features from the BVP, such as heart rate variability (HRV), are influenced by both sympathetic and parasympathetic activity [25] and, therefore, give only limited information about a person's sympathetic activity [26], [27]. EDA, on the other hand, which measures a person's sweat response, is considered a direct marker of sympathetic activity and is commonly used for psychophysiological evaluations [28], [29].

Previous work using minimally invasive methods has shown that repeated arousal stimuli induced by electrical stimulation are followed by an increase in sympathetic nerve activity, blood flow, and EDA in the forehead [30]. Other work obtained similar results and found that facial blood flow changes due to pain are not dependent on regional (orofacial) stimulation to occure [31]. Furthermore, analysis of thermal imagery found that arousal-induced sweat responses can be detected without contact with the body using thermal cameras [32]. However, thermal cameras are not widely available. Bhamborea et al. [19] published early proof-of-concept results that EDA could be inferred from the palm using only an RGB camera by counting the specular reflections from the skin. Building on that work [19] and the correlation between blood flow and EDA on the forehead [30], [31], subsequent work has shown that sympathetic arousal can also be inferred from videos of the face using only a regular RGB camera by measuring the peripheral blood flow to the forehead [20]. While this work showed first proof that it is possible to remotely extract a person's sympathetic arousal from a video of the face, the mean correlations across participants were moderate, and the standard deviation (STD) was high. Nevertheless, we were inspired by these results and aimed to develop a more robust method and release a dataset for remote sympathetic arousal prediction that is also available to other researchers. As this previous work [20] indicates that they measure changes in blood flow that correlate with EDA, we will refer to remotely measuring sympathetic arousal instead of measuring EDA.

For camera-based physiological measurements, such as rPPG, supervised neural architectures are state-of-the-art [22], [23], [33]. The spatial information to predict sympathetic arousal from the face should be similar to the spatial information for rPPG prediction. However, the typical frequency band of the sympathetic component of the EDA signal is between 0.045–0.25 Hz [34], which is considerably lower than the frequency band of 0.7–2.5 Hz from the HR [33], [35]. Therefore, we build upon previous work [23], which utilizes 3D convolutional layers to learn temporal domain features and adapt our network architecture such that it can capture the slow-changing temporal characteristics of the EDA signal.

III. Dataset

A. Recruiting and Recording

We recorded a dataset of $N = 20$ participants (5 female, 15 male, ages 19–36, $\mu = 26.7$ and $\sigma = 3.9$) to investigate remote sympathetic arousal prediction. Based on the Fitzpatrick scale [36], 5 participants had skin type II, 10 skin type III, 3 skin type V, and 2 skin type VI. The dataset captures 9.5 minutes of video recording (with disabled white balancing, autofocus, and auto-exposure) of participants' faces and hands, with synchronized EDA recordings from the finger and PPG recordings from the fingers and foreheads.

B. Apparatus

The participants placed their heads on a chin rest and their hands on a table with their palms facing upwards (see Fig. 1). The hands were secured with a belt over the thumb to minimize any motion in the videos and we kept the lighting and temperature in the room constant throughout the study. We recorded the videos using two Basler acA1300-200uc cameras pointed toward the participants' faces and hands and the physiological signals from a synchronized BIOPAC MP160 that triggered the cameras by wire. The design of the study is based on that of previous work [20], as they have shown successfully that with their setup, it is possible to remotely measure sympathetic arousal. To the best of our knowledge, our dataset is the first for remote sympathetic arousal prediction that will be available to other researchers.

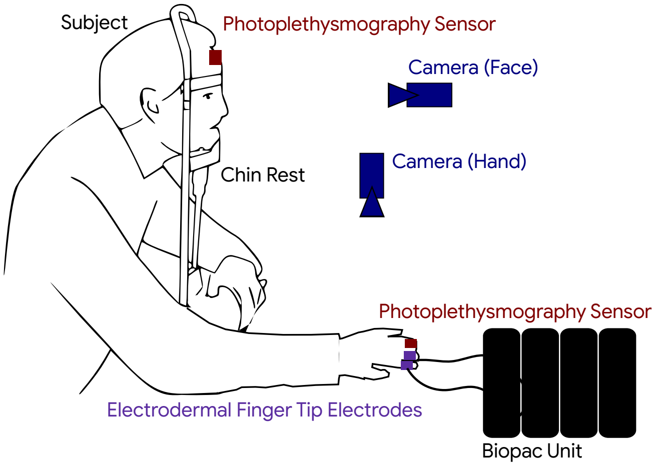
Subject
Photoplethysmography Sensor
Camera (Face)
Camera (Hand)
Chin Rest
Photoplethysmography Sensor
Electrodermal Finger Tip Electrodes
Biopac Unit

Fig. 1. The apparatus used for our study.

C. Study Protocol

The protocol alternated between periods of resting (2 minutes) and periods of physical stress due to pain during which the participants pinched their skin (self-pinching, 30 seconds) to stimulate an EDA response, starting with a period of rest. To ensure that all participants had significant changes in EDA during the study, we used a dependent t-test for paired samples ($p > 0.05$), comparing the EDA levels during the periods of rest to the periods of stress. Two (10%) of the 20 participants did not have significant EDA responses during the study (consistent with previous work [20], [30]), and we did not consider them for the following evaluation.

IV. Methods

Our work comprises two main components. First, we aim to remotely estimate a person's sympathetic arousal corresponding to a contact-based EDA signal using only facial videos as input. We use a deep-learning-based approach as these approaches have shown stronger performance than signal-processing-based approaches for related problems such as rPPG prediction [22], [23], [33], [37]. Second, we train a classification model that uses our remotely predicted signals to detect if a person experiences physical stress due to pain.

A. Remote Sympathetic Arousal Prediction

1) Proposed Architecture: As the backbone of our architecture, we use a 3D CNN [23] with a temporal input length of $T = 768$ frames. While such 3D CNN-based architectures have achieved impressive performance for the task of video-based HR prediction, most of the used 3D CNN architectures [23], [33] treat all input frames equally, ignoring that different frames may provide different contributions to the target prediction. For example, frames with more motion might convey less information than frames with less motion. To address this problem, we propose a temporal attention module (TAM) that allows our model to learn to discriminate between more and less important features along the temporal dimension. Each TAM block is composed of a 3D average pooling, a 3D convolutional layer (kernel size 1, stride 1, padding 0), and finally, a multi-layer perceptron (MLP, using the ReLU activation function) with a reduction rate $r = 16$ followed by a sigmoid activation function (see Fig. 2). Given an image feature map $\mathbf{F_{in}} \in \mathbb{R}^{C \times T \times w \times h}$ as input, a TAM block infers a 1D temporal attention map $\mathbf{A_t} \in \mathbb{R}^{T}$, which is then broadcasted (copied) along the spatial and channel dimension during multiplication. The final output $\mathbf{F_{out}}$ of the attention process can be summarized as:

$$\mathbf{F_{out}} = \mathbf{A_t} \otimes \mathbf{F_{in}}, \tag{1}$$

where $\otimes$ denotes element-wise multiplication. We add one TAM block before each temporal up-sampling step. This approach is inspired by the success of sequentially using attention maps along the channel and spatial dimensions [38]. Fig. 2 shows the final architecture and the proposed TAM block. The total number of FLOPs for one batch using our model is about 100 GigaFlops and the total number of parameters of our model is about 790 k. Of these, the two TAM blocks account for about 23 k parameters, which represents only about 3% of the total number of parameters.

2) Implementation Details: First, we detect the participants' faces using OpenCV's cascade classifier [39], crop the images to the bounding boxes, and then resize the images to a resolution of 72×72 (similarly as in previous work [20], [23]). Then, we downsample the frame rate of the videos from 100 Hz to 10 Hz as a frame rate of 10 Hz is sufficient for the typical frequency band of the sympathetic component of the EDA signal (between 0.045–0.25 Hz [34]). Finally, we take the consecutive difference between the frames and standardize them by dividing them through the STD of the pixel intensity values [22]. We process the ground truth EDA signals in the same fashion and use them as labels for our model.

Afterward, we trained our model using leave-one-subject-out (LOSO) cross-validation, during which we iteratively held out the data of one participant as test set, one as validation set, and use the data of the remaining participants as training set. We used a batch size of 4 for 30 epochs, a learning rate of 0.001, and the mean squared error as the loss function. To validate the stability of the models, we report the mean obtained correlations across all participants and three random seeds, which helps to ensure that our experiments are reproducible and do not depend on a single random initialization, such as the weight initialization of the neural network. We trained our model on a GeForce RTX 4090 with a runtime of about 9 hours for all subjects and three random seeds.

3) Evaluation: To compare the similarity between our predicted and the ground truth signal, we use the Spearman correlation as proposed in previous work [20]. We evaluate the performance of our model to predict the raw EDA signal and the slower-acting tonic component of the EDA signal, which we obtain using the convex optimization approach [40]. As previous work has found that longer temporal inputs help to remotely predict sympathetic arousal [20], we also evaluate how different input window sizes ranging from $T = 256$ to $T = 1024$ frames (corresponding to 25.6 to 102.4 seconds) influence the performance of our network. In addition, we

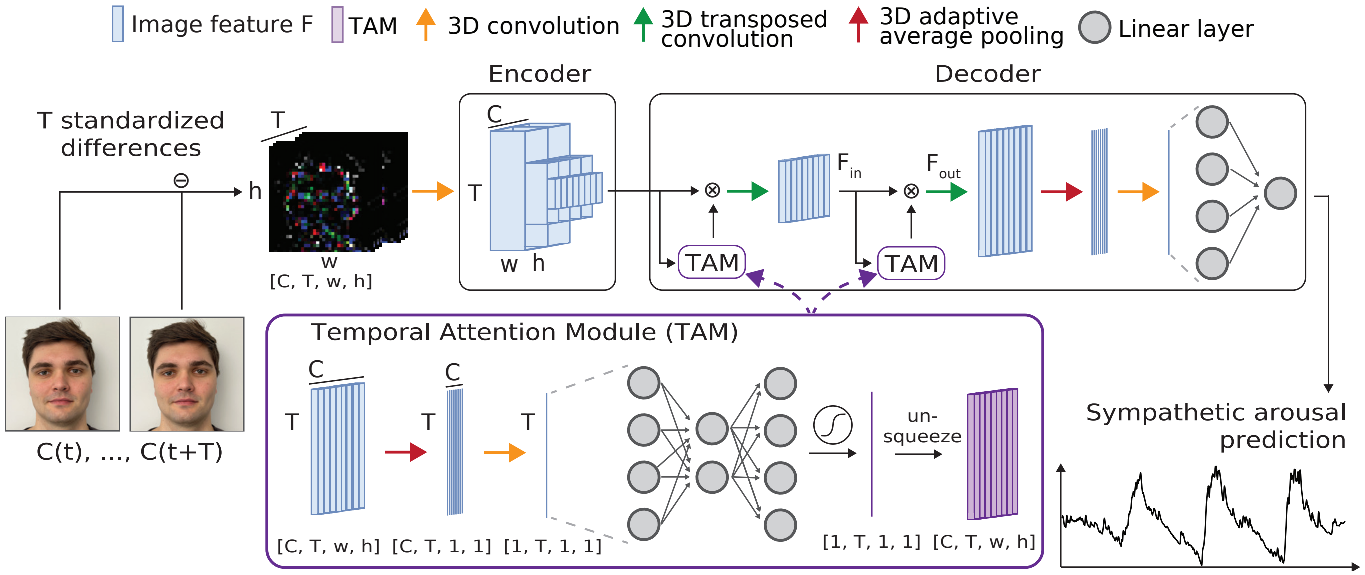
Image feature F
TAM
3D convolution
3D transposed convolution
3D adaptive average pooling
Linear layer
Encoder
Decoder
T standardized differences
⊖
h
T
w
[C, T, w, h]
C
T
w h
⊗
F in
⊗
F out
TAM
TAM
Temporal Attention Module (TAM)
C
T
C
T
T
[C, T, w, h] [C, T, 1, 1] [1, T, 1, 1]
un-squeeze
[1, T, 1, 1] [C, T, w, h]
C(t), ..., C(t+T)
Sympathetic arousal prediction

Fig. 2. Our proposed neural architecture to predict sympathetic arousal from facial videos.

implement four baseline networks (one Transformer-based model and three CNNs) that are used for remote HR prediction to compare the performance of our proposed model to the performance of these established networks. We trained all models in the same fashion as our proposed model. Furthermore, to compare our method with the current baseline, we also implemented the *blood pulsation amplitude* method (current baseline, a signal-processing-based approach) [20] and evaluated it on our dataset. A valid concern is that our model learns to predict small facial motions, such as micro-expressions, which could potentially occur during phases of stress. We, therefore, also calculate the Spearman correlation between our predicted signals and the magnitude of the optical flow of the face.

TABLE I
THE CALCULATED STATISTICAL FEATURES FROM THE PPG/RPPG AND THE CONTACT EDA/ PREDICTED SYMPATHETIC AROUSAL SIGNALS.

Signal	Feature	Description
PPG/ rPPG	μ_{HR}	Mean HR
	min_{HR}	Minimum HR
	max_{HR}	Maximum HR
	μ_{HRChange}	Mean change of the HR
	SDNN	STD of NN intervals
EDA/ predicted sympathetic arousal	μ_{EDA}	Mean
	σ_{EDA}	STD
	min_{EDA}	Minimum value
	max_{EDA}	Maximum value
	$\mu_{\text{EDAChange}}$	Mean of consecutive change

B. Physical Stress Detection

To assess how much value our remotely predicted sympathetic arousal signal adds to the downstream task of physical stress detection due to pain, we perform a classification on the used dataset. The goal of the classification is to distinguish between the non-stressful periods (resting) and the 30 second stressful periods (self-pinching). We perform the classification separately using only the ground truth signals obtained from the contact measurement device (EDA and PPG) and the camera-based predicted signals (our remotely predicted sympathetic arousal trained with tonic EDA and a remotely predicted rPPG signal). To predict the rPPG signal from the camera, we use the original PhysNet network [23] trained on our dataset with a LOSO cross-validation approach. For both modalities, the contact-based and camera-based inputs, we each run the stress detection once using only the PPG/ rPPG signal, once using only the EDA/ remote sympathetic arousal signal, and once with both signals together. In this way, we aim to analyze the importance of the EDA/remote sympathetic arousal signal compared to the PPG/rPPG signal in this study setting. We divide the 9.5 minute recordings of each participant into 19 windows (3 windows of pinching and 16 windows of resting) of 30 seconds each without overlap. For each window, we extract 10 commonly used features for stress detection from the PPG/rPPG and EDA/sympathetic arousal signals, such as the mean HR or mean EDA (see Table I) [41]–[43]. As a classifier, we use the Gradient Boosting (GB) classifier. To train our classification algorithms, we again use LOSO cross-validation. Given the unbalanced number of stress and rest windows, we compute the balanced accuracy and F1-score to evaluate our model performance.

V. RESULTS

A. Remote Sympathetic Arousal Prediction

We show the mean correlation ρ and the standard deviation of our results across all participants and three different random seeds in Table II. Using our proposed model with an input window size of 768 frames, we obtained a mean correlation of 0.73 ± 0.01 (raw EDA)/0.77 ± 0.02 (tonic EDA) between our predicted signal and the ground truth EDA signal over all participants. This is an improvement of 40%/48% compared to using the current baseline method (Traditional Method [20], a signal-processing-based approach) on our dataset. Also, the STD decreases using our method from 0.24/0.26 to 0.19/0.20.

We further evaluated the influence of the input window size on the model performance. The mean correlation gradually increases from a mean correlation of 0.37/0.47 using a window size of 256 frames to a mean correlation of 0.73/0.77 using a window size of 768 frames. When using a larger window size of 1024 frames, the mean correlation decreases to 0.65/0.71. The other implemented network structures, which are usually used for rPPG measurements, showed much lower performance than our introduced model, with mean correlations between 0.23 and 0.51. Furthermore, the Spearman correlation between the calculated magnitude of the dense optical flow of the facial video and our predicted signals is 0.23, indicating that our model does not simply learn to predict facial motions.

To qualitatively cross-check the results, we plotted our predicted sympathetic arousal and the ground truth EDA signals for all participants. In Fig. 3, we show four predicted signals and the corresponding ground truth signals. We can see that our predicted signals closely follow the overall trend of the ground truth signal. However, for individual participants, our model is currently only capable of accurately predicting the global trend and not smaller phasic changes, as we show in the bottom-right plot of Fig. 3.

B. Physical Stress Detection

Table III summarizes the results of our physical stress (due to pain) classification using the GB classifier. A simple baseline classifier, which always predicts rest, would achieve a balanced accuracy (BACC) of 0.5 and an F1 score of 0.4. We obtain very similar maximum BACC and F1 scores for both modalities, the camera-based signals and the predicted camera-based signals. For both modalities, we obtain the highest BACC using only the features from the EDA/our remotely predicted sympathetic arousal signal. With only the contact-based signals, the highest BACC is 0.94, and the highest F1 score is 0.89. For the camera-based signals, the highest BACC is 0.90, and the highest F1 score is 0.83. However, for both the contact-based and camera-based signals, the BACC and F1 scores drop considerably when using only the PPG/rPPG signal compared to using the EDA/remotely predicted sympathetic arousal. For the contact-based signals, the BACC drops to 0.57 and the F1 score to 0.18 and for the camera-based signals, the BACC drops to 0.56 and the F1 score to 0.17. Using our remotely predicted sympathetic arousal improves the balanced accuracy by 61% compared to only using the remotely predicted rPPG signal.

VI. Discussion

A. Remote Sympathetic Arousal Prediction

In our quantitative analysis, we have shown that we achieve a 40% (raw EDA)/48% (tonic EDA) higher mean correlation of 0.73/0.77 across all participants predicting the raw/tonic EDA signal using our introduced model compared to the current state-of-the-art method, which uses a signal processing approach [20]. At the same time, we decrease the STD of the correlation across all participants from 0.24/0.26 to 0.19/0.20. Our qualitative analysis (see Fig. 3) also shows how closely our predicted sympathetic arousal follows the global trend of the ground truth EDA signal. Furthermore, we see a substantial performance improvement of our network using the TAM blocks compared to using other 3D CNN architectures like PhysNet [23]. This indicates that the TAM blocks help to learn the network to discriminate between more and less important features. Also, we evaluated the performance of our network using different window input sizes. The mean correlation gradually increases from a mean correlation of 0.37/0.47 using a window size of 256 frames (corresponding to 25.6 seconds) to a mean correlation of 0.73/0.77 using a window size of 768 frames (corresponding to 76.8 seconds) and then decreases again. This is consistent with previous work that obtained the best performance using a window size of 60 seconds [20]. In addition, the main spectral power density of an EDA signal lies in the frequency band between 0.045–0.25 Hz (corresponding to 4 to 22.2 seconds) [34], indicating that a window size of 22.2 seconds is beneficial to predict EDA. Our qualitative analysis shows that our obtained correlations of 0.73/0.77 indeed reflect that our predicted signals capture the overall trend of the ground-truth EDA signals accurately. However, we also recognize that our model is not yet capable of predicting smaller changes, and while our proposed model considerably improved the mean performance, a standard deviation of 0.19/0.20 still means that our model is not yet able to accurately predict sympathetic arousal for individual participants. Additionally, to show that our network does not learn to simply predict motions that could occur during phases of stress, we have calculated the correlation between our predicted signals and the magnitude of the dense optical flow of the videos of the participants' faces. We have obtained a mean correlation of only 0.23 across all participants, indicating that our network does not simply predict motion.

Finally, previous work suggests that we do not measure actual sweat responses when predicting sympathetic arousal from facial videos but changes in blood flow that correlate with sympathetic arousal (see Section II) [20]. Therefore, we expect that our measured signal from the face and the EDA signal do not perfectly match, e.g., due to small temporal offsets between the two signals. However, as we discuss below, we believe that our stress classification results show that for the downstream task of detecting physical stress, we do not need to be able to reconstruct the ground truth EDA signal perfectly.

B. Physical Stress Detection

To help reveal where the predicted sympathetic arousal has utility for downstream inferences, we performed a physical stress (due to pain) detection experiment. Using camera-based predicted sympathetic arousal, we obtain a maximum BACC/F1 score of 0.90/ 0.83 for detecting physical stress due to pain with a GB classifier using our remotely predicted sympathetic arousal. This is almost as high as using the contact-based EDA signal with a maximum BACC/F1 score of 0.94/ 0.89. Furthermore, we see in Table III that using only

TABLE II

MEAN AND STD OF THE SPEARMAN CORRELATIONS (ρ) OVER THREE RANDOM SEEDS BETWEEN THE PREDICTED SYMPATHETIC AROUSAL AND THE GROUND TRUTH EDA SIGNALS ACROSS ALL PARTICIPANTS. OUR METHOD IMPROVES THE MEAN PERFORMANCE BY 40% (RAW EDA)/48% (TONIC EDA) TO 0.73/0.77 AND DECREASES THE STD BY 0.05/0.06 COMPARED TO THE TRADITIONAL METHOD (CURRENT BASELINE) [20].

		Raw		**Tonic**	
Method	**Window**	**Mean ρ**	**STD**	**Mean ρ**	**STD**
TS-CAN [33]	768	0.23 ± 0.03	0.32 ± 0.02	0.33 ± 0.05	0.33 ± 0.04
PhysFormer [37]	768	0.28 ± 0.02	0.24 ± 0.02	0.23 ± 0.02	0.28 ± 0.03
DeepPhys [22]	768	0.31 ± 0.05	0.26 ± 0.02	0.39 ± 0.06	0.28 ± 0.03
PhysNet [23]	768	0.43 ± 0.02	0.25 ± 0.03	0.51 ± 0.01	0.31 ± 0.01
Traditional Method [20]	—	0.52	0.24	0.52	0.26
Ours	256	0.37 ± 0.06	0.22 ± 0.04	0.47 ± 0.03	0.21 ± 0.04
Ours	384	0.49 ± 0.04	0.25 ± 0.00	0.58 ± 0.03	0.26 ± 0.02
Ours	512	0.64 ± 0.00	0.23 ± 0.02	0.63 ± 0.00	0.29 ± 0.03
Ours	**768**	$\mathbf{0.73} \pm 0.01$	$\mathbf{0.19} \pm 0.01$	$\mathbf{0.77} \pm 0.02$	$\mathbf{0.20} \pm 0.01$
Ours	1024	0.65 ± 0.00	0.28 ± 0.02	0.71 ± 0.02	0.29 ± 0.01
Improvement of ours over best previous method		**+0.21**	**−0.05**	**+0.25**	**−0.06**

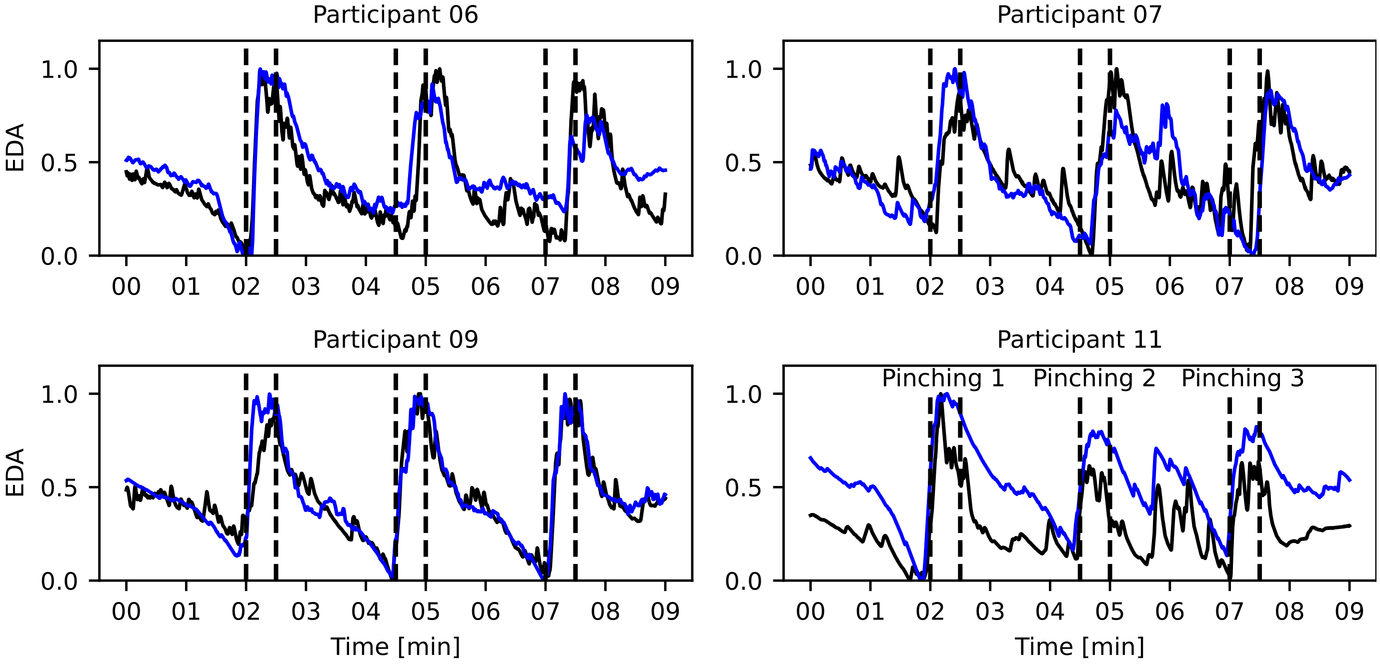

Fig. 3. Visual comparison between our predicted sympathetic arousal (blue) and the ground truth tonic EDA signal (black) for four participants. At minutes 2, 4.5, and 7, the participants are instructed to pinch themselves for 30 seconds to cause a sympathetic stress response. Note that for individual participants, such as participant 11, the model is currently only able to predict the global trend accurately. This difference is attributed to the nature of our method, which estimates sympathetic arousal by analyzing blood flow changes rather than measuring absolute EDA values.

either the contact-based PPG signals or only the remotely predicted rPPG signals, an accurate physical stress (due to pain) prediction is not possible for our dataset. A relatively modest stimulus like self-pinching does not seem to change features related to the blood volume pulse (BVP), such as heart rate, enough to allow for an accurate physical stress prediction. Very similar results were reported in related works. With contact-based signals, accuracies between 0.51 and 0.75 were obtained using only heart rate features (compared to 0.57 for our dataset) and accuracies of up to 0.95 with multi-modal features using the heart rate and the EDA signal to detect pain (compared to 0.94 for our dataset) [41], [44]. Previous work also obtained similar results using only features obtained from a remotely predicted rPPG signal, achieving accuracies between 0.59 and 0.63 (compared to 0.56 for our dataset) [45], [46]. This highlights the limitations of relying solely on BVP and the importance of having a second physiological metric, such as our proposed remote sympathetic arousal, to detect physical stress.

C. Limitations and Future Work

We recognize that our model and dataset have certain limitations. First, our study is highly controlled. To minimize any motion or lighting changes, we placed the participants' heads on a chin rest and kept the lighting constant in the room. Therefore, only limited conclusions can be drawn about the

TABLE III

BALANCED ACCURACIES (BACC) AND F1 SCORES USING ONLY THE CONTACT-BASED SIGNALS (PPG AND EDA), ONLY OUR CAMERA-BASED PREDICTED SIGNALS (RPPG AND REMOTE SYMPATHETIC AROUSAL (RSA)), AND BOTH TOGETHER. NOTE THAT WE CAN ONLY PREDICT PHYSICAL STRESS ACCURATELY USING EDA/OUR PREDICTED RSA AND NOT USING THE PPG/RPPG SIGNALS.

	PPG/rPPG	EDA/rSA	BACC	F1
Contact (reference measurement)	✓	✗	0.57	0.18
	✗	✓	**0.94**	**0.89**
	✓	✓	0.93	0.88
Camera (estimated)	✓	✗	0.56	0.17
	✗	✓	**0.90**	**0.83**
	✓	✓	0.88	0.80
Baseline	—	—	0.50	0.40

generalizability of our approach to more real-world situations. However, we believe that it is essential to first establish a dataset that makes it possible to evaluate the feasibility of possible methods under more controlled conditions. In future work, we aim to extend our dataset to include more natural scenarios and to evaluate our model's robustness to such conditions. Second, while we showed in our qualitative analysis that our model can predict the global trend of the EDA signal, we also found that it is not yet capable of predicting the small phasic components. We believe that one promising approach could be to investigate different loss functions, which give greater weight to the errors of the phasic component. This could help to improve the model's sensitivity to the more rapid phasic fluctuations. Finally, our dataset comprises data from 5 female and 15 male participants. We acknowledge that we should have paid more attention to a balanced ratio to create a balanced dataset and to potentially also allow for an investigation of the performance differences of our approach for female and male participants. We aim to correct this oversight in the future by recording further participants.

D. Broader Impacts

Perhaps the most obvious application for measuring sympathetic arousal remotely is stress management. Previous work has shown that EDA/sympathetic arousal measurements can be used to help people better understand their stress patterns [47]. In addition, currently used wearable sensors, such as smartwatches, can be very inconvenient for the user, or it might not be possible to wear them for safety reasons, e.g., for assembly line workers. Using a camera in such cases could offer unique opportunities for deployment, otherwise hard to achieve. Finally, we believe that it is important to consider the potential for a new technology such as ours to be deployed with negligence or by a bad actor. While people might be able to hide their emotions by not expressing them, they are, in general, not able to control their physiological states. Therefore, it is important that there are mechanisms in place to be aware of remote measurements and consent to them.

VII. CONCLUSION

In this paper, we have demonstrated that it is possible to improve the performance of remote sympathetic arousal prediction from a video of a person's face by using a 3D CNN architecture tailored to the temporal dynamics of sympathetic arousal. To evaluate our approach, we contribute the first dataset specifically designed for the task of remotely predicting sympathetic arousal and make it available to other researchers. Using LOSO cross-validation, we have demonstrated that our proposed network obtains a mean correlation of up to 0.73 (raw EDA)/0.77 (tonic EDA) between the predicted sympathetic arousal and the ground truth EDA signal, marking a 40%/48% improvement compared to previous work. However, we also recognize that our model is not yet capable of predicting more detailed phasic changes of the EDA signal for individual participants. Furthermore, we trained a GB classifier with features extracted from the contact EDA and PPG signals and our remotely predicted sympathetic arousal and rPPG signals to detect physical stress caused by pain. We achieved a mean BACC of 0.90 in predicting physical stress using only our remotely predicted signals. Our stress classification experiments also revealed that using contact PPG and remotely predicted rPPG signals alone does not yield accurate results for physical stress detection due to pain in our dataset. This underlines the importance of our proposed approach to offer an alternative physiological measurement to accurately predict physical stress. We hope that our contributed network and dataset can assist other researchers in exploring various signal processing and machine learning techniques for developing accurate remote sympathetic arousal prediction models.

REFERENCES

[1] D. Spathis, I. Perez-Pozuelo, S. Brage, N. J. Wareham, and C. Mascolo, "Self-supervised transfer learning of physiological representations from free-living wearable data," in *Proceedings of the Conference on Health, Inference, and Learning*, 2021, pp. 69–78.

[2] G. Dong, L. Cai, D. Datta, S. Kumar, L. E. Barnes, and M. Boukhechba, "Influenza-like symptom recognition using mobile sensing and graph neural networks," in *Proceedings of the conference on health, inference, and learning*, 2021, pp. 291–300.

[3] M. Moebus, S. Gashi, M. Hilty, P. Oldrati, and C. Holz, "Meaningful digital biomarkers derived from wearable sensors to predict daily fatigue in multiple sclerosis patients and healthy controls," *Iscience*, vol. 27, no. 2, 2024.

[4] M. Moebus and C. Holz, "Personalized interpretable prediction of perceived sleep quality: Models with meaningful cardiovascular and behavioral features," *Plos one*, vol. 19, no. 7, p. e0305258, 2024.

[5] J. Dunn, R. Runge, and M. Snyder, "Wearables and the medical revolution," *Personalized medicine*, vol. 15, no. 5, pp. 429–448, 2018.

[6] N. Sharma and T. Gedeon, "Objective measures, sensors and computational techniques for stress recognition and classification: A survey," *Computer methods and programs in biomedicine*, vol. 108, no. 3, pp. 1287–1301, 2012.

[7] J. Chen, M. Abbod, and J.-S. Shieh, "Pain and stress detection using wearable sensors and devices—a review," *Sensors*, vol. 21, no. 4, p. 1030, 2021.

[8] J. F. Knight, D. Deen-Williams, T. N. Arvanitis, C. Baber, S. Sotiriou, S. Anastopoulou, and M. Gargalakos, "Assessing the wearability of wearable computers," in *2006 10th IEEE International Symposium on Wearable Computers*, 2006, pp. 75–82.

[9] D. McDuff, "Camera measurement of physiological vital signs," *ACM Computing Surveys*, vol. 55, no. 9, pp. 1–40, 2023.

[10] A. M. Ansary, J. N. Martinez, and J. D. Scott, "The virtual physical exam in the 21st century," *Journal of Telemedicine and Telecare*, vol. 27, no. 6, pp. 382–392, 2021, pMID: 31690169. [Online]. Available: https://doi.org/10.1177/1357633X19878330

[11] D. J. McDuff, J. Hernandez, S. Gontarek, and R. W. Picard, "Cogcam: Contact-free measurement of cognitive stress during computer tasks with a digital camera," in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, 2016, pp. 4000–4004.

[12] W. Verkruysse, L. O. Svaasand, and J. S. Nelson, "Remote plethysmographic imaging using ambient light." *Optics express*, vol. 16, no. 26, pp. 21 434–21 445, 2008.

[13] M.-Z. Poh, D. McDuff, and R. W. Picard, "Non-contact, automated cardiac pulse measurements using video imaging and blind source separation." *Optics express*, vol. 18, no. 10, pp. 10 762–10 774, 2010.

[14] A. S. P. Jansen, X. V. Nguyen, V. Karpitskiy, T. C. Mettenleiter, and A. D. Loewy, "Central command neurons of the sympathetic nervous system: Basis of the fight-or-flight response," *Science*, vol. 270, no. 5236, pp. 644–646, 1995. [Online]. Available: https://www.science.org/doi/abs/10.1126/science.270.5236.644

[15] G. Giannakakis, D. Grigoriadis, K. Giannakaki, O. Simantiraki, A. Roniotis, and M. Tsiknakis, "Review on psychological stress detection using biosignals," *IEEE Transactions on Affective Computing*, vol. 13, no. 1, pp. 440–460, 2019.

[16] A. Mobascher, J. Brinkmeyer, T. Warbrick, F. Musso, H.-J. Wittsack, R. Stoermer, A. Saleh, A. Schnitzler, and G. Winterer, "Fluctuations in electrodermal activity reveal variations in single trial brain responses to painful laser stimuli—a fmri/eeg study," *Neuroimage*, vol. 44, no. 3, pp. 1081–1092, 2009.

[17] B. T. Susam, M. Akcakaya, H. Nezamfar, D. Diaz, X. Xu, V. R. de Sa, K. D. Craig, J. S. Huang, and M. S. Goodwin, "Automated pain assessment using electrodermal activity data and machine learning," in *2018 40th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC)*. IEEE, 2018, pp. 372–375.

[18] M. van Dooren, J. G.-J. de Vries, and J. H. Janssen, "Emotional sweating across the body: Comparing 16 different skin conductance measurement locations," *Physiology & Behavior*, vol. 106, no. 2, pp. 298–304, 2012. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0031938412000613

[19] M. J. Bhamborae, P. Flotho, A. Mai, E. N. Schneider, A. L. Francis, and D. J. Strauss, "Towards contactless estimation of electrodermal activity correlates," in *2020 42nd Annual International Conference of the IEEE Engineering in Medicine & Biology Society (EMBC)*. IEEE, 2020, pp. 1799–1802.

[20] B. Braun, D. McDuff, T. Baltrusaitis, and C. Holz, "Video-based sympathetic arousal assessment via peripheral blood flow estimation," *Biomedical Optics Express*, vol. 14, no. 12, pp. 6607–6628, 2023.

[21] M. Huelsbusch and V. Blazek, "Contactless mapping of rhythmical phenomena in tissue perfusion using ppgi," in *Medical Imaging 2002: Physiology and Function from Multidimensional Images*, vol. 4683. International Society for Optics and Photonics, 2002, pp. 110–117.

[22] W. Chen and D. McDuff, "Deepphys: Video-based physiological measurement using convolutional attention networks," in *Proceedings of the european conference on computer vision (ECCV)*, 2018, pp. 349–365.

[23] Z. Yu, X. Li, and G. Zhao, "Remote photoplethysmograph signal measurement from facial videos using spatio-temporal networks," *arXiv preprint arXiv:1905.02419*, 2019.

[24] B. Braun, D. McDuff, and C. Holz, "How suboptimal is training rppg models with videos and targets from different body sites?" *arXiv preprint arXiv:2403.10582*, 2024.

[25] G. A. Reyes del Paso, W. Langewitz, L. J. M. Mulder, A. van Roon, and S. Duschek, "The utility of low frequency heart rate variability as an index of sympathetic cardiac tone: A review with emphasis on a reanalysis of previous studies," *Psychophysiology*, vol. 50, no. 5, pp. 477–487, 2013. [Online]. Available: https://onlinelibrary.wiley.com/doi/abs/10.1111/psyp.12027

[26] G. Billman, "The lf/hf ratio does not accurately measure cardiac sympatho-vagal balance," *Frontiers in Physiology*, vol. 4, 2013. [Online]. Available: https://www.frontiersin.org/articles/10.3389/fphys.2013.00026

[27] B. L. Thomas, N. Claassen, P. Becker, and M. Viljoen, "Validity of commonly used heart rate variability markers of autonomic nervous system function," *Neuropsychobiology*, vol. 78, no. 1, pp. 14–26, Feb. 2019.

[28] M. E. Dawson, A. M. Schell, and D. L. Filion, *The Electrodermal System*, 4th ed., ser. Cambridge Handbooks in Psychology. Cambridge University Press, 2016, p. 217–243.

[29] H. F. Posada-Quintero and K. H. Chon, "Innovations in electrodermal activity data collection and signal processing: A systematic review," *Sensors*, vol. 20, no. 2, 2020. [Online]. Available: https://www.mdpi.com/1424-8220/20/2/479

[30] M. Nordin, "Sympathetic discharges in the human supraorbital nerve and their relation to sudo- and vasomotor responses." *The Journal of Physiology*, vol. 423, no. 1, pp. 241–255, 1990. [Online]. Available: https://physoc.onlinelibrary.wiley.com/doi/abs/10.1113/jphysiol.1990.sp018020

[31] O. Vassend and S. Knardahl, "Effects of repeated electrocutaneous pain stimulation on facial blood flow," *Biological Psychology*, vol. 68, no. 2, pp. 163–178, 2005. [Online]. Available: https://www.sciencedirect.com/science/article/pii/S0301051104000985

[32] D. Shastri, M. Papadakis, P. Tsiamyrtzis, B. Bass, and I. Pavlidis, "Perinasal imaging of physiological stress and its affective potential," *IEEE Transactions on Affective Computing*, vol. 3, no. 3, pp. 366–378, 2012.

[33] X. Liu, J. Fromm, S. Patel, and D. McDuff, "Multi-task temporal shift attention networks for on-device contactless vitals measurement," *Advances in Neural Information Processing Systems*, vol. 33, pp. 19 400–19 411, 2020.

[34] H. F. Posada-Quintero, J. P. Florian, A. D. Orjuela-Cañón, T. Aljama-Corrales, S. Charleston-Villalobos, and K. H. Chon, "Power spectral density analysis of electrodermal activity for sympathetic function assessment," *Annals of biomedical engineering*, vol. 44, no. 10, pp. 3124–3135, 2016.

[35] D. H. Spodick, P. Raju, R. L. Bishop, and R. D. Rifkin, "Operational definition of normal sinus heart rate," *The American journal of cardiology*, vol. 69, no. 14, pp. 1245–1246, 1992.

[36] T. B. Fitzpatrick, "The validity and practicality of sun-reactive skin types i through vi," *Archives of dermatology*, vol. 124, no. 6, pp. 869–871, 1988.

[37] Z. Yu, Y. Shen, J. Shi, H. Zhao, P. H. Torr, and G. Zhao, "Physformer: Facial video-based physiological measurement with temporal difference transformer," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2022, pp. 4186–4196.

[38] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "Cbam: Convolutional block attention module," in *Proceedings of the European conference on computer vision (ECCV)*, 2018, pp. 3–19.

[39] G. Bradski, "The OpenCV Library," *Dr. Dobb's Journal of Software Tools*, 2000.

[40] A. Greco, G. Valenza, A. Lanata, E. P. Scilingo, and L. Citi, "cvxeda: A convex optimization approach to electrodermal activity processing," *IEEE transactions on biomedical engineering*, vol. 63, no. 4, pp. 797–804, 2015.

[41] D. Cho, J. Ham, J. Oh, J. Park, S. Kim, N.-K. Lee, and B. Lee, "Detection of stress levels from biosignals measured in virtual reality environments using a kernel-based extreme learning machine," *Sensors*, vol. 17, no. 10, p. 2435, 2017.

[42] P. Bobade and M. Vani, "Stress detection with machine learning and deep learning using multimodal physiological data," in *2020 Second International Conference on Inventive Research in Computing Applications (ICIRCA)*. IEEE, 2020, pp. 51–57.

[43] A. Arsalan and M. Majid, "Human stress classification during public speaking using physiological signals," *Computers in biology and medicine*, vol. 133, p. 104377, 2021.

[44] S. D. Subramaniam and B. Dass, "Automated nociceptive pain assessment using physiological signals and a hybrid deep learning network," *IEEE Sensors Journal*, vol. 21, no. 3, pp. 3335–3343, 2020.

[45] V. Kessler, P. Thiam, M. Amirian, and F. Schwenker, "Pain recognition with camera photoplethysmography," in *2017 Seventh International Conference on Image Processing Theory, Tools and Applications (IPTA)*. IEEE, 2017, pp. 1–5.

[46] R. Yang, Z. Guan, Z. Yu, X. Feng, J. Peng, and G. Zhao, "Non-contact pain recognition from video sequences with remote physiological measurements prediction," *arXiv preprint arXiv:2105.08822*, 2021.

[47] F.-T. Sun, C. Kuo, H.-T. Cheng, S. Buthpitiya, P. Collins, and M. Griss, "Activity-aware mental stress detection using physiological sensors," in *Mobile Computing, Applications, and Services*, M. Gris and G. Yang, Eds. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, pp. 211–230.