UNIVERSITY OF CALIFORNIA
Los Angeles

Equitable Thermal Imaging through Single-Shot Correction of Solar Loading

A thesis submitted in partial satisfaction
of the requirements for the degree
Master of Science in Electrical and Computer Engineering

by

Ellin Qing Lan Zhao

2024

ABSTRACT OF THE THESIS

Equitable Thermal Imaging through
Single-Shot Correction of Solar Loading

by

Ellin Qing Lan Zhao
Master of Science in Electrical and Computer Engineering
University of California, Los Angeles, 2024
Professor Achuta Kadambi, Chair

Thermal cameras enable rapid non-contact detection of body temperature and are widely deployed for fever screening, but are inaccurate in unconstrained environments. Previous works have studied the impact of ambient temperature on thermal measurements, and proposed physics-agnostic correction of environmental effects. No previous studies have considered *solar loading*, the increase in skin temperature due to solar radiation. Solar loading results in spurious fever detection and is skin tone dependent, introducing inequity in non-contact fever detection. We propose a method to improve fever detection by removing the solar loading effect from thermal images of the face. We correct solar loading using only one frame of data using a physics-informed neural network that leverages the skin temperature forward model. Our model reduces solar loading mean absolute error (MAE) by 70.5% and achieves 100% specificity in fever detection. We ensure our model is robust by collecting a diverse dataset of 100 subjects with thermal and RGB images and skin tone measurements. Our works shows that it is possible to correct complex thermal perturbations to enable robust and equitable human thermography.

The thesis of Ellin Qing Lan Zhao is approved.

Jonathan Chau-Yan Kao

Aydodan Ozcan

Achuta Kadambi, Committee Chair

University of California, Los Angeles

2024

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGMENTS

This thesis revises the publication listed below. I thank the co-authors of these works (Pradyumna Chari, Sasha Vilesov, Laleh Jalilian, Achuta Kadambi).

Ellin. Q. Zhao, Alexander Vilesov, Shreeram Athreya, Pradyumna Chari, Laleh Jalilian, and Achuta Kadambi. "Making thermal imaging more equitable and accurate: Resolving solar loading biases", *arXiv preprint*, arXiv:2304.08832, April 2023.

I thank my advisor, Professor Achuta Kadambi, and the members of the the Visual Machines Group at UCLA for their support and helpful discussions. Finally, I thank my undergraduate mentees for their hard work in collecting an extensive dataset over 3 years.

CHAPTER 1

Introduction

Infrared sensing enables rapid non-contact measurement of body temperature, an important vital sign that signals the presence of infections and illnesses. In recent years, infrared thermometers (IRTs) have been widely used in non-medical settings to curtail the spread of diseases. In a typical scenario, a person is screened for fever upon entering a building. The person is measured using an IRT, and the measured temperature determines how the person is triaged. IRTs enable remote health sensing by preventing disease transmission through contact-based measurement devices, and are important tools for protecting public health.

For all its benefits, infrared thermography comes with a major limitation—it is accurate only when a person is at equilibrium with room temperature conditions. We call this equilibrium the *baseline* state. Infrared thermometers cannot measure body temperature directly; instead, they measure skin temperature and convert the value to an estimated body temperature. The conversion is derived from data taken from subjects in a baseline state, and as such, thermometers are only accurate when operated in these conditions [17]. Unfortunately, infrared thermometer usage diverges from this required condition: people are typically measured upon entering a building after they have been exposed to variable outdoor conditions, such as hot air or sunlight.

Given the widespread use of IRTs, it is critical to understand their limitations and resolve device inaccuracies. Previous works have shown that IRTs perform poorly after exposure to hot or cold air temperatures [16, 44, 36]. Thus far, no works have studied how sunlight affects thermometer measurements despite the prevalence of this issue. In our study, we found that

sunlight exposure, which we call *solar loading*, causes IRTs to overestimate temperature by 2-3°C, leading to false positive fevers and poor precision as shown in Figure 6.2. Solar loading error is coupled with skin tone because darker skin absorbs more solar radiation and heats up faster, resulting in worse fever precision for people with darker skin. The current solution to solar loading is to wait for up to 30 minutes and then retake the temperature, which is not practical [12, 2, 51, 41].

We present the first study on solar loading in the context of infrared thermography. We collect an extensive dataset with 100 subjects to study solar loading. We record thermal data using multiple IRTs and a thermal camera, and record objective skin tone measurements using a colormeter. Our dataset is diverse in skin tones, and contains over 140,000 paired RGB-thermal images and 900 thermometer measurements. The dataset will be publicly available and may be helpful for studying thermal methods in unconstrained environments [21]. We present a single-shot method of baseline temperature reconstruction that reduces solar loading error by 70.5%. We use a physics-informed neural network (PINN) to reconstruct physiologically accurate baseline images from a single solar loaded image. The model is constrained to find solutions that are both faithful to the ground truth and are stationary—that is, images that abide by the physical meaning of a baseline image.

We first review related works on infrared thermography and bias in medical devices. We then introduce the bioheat equation, which describes skin temperature as a system of partial differential equations (PDEs). We show the difficulties in using the PDEs to constrain a neural network and derive a stationary constraint that can easily be integrated into the neural network optimization. Since there are no previous works for comparison, we additionally present a temporal solution to solar loading against which we compare our single-shot method. Finally, using our collected dataset we show that (1) solar loading exhibits a skin tone bias and (2) our physics-informed model can recover baseline temperature accurately and equitably.

CHAPTER 2

Related Works

We aim to improve human infrared thermography by studying skin tone bias and correcting solar loading errors in a physically-constrained manner. We first review studies on infrared thermography and skin tone biases in remote health sensing. We then discuss general efforts to incorporate physics into machine learning models and specific applications to health sensing.

2.1 Infrared Thermography

Infrared thermometers detect fevers by measuring skin temperature increase during febrile states. Due to the rise in IRT usage during the COVID-19 pandemic, questions have been raised about the accuracy and biases of the devices. [1] found that age and gender affect IRT accuracy. Studies on skin tone bias have conflicting conclusions. [27] compared IRT results across human-annotated skin tone categories: light, medium and dark. No skin tone bias was found, although the dataset was heavily skewed towards light skin. [45] compared the performance of IRTs across measurement locations and ethnic groups (Black, White and mixed). Again, no bias was found. However, [4] showed that IRTs, specifically temporal artery thermometers, underestimate temperature in Black patients. Participants were grouped by self-reported ethnicity (Black and White). The authors hypothesized that this bias is caused by different thermal emissivities in darker skin, but [9] show that skin pigmentation does not affect emissivity. These conflicting studies show that race, ethnicity and subjective skin tone labeling preclude an understanding of bias in IRTs.

Lack of Solar Loading Studies IRTs measure hot areas of the face to be robust to environmental effects, but they are still are inaccurate when the ambient temperature deviates from room temperature [44, 16, 15, 32, 37, 47]. Recent works improve IRT operation in different environmental conditions by fitting relevant temperature variables to a linear function [42, 36]. To the best of our knowledge, these are no prior works that study or correct solar loading error in IRTs.

2.2 Bias in Health Sensing

Seminal work by [7] revealed racial and gender disparities in computer vision algorithms. Unfortunately, biases arise in other technologies, such as optical health sensing devices. Optical biases arise when the light transport is inadequately characterized for certain groups, such as people with dark skin [26]. For example, pulse oximeters and remote photo-plethysmography, which measure blood oxygenation and heart rate respectively, are known to perform worse for darker skin tones due to poor signal-to-noise ratio [31, 43]. Biases can be resolved computationally by balancing skin tones in a dataset with synthetic data, or physically by using multi-modal data [50, 49].

2.3 Physics-Informed Methods

The skin temperature forward model has been studied extensively, providing a set of constraints for correcting solar loading. We review methods that incorporate physics into neural networks. In a standard physics-informed neural network (PINN), the model takes in a coordinate (x, t) and outputs the field value at that location. The governing equations and boundary conditions are used regularizers during training to constrain the model outputs [11, 8]. The same idea can be applied physical models that are not PDEs. [3] use a deep learning model to decompose images into blood and melanin parameters by incorporating

biophysical priors. Beyond machine learning, the finite element method can also be used for physics-guided inversion [13].

We discuss specific applications of physics-informed methods. [18] and [23] use the bioheat model to recover physiological parameters, such as blood flow, vessel location and tumor size. Beyond medical applications, the bioheat equation has been used in computer vision to improve facial recognition. [52] use blood perfusion maps obtained by approximately inverting the bioheat equation as features for facial recognition. Similarly, [5] and [40] extract vessels from thermal images using anisotropic filtering and morphological operations to perform facial recognition. These methods are less explicit in using the bioheat equation; they simply reference heat diffusion from vessels to justify why they equate edges in thermal images as blood vessels. Beyond the heat equation, [29] estimate blood perfusion using attention masks on areas of face with strong pulse signals.

CHAPTER 3

Background

To resolve solar loading biases in thermography, we must understand the operation of thermography devices and the appearance of skin temperature during baseline and solar loaded states.

3.1 Camera-Based Thermography

Infrared thermometers measure skin temperature using either a point sensor or camera, and convert the skin temperature to body temperature using an empirically determined function. [53] provide examples of body temperature functions for different measurement locations. IRTs commonly measure skin at the forehead or medial canthus. These locations are highly perfused by arterial blood resulting in a skin temperature that is highly correlated with body temperature. Point sensors record at a distance of 2-3 cm from the forehead, while camera-based systems record the entire face at a distance of up to 3 meters [17]. Fevers can be detected from further distances by compensating for temperature falloff [10]. Although camera-based IRTs capture the entire face, only the pixels at the canthus or forehead values are used [38]. The IEC standard for IRT accuracy is ± 0.5°C for commercial devices and ± 0.3°C for clinical devices [24].

3.2 Skin Temperature

Skin temperature is affected by many factors, and the two major factors are blood circulation and environmental conditions. Blood, at temperature T_c, circulates from the body's core and perfuses tissue and skin [22]. Blood perfusion, varies spatially causing skin near major vessels to be hotter.

Other heat sources affecting skin temperature are metabolic heat generation, m, and external heat, q_{ext}, such as solar radiation. To summarize these effects, we define tissue temperature over a 3D domain as $u(\mathbf{x}, t)$ where $\mathbf{x} = [x, y, z]^\top$. Tissue temperature is described by Penne's bioheat equation:

$$\rho c \frac{\partial u}{\partial t} = k\nabla^2 u + b(\mathbf{x})(T_c - u) + m + q_{\text{ext}}(\mathbf{x}), \tag{3.1}$$

where ρ, c, k are skin density, heat capacity and thermal conductivity. Blood flow is defined as $b = \rho_b c_b v_b$, where ρ_b, c_b, v_b are blood density, specific heat and flow rate. Notation is summarized in Table 3.1. The external heating, q_{ext}, will be discussed in detail later. At $u(z = 0)$, flux from the environment—air radiation and convection—must equal flux from the skin, yielding the following Robin boundary condition:

$$k \frac{\partial u}{\partial z}\bigg|_{z=0} = h\,(u(z{=}0) - T_e), \tag{3.2}$$

where h is the combined radiative and convective coefficient and T_e is the air temperature.

Equation (3.1) describes temperature under steady state and transient conditions. During steady state, temperature does not change over time ($\frac{\partial u}{\partial t} = 0$) and there is no external heating. Skin temperature can be transient during both heating and cooling.

3.3 Skin Color

Skin heating, in certain ranges of the electromagnetic spectrum, is skin tone dependent. We describe how skin tone affects heating and how we characterize skin tone for our bias

analysis. When radiation impinges on the skin, a portion of the radiation is absorbed and converted to thermal energy. The amount of absorbed energy depends on the radiation spectrum and the skin's absorption coefficient, μ_a. The sun emits nearly half of its energy as visible light, which is absorbed by melanin in the skin. Melanin concentration varies across the population, causing some skin heats up more than others. The skin absorption coefficient is a linear combination of melanin absorption, $\mu_{a,\text{mel}}$ and baseline epidermal skin absorption, $\mu_{a,\text{epi}}$:

$$\mu_a = f_{\text{mel}}\mu_{a,\text{mel}} + (1 - f_{\text{mel}})\mu_{a,\text{epi}}, \tag{3.3}$$

where f_{mel} is the volume fraction of melanin. [25]. Later, we will see how μ_a affects skin heating.

3.3.1 Skin tone measurement

To assess bias in infrared thermometry, we require a means of measuring skin color. Previous studies on bias have used ethnicity or the Fitzpatrick skin type (FST) scale to characterize skin color [31, 45, 4]. The Fitzpatrick scale was developed for phototherapy and divides skin into 6 categories, which are skewed towards light skin, and there is no objective method for determining FST [33]. Labeling skin color using ethnicity or FST does not isolate the primary source of skin color: melanin. Skin tone can be objectively measured using a skin colormeter (DSM III Colormeter, Cortex Technology), which measures the amount of light reflected from the skin to determine melanin concentration. The melanin value returned by the colormeter is not necessarily equal to μ_a.

3.4 Problem Statement

Our aim is to recover the baseline skin temperature after solar loading using only one thermal image. The forehead temperature can be extracted from the reconstructed baseline and used for downstream tasks, such as fever detection. The problem of recovering baseline

temperature is highly under-determined because skin temperature is a 4D function of space and time, but we only observe the 2D skin surface at a single time. In the following chapter, we describe our approach for recovering baseline temperature.

Notation	Meaning	Units
$\mathbf{U}$	skin temperature image	$^\circ$C
$\mathbf{U}_0$	baseline skin temperature image	$^\circ$C
$\hat{\mathbf{U}}_0$	estimated baseline skin temperature image	$^\circ$C
$\mathbf{S}$	solar loading increase image	$^\circ$C
$\mathbf{B}$	blood flow map	$\mathrm{W\,m^{-3}\,^\circ C^{-1}}$
$\mathbf{R}$	residual image	$^\circ$C
$\mathbf{I}$	RGB image	-
u	tissue temp	$^\circ$C
u_0	baseline tissue temperature	$^\circ$C
s	solar loading increase	$^\circ$C
T_c	core temperature	$^\circ$C
T_e	ambient temperature	$^\circ$C
ρ, ρ_b	density (skin, blood)	$\mathrm{kg\,m^{-3}}$
c, c_b	heat capacity (skin, blood)	$\mathrm{J\,kg^{-1}\,^\circ C^{-1}}$
v_b	blood perfusion rate	$\mathrm{s^{-1}}$
b	blood flow term	$\mathrm{W\,m^{-3}\,^\circ C^{-1}}$
k	thermal conductivity	$\mathrm{W\,m^{-1}\,^\circ C^{-1}}$
h	convective heat transfer coefficient	$\mathrm{W\,m^{-2}\,^\circ C^{-1}}$
q_{ext}	external heating	$\mathrm{W\,m^{-2}}$
f_{mel}	volume fraction of melanin	-
μ_a	skin absorption coefficient	$\mathrm{m^{-1}}$
$\mu_{a,\text{mel}}$	melanin absorption coefficient	$\mathrm{m^{-1}}$
$\mu_{a,\text{epi}}$	epidermal absorption coefficient	$\mathrm{m^{-1}}$

Table 3.1: **Summary of mathematical notation for temperature and biophysical parameters.** Left columns shows the symbol and right column describes the notation.

CHAPTER 4

Theory

Given a solar loaded image, we want to recover the underlying baseline image. This is difficult because we are trying to separate temperature transients from an image using only spatial information. Fortunately, baseline skin temperature has a canonical appearance due to physiological factors, which can be captured using a deep learning model. We show the canonical baseline pattern in Figure 4.1; the set of baseline images is low rank compared to the solar loaded images [35]. The optimal reconstructed baseline should match the temperature magnitude of the ground truth to enable accurate fever detection. By definition, the estimated baseline should also be stationary (i.e., not time-varying). Drawing inspiration from recent physics-informed neural networks, we use the bioheat equation to ensure that the model outputs a stationary solution.

Our method uses a frequency-based loss to ensure that the reconstructed baseline is stationary. In the following, we describe the skin temperature forward model and show why it is difficult to constrain a model to learn stationary solutions. Next, we show that viewing the problem from the frequency domain obviates these challenges. We formulate a frequency-based loss function that enables a model to learn stationary solutions, and show that the loss has a physical interpretation related to blood flow.

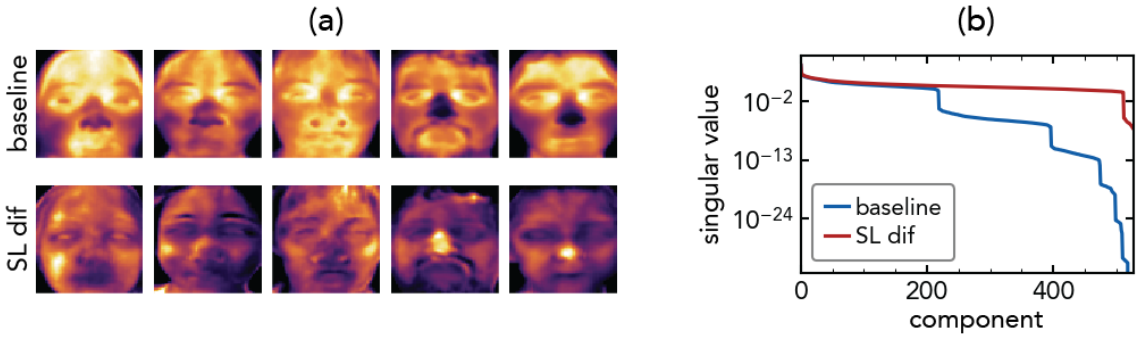

Figure 4.1: **Baseline skin temperature has a canonical appearance due to blood perfusion, but solar loading has a wide range of appearances.** **(a)** Example baseline and solar loaded images. **(b)** Singular values for decompositions of baseline and solar loading increase. Each baseline image is flattened into column and a matrix is created from all the images. The matrix is then decomposed using the singular value decomposition. The same process is repeated for the solar loaded data.

4.1 Solar Loading Pattern

Before we discuss stationary skin temperature, we describe the solar loading effect and its link to skin tone. During solar loading, temperature increases proportional to skin tone and incident radiation. Radiation from the sun with power P_0 reaches the face from direction ℓ. The radiation incident at each (x, y) on the face depends on the surface normal, $\mathbf{n}(x, y)$, and follows the cosine law [46]:

$$P(x, y) = P_0 \, \ell^\top \mathbf{n}(x, y). \tag{4.1}$$

Solar radiation is transmitted through the skin, and is attenuated as it travels deeper into the skin. Transmitted radiation depends on the skin's absorption coefficient, μ_a, and is described by Beer-Lambert law:

$$q_{\text{ext}}(\mathbf{x}) = P(x, y) e^{-\mu_a z}. \tag{4.2}$$

The Beer-Lambert law approximately holds for skin. There is some deviation due to light scattering from blood, but we are interested in melanin, which is contained in the bloodless

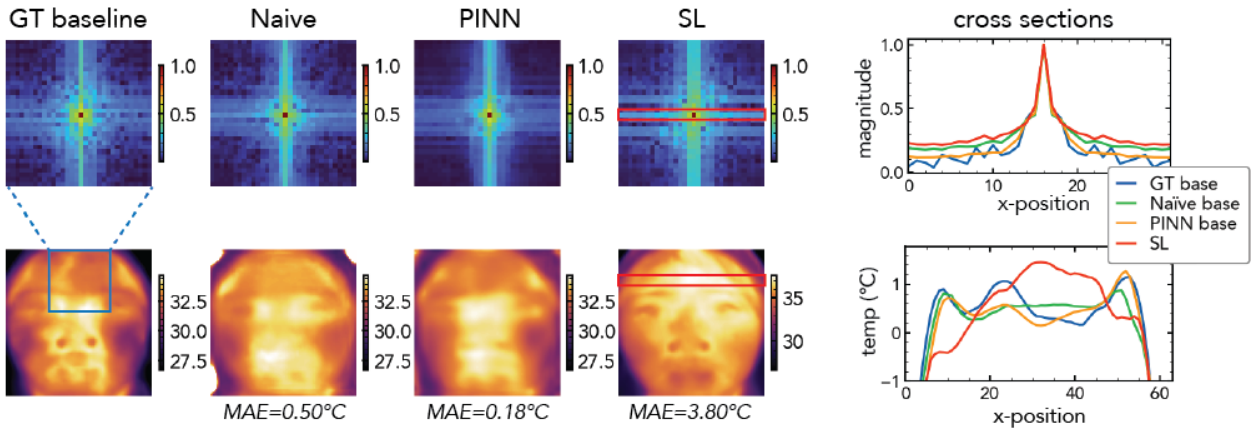

Figure 4.2: **Forehead temperature estimation is improved when the optimization objective includes the spectrum of the thermal images.** The naïve only uses appearance loss and the PINN uses both appearance and frequency loss. The reconstructions appear similar, but the PINN has a lower forehead temperature MAE. The PINN replicates the ground truth spectrum more accurately and captures small temperature variations from blood, which can be seen in the cross sections. This results in a better forehead temperature estimate.

epidermis [34]. Examples of solar loading patterns in Figure 4.1(a). The spatial pattern depends on face geometry, leading to a wide range of possible solar loading patterns, compared to baseline patterns.

4.2 Baseline Skin Temperature

The bioheat equation describes skin temperature under a variety of conditions including the baseline state. During the baseline, skin temperature is stationary: $\frac{\partial u_0}{\partial t} = 0$ as $t \to \infty$. Baseline skin temperature, u_0 satisfies the governing equation,

$$\rho c \frac{\partial u_0}{\partial t} = k\nabla^2 u_0 + b(T_c - u_0) + m. \tag{4.3}$$

One way to ensure an estimated solution is stationary is use the forward Euler method;

that is, advance the estimated solution by a few time-steps and penalize deviations from the initial value. However, there are certain challenges to executing this. Before discussing these challenges, we will modify Equation (4.3) to apply to 2D thermal images. First, we rewrite Equation (4.3) using finite differencing on t, x, y, z:

$$\frac{\partial u_0}{\partial t} \approx \frac{u_0(\mathbf{x}, t+1) - u_0(\mathbf{x}, t)}{\Delta t} \tag{4.4}$$

$$\frac{\partial^2 u_0}{\partial i^2} \approx \frac{u_0(i-1, t) - 2u_0(i, t) + u_0(i+1, t)}{\Delta i^2} \tag{4.5}$$

$$\nabla^2 u_0 \approx \sum_{i \in \{x,y,z\}} \frac{\partial^2 u_0}{\partial i^2}. \tag{4.6}$$

The baseline temperature equation in discretized time steps is:

$$u_0(t+1) = u_0(t) + ak\nabla^2 u_0(t) + ab(T_c - u_0(t)) + am, \tag{4.7}$$

where we have defined $a = \frac{\Delta t}{\rho c}$. Next, we combine the first two terms on the right hand side, $u_0 + ak\nabla^2 u_0$. These terms constitute a blurring of u_0. We can see this by using the spatial finite differencing and rewriting the expression as a convolution. We set $\rho = 1, c = 3.77, k = 0.21$ and assume that $\Delta x = \Delta t = 1$ [19]. The convolution filter has shape 3×3×3, but for brevity we write the filter in only one direction, x:

$$u_0 + ak\nabla^2 u_0 \approx u_0 * \begin{bmatrix} ak & 1-2ak & ak \end{bmatrix} \tag{4.8}$$

$$\approx u_0 * \begin{bmatrix} 0.06 & 0.88 & 0.06 \end{bmatrix} \tag{4.9}$$

$$\approx u_0 * g_x. \tag{4.10}$$

We call the full 3D filter g [1]. Since u_0 is the stationary solution, we have $u_0(t+1) = u_0(t)$. Plugging this into Equation (4.7):

$$u_0 = g * u_0 + ab(T_c - u_0) + am, \tag{4.11}$$

[1] In the z-direction, the filter will be truncated because the skin cannot diffuse heat beyond the skin surface

where we have dropped t. This equation describes a continuous temperature distribution, but we are working with 2D images. We define the i, j element of the following matrices as:

$$\mathbf{U}_0[i,j] = u_0(i\Delta x, j\Delta y, z = 0) \tag{4.12}$$

$$\mathbf{B}[i,j] = b(i\Delta x, j\Delta y, z = 0). \tag{4.13}$$

Let $\mathbf{G}$ be the 2D blur filter g in the x and y directions. Equation (4.11) becomes:

$$\mathbf{U}_0 \approx \mathbf{G} * \mathbf{U}_0 + a\mathbf{B} \odot (T_c - \mathbf{U}_0) + am, \tag{4.14}$$

where $\odot$ is the Hadamard product. We ignore the diffusion in the z-direction, so Equation (4.14) is an approximation.

Challenges For a given $\hat{\mathbf{U}}_0$, we can ensure it is stationary by advancing by a time step using Equation (4.14). However, we do not know $\mathbf{B}, T_c$ or m. Assuming we have paired $(\mathbf{U}, \mathbf{U}_0)$ data, we may invert for $\hat{\mathbf{B}}$ using $\mathbf{U}_0$ and use $\hat{\mathbf{B}}$ to constrain reconstructions from $\mathbf{U}$. T_c and m can be constrained to physically accurate values. This approach requires the images to be well aligned pixel-wise, otherwise we will incur errors in our stationary check. We may block reduce $(\mathbf{U}, \mathbf{U}_0)$ to reduce alignment issues, but blood flow, $\mathbf{B}$, presents as thin edges so the loss in spatial resolution again can lead to errors [40, 5]. For these reasons, it is difficult to apply the stationary constraint in the spatial domain.

4.3 Spectral Stationary Constraint

We can circumvent the difficulties with the stationary constraint by working in the frequency domain. Any estimate of $\mathbf{U}_0$ should satisfy Equation (4.14) in order to be a stationary solution. The same applies after taking the Fourier transform, $\mathcal{F}\{\cdot\}$, of the equation:

$$\mathcal{F}\{\mathbf{U}_0\} = am\,\delta(\omega) + \mathcal{F}\{\mathbf{G} * \mathbf{U}_0\} + a\mathcal{F}\{\mathbf{B} \odot (T_c - \mathbf{U}_0)\}, \tag{4.15}$$

where δ is the Dirac delta. The *spectral* steady state constraint states that the spectrum of $\mathbf{U}_0$ can be decomposed based on frequency bands. The DC term and blurring make up

the low frequency content, so the remaining blood-dependent term must capture the high frequency content in $\mathbf{U}_0$.

The optimal reconstruction should match the appearance of the baseline and be stationary. We next discuss how each frequency band can be used to satisfy the optimization objectives.

Low Frequencies The estimated image is matched to the baseline appearance using the appearance loss, $\mathcal{L}_{\text{app}}$, which is the mean squared error (MSE) between the estimate and ground truth:

$$\mathcal{L}_{\text{app}}(\hat{\mathbf{U}}_0, \mathbf{U}_0) = \left\| \hat{\mathbf{U}}_0 - \mathbf{U}_0 \right\|_F^2 \tag{4.16}$$

It is known that minimizing the MSE results in smooth solutions. That is,

$$\arg\min_{\hat{\mathbf{U}}_0} \mathcal{L}_{\text{app}}(\hat{\mathbf{U}}_0, \mathbf{U}_0) = \arg\min_{\hat{\mathbf{U}}_0} \left\| \mathbf{G} * \hat{\mathbf{U}}_0 - \mathbf{G} * \mathbf{U}_0 \right\|_F^2 \tag{4.17}$$

Minimizing the appearance loss inherently minimizes the difference between the estimated low frequency component in to the ground truth low frequencies in Equation (4.15).

High Frequencies Although the appearance loss minimizes the error between the low frequency components, the stationary constraint may not be satisfied. To ensure the constraint is satisfied, we can add an additional loss on the high frequency features:

$$\mathcal{L}_s(\hat{\mathbf{U}}_0, \mathbf{U}_0) = \left\| \mathcal{F}\{\hat{\mathbf{U}}_0 - \mathbf{G} * \hat{\mathbf{U}}_0\} - \mathcal{F}\{\mathbf{U}_0 - \mathbf{G} * \mathbf{U}_0\} \right\|_F^2. \tag{4.18}$$

This the the stationary loss. When both the appearance and stationary losses are minimized, the estimate satisfies the stationary constraint. To better understand what this loss is optimizing, we define the estimated baseline as $\hat{\mathbf{U}}_0 = \mathbf{U}_0 + \mathbf{R}$, where $\mathbf{R}$ is the residual. The stationary loss becomes:

$$\mathcal{L}_s(\hat{\mathbf{U}}_0, \mathbf{U}_0) = \|\mathcal{F}\{\mathbf{B} \odot \mathbf{R}\}\|_F^2 \tag{4.19}$$

where $\odot$ is a Hadamard product. We move back to the spatial domain. Using Parseval's identity and the definition of mean squared error, we have:

$$\|\mathcal{F}\{\mathbf{B} \odot \mathbf{R}\}\|_F^2 = 2\pi \|\mathbf{B} \odot \mathbf{R}\|_F^2 \tag{4.20}$$

$$= 2\pi \sum_{i=0}^{n} \sum_{j=0}^{n} b_{ij}^2 r_{ij}^2. \tag{4.21}$$

Since $\mathbf{B}$ is fixed, a trivial solution that minimizes the summation is $\mathbf{R} = 0$. Assume that $\mathbf{R}$ is not zero. The summation is minimized when r_{ij} is low where b_{ij} is high and vice versa, forcing the residual to be uncorrelated with blood flow. This is important because the parts of the face that are used for fever screening are highly perfused areas, such as the forehead.

Full Physics-informed Loss Function One problem with the frequency loss is that it is applied to the entire image. Some parts of the face may have more blood flow and high frequency content, but there is no notion of where the high frequency content belongs spatially. To alleviate this, we can apply the loss to patches of the image:

$$\mathcal{L}_{s,\text{tot}} = \sum_{i=0}^{n_p} \mathcal{L}_s(p_i(\hat{\mathbf{U}}_0), p_i(\mathbf{U}_0)) \tag{4.22}$$

where n_p is the number of patches and $p_i(\cdot)$ extracts the i-th patch. The full physics-informed loss is:

$$\mathcal{L}(\hat{\mathbf{U}}_0, \mathbf{U}_0) = \mathcal{L}_{\text{app}}(\hat{\mathbf{U}}_0, \mathbf{U}_0) + \gamma \mathcal{L}_{s,\text{tot}}(\hat{\mathbf{U}}_0, \mathbf{U}_0), \tag{4.23}$$

where γ is a weight for the second loss term.

Summary From the bioheat equation, we obtained a stationary constraint on baseline solutions. We could not directly apply the constraint in the spatial domain because the blood flow is unknown. We showed that the stationary constraint can be enforced using the high frequency content of the images. We need to select width of the blur kernel so that the constraint is stable, but this is far easy than dealing with the unknown blood map in the spatial domain. We show how the frequency loss improves reconstructed temperature in Figure 4.2.

4.4 Elevated Body Temperature

For completeness, we describe skin temperature for febrile states, although we do not explicitly use this in our model. During a fever, body temperature is elevated from a healthy temperature causing the skin temperature to increase. A healthy body temperature is approximately 37°C and a fever is 38°C or higher. We define the difference between the febrile and healthy body temperature as ΔT_c, and difference for skin temperature as Δu. The spatial pattern of the skin temperature increase can be described by Equation (3.1):

$$0 = k\nabla^2(\Delta u) + b(\mathbf{x})(\Delta u - \Delta T_c). \tag{4.24}$$

We observe that the spatial pattern of blood drives the change in febrile skin temperature. This is different from solar loading where skin temperature changes based on face geometry and sun location.

CHAPTER 5

Implementation

To evaluate solar loading, we collect a new dataset of face temperatures in different conditions. An overview of our dataset is given in Table 5.1 We describe our dataset and our neural network implementation details.

5.1 Dataset

Our dataset contains thermal data (camera, thermometers) for baseline and solar loaded states. Each subject's skin color is recorded as melanin index, using the colormeter described in Section 3.3.1, and through manually annotated Fitzpatrick skin type. The dataset contains 100 subjects of diverse skin tones to allow us to evaluate bias. The skin tone distribution is shown in Figure 5.1. While we use this dataset specifically for solar loading, it can be used to analyze thermal methods under transient conditions, such as thermal face recognition. In the following, we describe the data collection protocol and how we generate a dataset for our supervised learning model.

5.1.1 Experimental Protocol

One purpose of the dataset is to evaluate solar loading increase over different skin tones. The solar loading effect depends not only on skin tone, but also on environmental conditions such as wind, solar angle, and cloudiness. To control for these covariates, we only collect data under the following conditions:

- Solar zenith close to 0 (12:00-15:00)
- Minimal to no clouds (solar power > 900 Wm^{-2})
- Minimal to no wind (wind speed < 3 ms^{-1})

Collecting data in a limited time of day also controls for diurnal variations, so that our equity analysis is not biased by body temperature fluctuations.

Data collection is split into three sessions: baseline, solar loading, cooling. The baseline and cooling sessions are completed indoors, and the solar loading session is completed outdoors. During the baseline portion, the subject is in a thermoneutral state. Body temperature does not change over the course of data collection, so we record it once using an oral thermometer. We record skin tone using a colormeter, and Fitzpatrick skin type is determined by the average value from 5 annotators performed on white-balanced RGB images taken during the baseline. To provide a point of comparison for solar loaded data, we record the subject's face temperature for 1 minute using a thermal camera.

Next, the subject is solar loaded outside for 5 minutes. The solar power is recorded to remove solar strength as a covariate for solar loading. Afterwards, the subject returns indoors and their face temperature is recorded for 5 minutes while they cool down. This mimics the scenario of a person being screened when they enter a building. The cooling data is what we use to train and evaluate our model. After each of the three sessions, we collect forehead measurements using 3 infrared thermometers. We additionally measure indoor air temperature and relative humidity using an anemometer.

5.1.2 Hardware Setup

We collect data using commercially available infrared thermometers and a custom camera setup. We show the devices used in Figure 5.2. We use three common forehead infrared thermometers (Welch Allyn 105801, ADC Adtemp 429, Joytech Sejoy DET-306). Ground truth

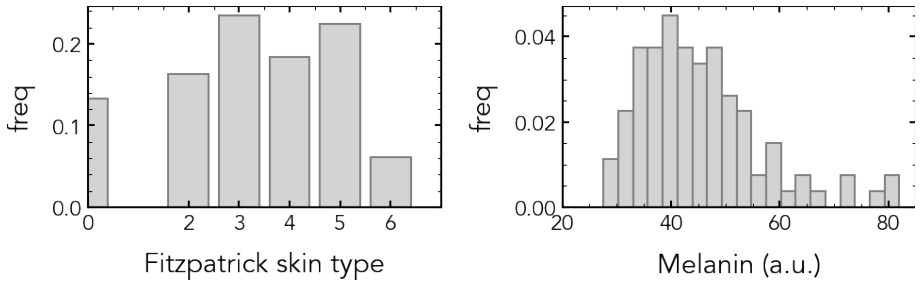

Figure 5.1: **Our dataset captures an diverse range of skin tones.** Skin tone is recorded as Fitzpatrick skin type (FST) and as a melanin index.

Type	**Value**	**Notes**
Skin tone	Melanin, erythema	-
	Fitzpatrick skin type	-
Body temperature	Infrared thermometers	6 readings
	Oral thermometer	-
Images	Thermal camera	1440 frames
	RGB camera	1440 frames
Indoor conditions	Solar power	-
	Air temperature	-
	Relative humidity	-

Table 5.1: **For each subject, we record skin tone labels, thermal data and the environmental conditions.** Our dataset contains 100 subjects, yielding over 140,000 paired RGB-thermal images in our dataset.

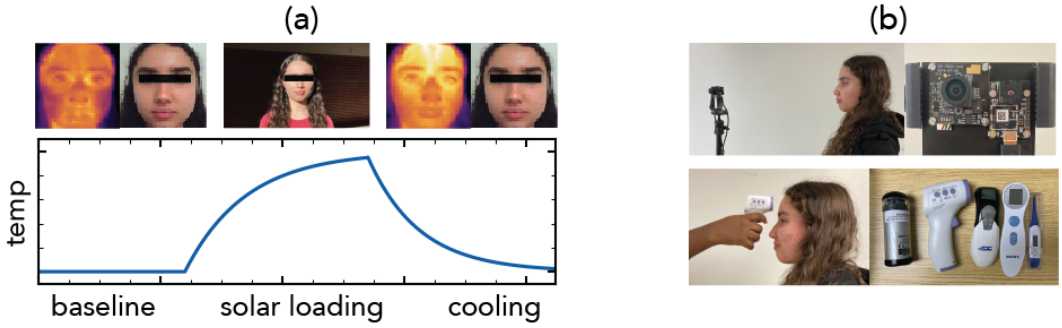

Figure 5.2: **Our dataset contains thermal data captured over multiple temperature states: baseline, solar heating and cooling. (a)** During the baseline and cooling sessions, we capture paired RGB and thermal images, as shown in the top row. **(b)** Images are recorded using a synchronized RGB-thermal camera system, shown in the top row. We record values from commercial instruments shown in the bottom row: colormeter, 3 infrared thermometers and 1 oral thermometer.

body temperature is recorded using an oral thermometer (Boncare Digital Thermometer MT-601A). Indoor conditions are recorded using a handheld anemometer (BTMeter Anemometer 866A) and solar radiation strength is recorded using a solar power meter (Tenmars Solar Meter TM-206).

We record images using a custom RGB-thermal camera setup. A thermal camera (FLIR Lepton 3.5 LWIR) is mounted next to a RGB camera (Arducam) on a 3D printed component. Images are taken at a sampling rate of 4 Hz. The RGB images are used to compute facial landmarks that are used for data preprocessing. The RGB and thermal images are not aligned pixel-wise, so we calculate the homography needed to register the images. We image a cold metal sheet placed behind a cardboard cutout checkerboard pattern, to create distinct thermal and RGB image features. We manually match the key points using MATLAB's image registration procedure. All RGB-derived data is warped to be aligned with the IR images. In lieu of a black-body reference, the wall in the background of the images is used as a pseudo-reference to help correct for camera flat-field correction effects. The camera setup

and example images are shown in Figure 5.2.

5.2 Paired Dataset Generation

The goal of solar loading correction is to take an input image, $\mathbf{U}$, and output the baseline state, $\mathbf{U}_0$. To realize this using a supervised approach, we need paired images $(\mathbf{U}, \mathbf{U}_0)$ for training. The paired imaged should be aligned to enable a model to learn temperature differences and not facial expression or head pose differences. We describe our approach for generating this dataset.

5.2.1 Data Cleaning

We filter out frames where the subject is not facing forwards to allow us to extract the forehead temperature, which is not affected by directional emissivity. All frames are cropped to the facial region and resized to 64×64. We use `dlib` to find the facial landmarks from the RGB images [28]. We check if the landmarks are symmetric; if there are not, the subject is not facing forwards and the frames are discarded. The thermal camera periodically performs a flat-field correction (FFC), during which the output frames are unusable. The FFC results in temperature spikes, which we can detect using an object of known temperature in the scene. In our case, the background temperature is known to be steady. We extract the background temperature over all frames and use `scipy.find_peaks` to detect peaks that match the calibration periodicity [14]. Data occurring during the FFC is not used.

5.2.2 Finding Optimal Pairs

Finding the best image pair using only thermal data is difficult because the temperature distribution changes after solar loading. Instead, we can find the best matches using the facial landmarks derived from RGB images. Assume $\mathbf{v}_i$ is a vector of landmarks derived

from the i-th RGB image. For the baseline session we have $\{(\mathbf{U}_0^i, \mathbf{v}_0^i)\}_{i=1}^{N_1}$ and for the cooling session we have $\{(\mathbf{U}^i, \mathbf{v}^i)\}_{i=1}^{N_2}$. For a given solar loaded image, $\mathbf{U}^i$, we find the best aligned $\mathbf{U}_0^*$ by:

$$\mathbf{U}_0^* = \min_j \left\| \mathbf{v}^i - \mathbf{v}_0^j \right\|_2 \tag{5.1}$$

5.2.3 Training and Test Sets

We split our dataset into training and test sets, and the training set is further divided for 8-fold cross validation. The training set contains 80 subjects, with a total of 18,390 paired frames for training, and the test set contains 20 subjects, with 5,017 paired frames. Each subject appears in exclusively in either the training or test set to prevent overfitting. We only use the first 2 minutes of cooling data because past this time, the amount of solar loading is low and it is ambiguous whether there is solar loading. We apply random horizontal flips to the images as data augmentation.

5.3 Model Details

5.3.1 Base Model Architecture

Our model is based on the UNet architecture, which is extensively used for image-to-image tasks such as denoising and segmentation [39, 20]. The encoder of the UNet consists of 4 downsampling layers, where each downsampling layer contains 2 convolutional blocks (convolution, batch normalization, ELU), followed by max pooling. The number of feature channels is 32, 64, 128, 128. The decoder follows a symmetric architecture, with the maxpooling replaced by upsampling. We apply a scaled sigmoid on the final model output to scale the output into a known range of temperature values ($26.7 \leq u \leq 43°C$). We use a batch size of 32 and train using the Adam optimizer with learning rate of 1e-4 and weight decay of 1e-4.

We train for 15 epochs. For the naïve model, the loss function is the appearance loss (MSE between the estimated and ground truth face images).

5.3.2 Physics-informed Network

For the physics-informed network, we use the stationary loss in addition to the appearance loss. For each estimate, we compute the 2D FFT of the image and filter out low frequencies using a Gaussian blur kernel with width of 5 pixels and a standard deviation of 5. The MSE between the filtered ground truth and estimate spectra constitutes the physics-informed loss. Our final model calculates the FFT on patches of size 32×32.

5.4 Evaluating Performance

Our work is the first to tackle the problem of solar loading and as such, there are no existing methods against which we can compare our results. We instead report solar loading errors on uncorrected data, and the results of our temporal and spatial solutions. We report the mean squared error (MSE), root mean squared error (RMSE) and mean absolute error (MAE) of our models. We compare the MAE to the fever threshold. Additionally, we report the mean absolute percentage error (MAPE), which is defined as:

$$\text{MAPE}(y, \hat{y}) = \frac{1}{n} \sum_{i=0}^{n} \left| \frac{y_i - \hat{y}_i}{y_i} \right|. \tag{5.2}$$

All errors are calculated using the forehead temperature. The goal of correcting solar loading is to reduce false positive fevers. To evaluate practical or clinical relevancy of our results, we report the specificity of our model:

$$\text{Specificity} = \frac{TN}{TN + FP}, \tag{5.3}$$

where TN and FP are the number of true negatives and false positives respectively. The fever threshold for forehead temperatures in prior works is between 35.1-35.6°C [30, 17], so

we use a fever threshold of 35.35°C. Since we have no fever datapoints, we do not report the sensitivity or precision values.

CHAPTER 6

Results

Method	Forehead (°C)					
	MSE (↓)	RMSE (↓)	MAE (↓)	MAPE (↓)	Specificity (↑)	Pearson R (↑)
No correction	4.946	2.224	2.048	0.062	0.451	0.607
Ours (10 sec temporal fit)	24.915	4.991	4.178	0.120	0.548	0.170
Ours (5 min temporal fit)	2.718	1.649	0.764	0.022	0.935	0.737
Ours (basic CNN)	0.602	0.776	0.650	0.020	1.000	0.659
Ours (PINN)	0.523	0.723	0.605	0.018	1.000	0.720

Table 6.1: **Our single-shot model achieves the best performance over all the error metrics, and shows high correlation with the ground truth.** The temporal correction model performs well with 5 minutes of data, but fails when given a duration that is more reasonable for fever screening (10 seconds).

We present the results of our physics-informed solar loading correction model. We emphasize that there is no existing method to compare our results to. To provide a better assessment of our single-shot correction, we include results from a temporal correction model.

6.1 Temporal Solution

Solar loading can be corrected with time series data by finding the steady state of the temperature decay.

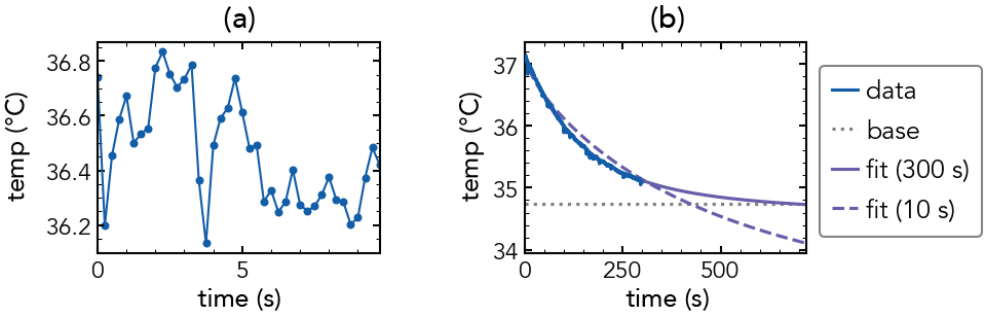

Figure 6.1: **Solar loading can be detected by fitting time series data to an exponential function. (a)** Variation in forehead temperature over 10 seconds. **(b)** Average temporal fits over all subjects. Fitting using 300 seconds of data yields accurate results, but using 10 seconds does not.

We extract temperature at the forehead while skin is cooling after solar loading, and fit a decaying exponential to the data. Specifically, we find $A > 0, B > 0$ and $\hat{u}_0$ such that:

$$u(t) = Ae^{-Bt} + \hat{u}_0. \tag{6.1}$$

We constrain the steady state to a valid range of temperatures, $32 \leq \hat{u}_0 \leq 37°C$. We find the best exponential fit to the data using `scipy .curve_fit`. The curve fitting using 10 seconds of data and 5 minutes of data is shown in Figure 6.1(b). Solar loading can be removed (MAE=0.764°C) using 5 minutes of data, but that requires 5 minutes of continuous measurement. Using a more reasonable time frame, such as 10 seconds, results in poor performance (MAE=4.991°C). This is because the noise in the temperature over 10 seconds obscures the cooling trend. For reference, the thermal camera used in this study has a NETD of 0.05°C [48]. Figure 6.1 shows the variation in temperature over 10 seconds and fitted curves averaged over the entire dataset.

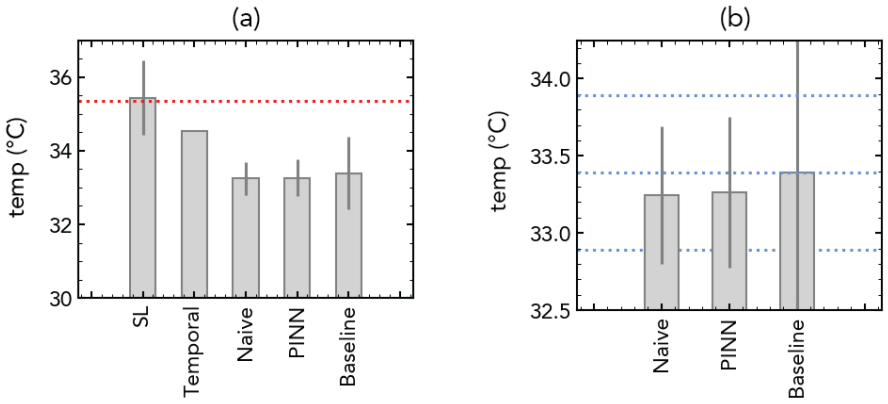

Figure 6.2: **Our spatial correction models accurately estimate baseline temperature enabling better fever detection.** **(a)** The estimated baseline temperatures are shown against the fever threshold (red line). Error bars show ±1 SD. No error bar is shown for the temporal method because the standard deviation is greater than the limits of the plot. **(b)** The results of our model are within ±0.5°C (blue lines) of the ground truth.

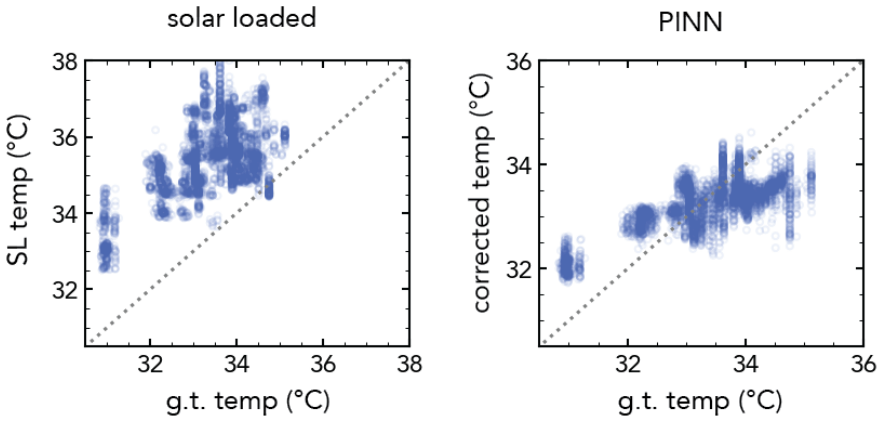

Figure 6.3: **Our baseline estimates improve the correlation correlation ($R = 0.723$) compared to the uncorrected solar loading ($R = 0.607$).** Although the solar loading correlation is high, the absolute error is also high.

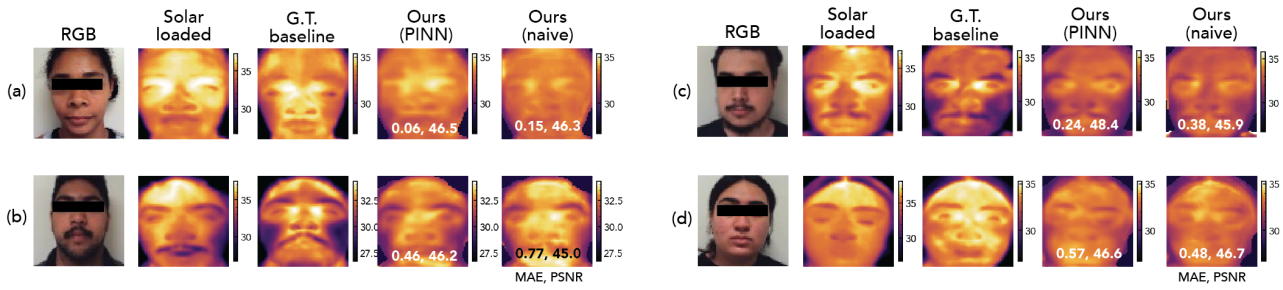

Figure 6.4: **Baseline reconstructions from our single-shot methods achieve both low temperature errors and high PSNR.** We show the model input (solar loaded image) and outputs (PINN and naïve reconstructions) for 4 subjects in the test set. Overall, the PINN achieves lower temperature error and higher PSNR than the naïve model.

6.2 Single-Shot Solution

We present results from our single-shot solar loading correction models. Both the naïve and physics-informed model significantly reduce the solar loading error and produce accurate reconstructions of facial temperature.

6.2.1 Forehead Temperature Estimation

Solar loading errors in temperature estimation typically occur when a person is being screened for fever. As such, we report the accuracy of forehead temperature estimation and fever detection. The results are summarized in Table 6.1. The PINN has the best performance across all but one metric, the Pearson correlation coefficient. Our temporal model has the highest correlation ($R = 0.737$) and the PINN has the next best correlation ($R = 0.720$). We show the correlation of the solar loaded and PINN corrected temperatures with the ground truth in Figure 6.3.

Both of the single-shot methods improve the solar loading error, and the PINN outperforms the basic model across all metrics. The PINN reduces the solar loading error by 70.5% and and the naive model reduces the error by 64.9%. The correction enables more

accurate fever detection. Without correction, only 45% of solar loaded data was classified as non-febrile, but our model correction results in 100% of subjects being correctly classified as non-febrile. In Figure 6.2, we show the corrected temperatures against the fever threshold of 35.35°C.

6.2.2 Reconstruction Quality

In addition to accurate forehead temperature estimation, our model produces reconstructions of the entire face that are faithful to the ground truth. Our aim is not to optimize image quality, but assessing the quality of the full face reconstructions shows that the model is not simply overfitting to the forehead temperature by removing a constant temperature offset. The reconstructions from our single-shot models are shown in Figure 6.4 along with the forehead MAE and image PSNR values. PINN outptus have lower forehead MAE and higher PSNR values than the naive model, in all but one of the examples. For Figure 6.4(d), the PINN has a higher MAE but it is only 0.09°C greater than the naive model error. We see that PSNR is not indicative of the forehead error: fig. 6.4(a) and (b) have similar PSNR values, but a large difference in temperature error. Figure 6.4(b) shows a clear difference in the reconstructions between the PINN and naive models. The naive model keeps heat centered on the forehead, similar to the input solar loaded image. The PINN, however, directs heat towards the edges of the forehead where the temporal arteries are located. This results in a more accurate temperature estimation: the PINN has a MAE value of 0.46°C, while the naive model has a MAE of 0.77°C.

6.3 Ablation Study

We have seen that the addition of a physics-informed loss improves the temperature estimation. We provide more results to this effect in an ablation study. The results are summarized in Table 6.2. Reported values are averaged over 8-fold cross validation on the training set.

	Forehead (°C)			
Method	MSE (↓)	RMSE (↓)	MAE (↓)	MAPE (↓)
No corr	5.483	2.335	2.134	0.064
Naïve	0.553	0.735	0.606	0.018
Naïve + fever aug	0.619	0.772	0.605	0.018
PINN (16x16)	0.622	0.764	0.596	0.018
PINN (32x32)	0.517	0.709	0.566	0.017
PINN (32x32) + fever aug	0.631	0.780	0.611	0.018

Table 6.2: **Ablation study on model hyperparameters.** We evaluate the performance of the physics-informed loss with varying patch sizes, and the performance of synthetic fever augmentation. The best metrics are highlighted in green.

6.3.1 Physics-Informed Loss

We compare model performance with and without the physics-informed loss. For the physics-informed loss, we vary the patch size to see how it improves the performance. We test patch sizes of 16×16 and 32×32. In theory, smaller patch sizes should yield better results (at the expense of runtime) because blood-dependent frequency features are better localized. However, we found that this was not the case. Using a 32×32 patch size resulted in the best performance across all metrics. This is likely because slight image misalignment at finer patch scales causes spuriously high loss values. The physics-informed model trained with no fever augmentation outperformed the naïve model's MAE by 0.04°C.

6.3.2 Synthetic Fever Augmentation

We hypothesized that the model may be over reliant on temperature magnitude and may not learn from the full spatial pattern of the face. We added synthetic fever images to the model to prevent overfitting to magnitude. We added a random constant temperature offset

in the range of 2-3°C to skin pixels in the baseline images, as was done in [54]. We found that this did not improve the results for both the naïve and physics-informed models. The fever augmentation degraded the PINN results to a higher extent, with the MAE increasing by 0.05°C. Our method of generating synthetic fever images may not be physically accurate, which results in poorer performance with the fever augmentation.

6.4 Equity Analysis

We described the theoretical basis for solar loading skin tone bias in Section 4.1, and we now show that (1) the bias is captured by our dataset and (2) our model reduces the bias. We show the correlations between solar loading increase and melanin index in Figure 6.5. For skin temperature, we average the forehead temperature increase over the first 20 seconds of cooling, and use this as the solar loading increase. For the IRTs, solar loading increase is the difference between the solar loading and baseline measurements. We use a non-parametric measure of correlation (Spearman correlation coefficient) because the specific relationship between solar loading increase and melanin is unknown.

In Figure 6.5(a), we show that there is a positive correlation between melanin and solar loading bias for both skin temperature and IRT measurements. The correlation is positive, albeit low, but it is a statistically significant result ($p < 0.1$). The IRT measurements show a stronger skin tone bias with ($R = 0.338$) than the skin temperature ($R = 0.133$). This difference may arise from the experimental procedure. The IRT measurements were taken immediately after sun exposure, while the skin temperature measurements were taken after the subject re-entered the building (10-20 seconds). The skin may be slightly cooled by the beginning of the skin temperature measurement, resulting in an apparently lower bias. Our model reduces the skin tone bias: as shown in Figure 6.5(b), the temperature errors are not correlated with melanin.

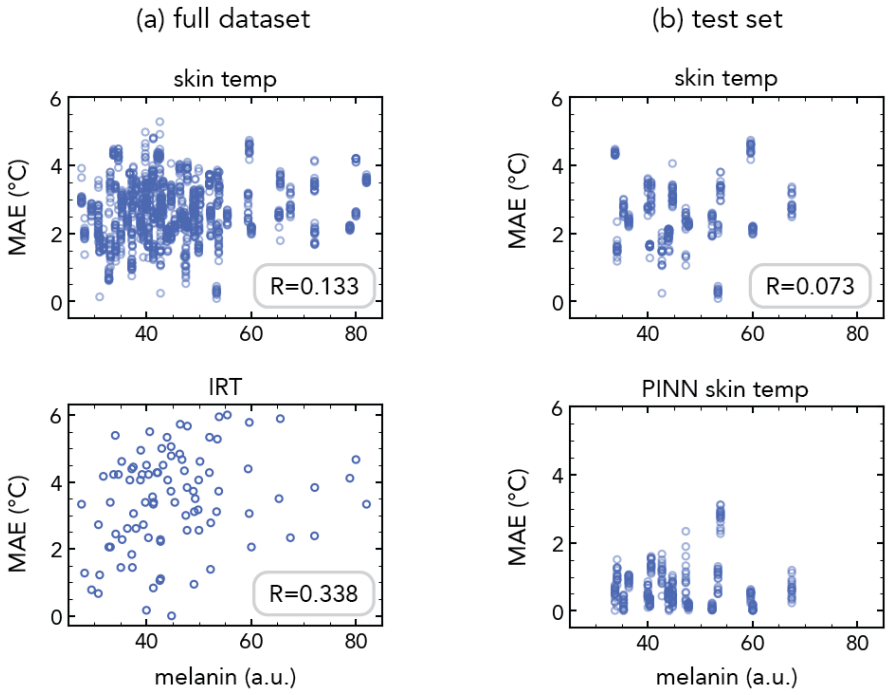

Figure 6.5: **Solar loading increase is correlated with melanin levels for skin and IRT temperature.** Our model corrects solar loading and removes bias. We report the Spearman correlation coefficient (p-value < 0.1).

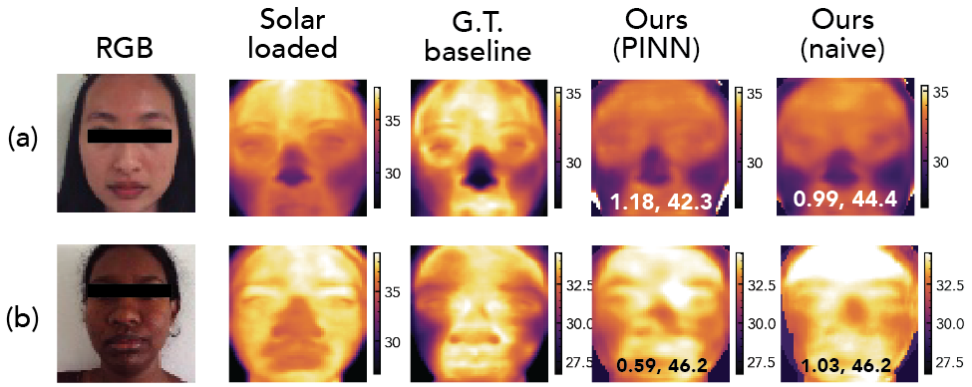

Figure 6.6: **Example reconstructions for cases with high temperature error.** We report the forehead MAE and PSNR. Although the PSNR is high, the temperature error is also high.

CHAPTER 7

Discussion

We have presented the first method to combat solar loading biases in thermography. We have shown that the bioheat equation can be used to constrain a model to learn realistic baseline images. Our model is able to reduce solar loading error by 70%, resulting in better specificity for fever detection. We collected an extensive dataset of thermal images under baseline and solar loaded conditions. The data and code will be made publicly available. As alternate modalities continue to be used in computer vision, this dataset may also be used for studying thermal face recognition and emotion analysis in unconstrained environments.

7.1 Limitations

This work only considers temperature transients induced by solar radiation, but there are other factors that cause skin temperature to deviate from its baseline state, such as air temperature and wind. These are not considered because they require a different physical model, and solar loading is an important enough problem to merit a standalone study. There exist other studies on ambient temperature effects, but none so far on solar loading. Our dataset does not contain any subjects with fevers, as that would require infectious disease controls as part of data collection. Expanding the dataset is a direction for future work.

7.2 Future Work

We only use temperature magnitude as input to our model, but thermal and RGB images provide a wealth of information about human health. Imaging photoplethysmography (iPPG) methods detect heart rate from RGB images, which can indicate exposure to transient conditions [6]. To detect fevers, our model removes transients from skin temperature, but it may be helpful to estimate body temperature directly to eliminate compounding errors in downstream systems. With a higher resolution thermal camera, we may also jointly estimate the blood flow map and the baseline skin temperature. Since this is the first study on solar loading, there is a multitude of directions for follow up works.

7.3 Ethics

We have studied the solar loading problem in a targeted manner by focusing on optical melanin biases, but thermography may be biased in other ways. It is important for future works to study bias along other axes.

7.4 Conclusion

We present a method to improve fever detection in unconstrained environments, namely under solar loading conditions. Our physics-informed model improves solar loading MAE by 70.5%, allowing for equitable fever detection. Our method preserves the rapidness of standard IRT measurement and expands the usability of the devices. Our hardware setup uses an inexpensive thermal camera—costing under $200 USD—and along with our open-source code, this setup is easily adoptable in place of standard IRTs. The benefit of our computational method is that it can be incorporated into existing camera-based thermography systems. Research on bias in imaging and optical devices is important, and this work studies an important application where skin tone bias arises.

REFERENCES

[1] Scott Adams, Tracey Bucknall, and Abbas Kouzani. An initial study on the agreement of body temperatures measured by infrared cameras and oral thermometry. *Scientific Reports*, 11(1), June 2021.

[2] ADC. Adtemp non-contact thermometer, model 429 user manual, jan 2017.

[3] Sarah Alotaibi and William Smith. Biofacenet: Deep biophysical face image interpretation, 2019.

[4] Sivasubramanium V. Bhavani, Zanthia Wiley, Philip A. Verhoef, Craig M. Coopersmith, and Ighovwerha Ofotokun. Racial differences in detection of fever using temporal vs oral temperature measurements in hospitalized patients. *JAMA*, 328(9):885, September 2022.

[5] Pradeep Buddharaju, Ioannis T. Pavlidis, Panagiotis Tsiamyrtzis, and Mike Bazakos. Physiology-based face recognition in the thermal infrared spectrum. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 29(4):613–626, April 2007.

[6] Mark J Buller, William J Tharion, Samuel N Cheuvront, Scott J Montain, Robert W Kenefick, John Castellani, William A Latzka, Warren S Roberts, Mark Richter, Odest Chadwicke Jenkins, and Reed W Hoyt. Estimation of human core temperature from sequential heart rate observations. *Physiological Measurement*, 34(7):781–798, June 2013.

[7] Joy Buolamwini and Timnit Gebru. Gender shades: Intersectional accuracy disparities in commercial gender classification. In Sorelle A. Friedler and Christo Wilson, editors, *Proceedings of the 1st Conference on Fairness, Accountability and Transparency*, volume 81 of *Proceedings of Machine Learning Research*, pages 77–91, New York, US, 23–24 Feb 2018. PMLR.

[8] Shengze Cai, Zhicheng Wang, Sifan Wang, Paris Perdikaris, and George Em Karniadakis. Physics-informed neural networks for heat transfer problems. *Journal of Heat Transfer*, 143(6), April 2021.

[9] Matthew Charlton, Sophie A. Stanley, Zoë Whitman, Victoria Wenn, Timothy J. Coats, Mark Sims, and Jonathan P. Thompson. The effect of constitutive pigmentation on the measured emissivity of human skin. *PLOS ONE*, 15(11):e0241843, November 2020.

[10] Jing Wei Chin, Kwan Long Wong, Tsz Tai Chan, Kristian Suhartono, and Richard H.Y. So. An infrared thermography model enabling remote body temperature screening up to 10 meters. In *2021 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 3870–3876, Virtual, 2021. IEEE.

[11] Salvatore Cuomo, Vincenzo Schiano Di Cola, Fabio Giampaolo, Gianluigi Rozza, Maziar Raissi, and Francesco Piccialli. Scientific machine learning through physics–informed neural networks: Where we are and what's next. *Journal of Scientific Computing*, 92(3), July 2022.

[12] Hein Daanen, Stephan Bose-O'Reilly, Matt Brearley, D. Andreas Flouris, Nicola M. Gerrett, Maud Huynen, Hunter M. Jones, Jason Kai Wei Lee, Nathan Morris, Ian Norton, Lars Nybo, Elspeth Oppermann, Joy Shumake-Guillemot, and Peter Van den Hazel. Covid-19 and thermoregulation-related problems: Practical recommendations. *Temperature*, 8(1):1–11, August 2020.

[13] Aniket Dashpute, Vishwanath Saragadam, Emma Alexander, Florian Willomitzer, Aggelos Katsaggelos, Ashok Veeraraghavan, and Oliver Cossairt. Thermal spread functions (tsf): Physics-guided material classification. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, Vancouver, CAN, June 2023. IEEE.

[14] Pan Du, Warren A. Kibbe, and Simon M. Lin. Improved peak detection in mass spectrum by incorporating continuous wavelet transform-based pattern matching. *Bioinformatics*, 22(17):2059–2065, July 2006.

[15] Cornelius Dzien, Wolfgang Halder, Hannes Winner, and Monika Lechleitner. Covid-19 screening: are forehead temperature measurements during cold outdoor temperatures really helpful? *Wiener klinische Wochenschrift*, 133(7–8):331–335, October 2020.

[16] Ufuk Erenberk, Emel Torun, Emin Ozkaya, Selcuk Uzuner, Aysegul Dogan Demir, and Rusen Dundaroz. Skin temperature measurement using an infrared thermometer on patients who have been exposed to cold. *Pediatrics International*, 55(6):767–770, December 2013.

[17] Josh Foster, Alex Bruce Lloyd, and George Havenith. Non-contact infrared assessment of human body temperature: The journal temperature toolbox. *Temperature*, 8(4):306–319, April 2021.

[18] M Garbey, A Merla, and I Pavlidis. Estimation of blood flow speed and vessel location from thermal video. In *Proceedings of the 2004 IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 2004. CVPR 2004*, Washington, DC, USA, 2004. IEEE.

[19] Raymond G. Gordon, Robert B. Roemer, and Steven M. Horvath. A mathematical model of the human temperature regulatory system - transient cold exposure response. *IEEE Transactions on Biomedical Engineering*, BME-23(6):434–444, November 1976.

[20] Steven Guan, Amir A. Khan, Siddhartha Sikdar, and Parag V. Chitnis. Fully dense unet for 2-d sparse photoacoustic tomography artifact removal. *IEEE Journal of Biomedical and Health Informatics*, 24(2):568–576, February 2020.

[21] Gabriel Hermosilla, Javier Ruiz-del Solar, Rodrigo Verschae, and Mauricio Correa. A comparative study of thermal face recognition methods in unconstrained environments. *Pattern Recognition*, 45(7):2445–2459, July 2012.

[22] Jordan Hristov. Bio-heat models revisited: Concepts, derivations, nondimensalization and fractionalization approaches. *Frontiers in Physics*, 7, November 2019.

[23] J Iljaž, L C Wrobel, T Gomboc, M Hriberšek, and J Marn. Solving inverse bioheat problems of skin tumour identification by dynamic thermography. *Inverse Problems*, 36(3):035002, February 2020.

[24] ISO. Medical electrical equipment part 2-59: Particular requirements for the basic safety and essential performance of screening thermographs for human febrile temperature screening. Standard, International Organization for Standardization, Geneva, CH, September 2017.

[25] Steven L Jacques. Optical properties of biological tissues: a review. *Physics in Medicine and Biology*, 58(11):R37–R61, May 2013.

[26] Achuta Kadambi. Achieving fairness in medical devices. *Science*, 372(6537):30–31, April 2021.

[27] Shahrukh Khan, Bridey Saultry, Scott Adams, Abbas Z. Kouzani, Kelly Decker, Robin Digby, and Tracey Bucknall. Comparative accuracy testing of non-contact infrared thermometers and temporal artery thermometers in an adult hospital setting. *American Journal of Infection Control*, 49(5):597–602, May 2021.

[28] Davis E. King. Dlib-ml: A machine learning toolkit. *Journal of Machine Learning Research*, 10:1755–1758, 2009.

[29] Daniel McDuff and Ewa M. Nowara. "warm bodies": A post-processing technique for animating dynamic blood flow on photos and avatars. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21, Yokohama, JP, May 2021. ACM.

[30] Daniel Kwok-keung Ng, Chung-hong Chan, Eric Yat-tung Chan, Ka-li Kwok, Pok-yu Chow, Wing-Fai Lau, and Jackson Che-Shun Ho. A brief report on the normal range of forehead temperature as determined by noncontact, handheld, infrared thermometer. *American Journal of Infection Control*, 33(4):227–229, May 2005.

[31] Ewa M. Nowara, Daniel McDuff, and Ashok Veeraraghavan. A meta-analysis of the impact of skin tone and gender on non-contact photoplethysmography measurements. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR) Workshops*, Seattle, WA, USA, June 2020. IEEE.

[32] Takuya Ogawa, Yukinori Tsukuda, Yuki Suzuki, Shigeto Hiratsuka, Ryo Inoue, and Norimasa Iwasaki. Utility of thermal image scanning in screening for febrile patients in cold climates. *Journal of Orthopaedic Science*, 27(6):1333–1337, November 2022.

[33] U.K. Okoji, S.C. Taylor, and J.B. Lipoff. Equity in skin typing: why it is time to replace the fitzpatrick scale. *British Journal of Dermatology*, 185(1):198–199, April 2021.

[34] Ilze Oshina and Janis Spigulis. Beer–lambert law for optical tissue diagnostics: current state of the art and the main limitations. *Journal of Biomedical Optics*, 26(10), October 2021.

[35] Yigang Peng, Arvind Ganesh, John Wright, Wenli Xu, and Yi Ma. Rasl: Robust alignment by sparse and low-rank decomposition for linearly correlated images. In *2010 IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, San Francisco, CA, US, June 2010. IEEE.

[36] Chu Chu Qiu, Jing Wei Chin, Kwan Long Wong, Tsz Tai Chan, Yu Dong He, and Richard H. Y. So. Remote mass facial temperature screening in varying ambient temperatures and distances. In *2023 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 6068–6076, Vancouver, CN, June 2023. IEEE.

[37] Nirmal Ravi, Mathura Vithyananthan, and Aisha Saidu. Are all thermometers equal? a study of three infrared thermometers to detect fever in an african outpatient clinic. *PeerJ*, 10:e13283, June 2022.

[38] E. F. J. Ring, H. Mcevoy, A. Jung, J. Zuber, and G. Machin. New standards for devices used for the measurement of human body temperature. *Journal of Medical Engineering and Technology*, 34(4):249–253, April 2010.

[39] Olaf Ronneberger, Philipp Fischer, and Thomas Brox. *U-Net: Convolutional Networks for Biomedical Image Segmentation*, page 234–241. Springer International Publishing, Munich, Germany, 2015.

[40] Ayan Seal, Debotosh Bhattacharjee, Mita Nasipuri, and Dipak Kumar Basu. Minutiae based thermal face recognition using blood perfusion data. In *2011 International Conference on Image Information Processing*, Shimla, India, November 2011. IEEE.

[41] Joytech Sejoy. Sejoy digital infrared forehead thermometer det-206/306, jan 2010.

[42] Adrian Shajkofci. Correction of human forehead temperature variations measured by non-contact infrared thermometer. *IEEE Sensors Journal*, 22(17):16750–16755, September 2022.

[43] Michael W. Sjoding, Robert P. Dickson, Theodore J. Iwashyna, Steven E. Gay, and Thomas S. Valley. Racial bias in pulse oximetry measurement. *New England Journal of Medicine*, 383(25):2477–2478, December 2020.

[44] Jeffrey F Spindel, Stephen Pokrywa, Nathan Elder, and Clayton Smith. The environment has effects on infrared temperature screening for covid-19 infection. *American Journal of Infection Control*, 49(11):1445–1447, November 2021.

[45] Wally auf der Strasse, Daniel Prado Campos, Celso Júnio Aguiar Mendonça, Jamil Faissal Soni, Joaquim Mendes, and Percy Nohama. Forehead, temple and wrist temperature assessment of ethnic groups using infrared technology. *Medical Engineering & Physics*, 102:103777, 2022.

[46] Kenichiro Tanaka, Hiroyuki Kubo, Nobuhiro Ikeya, Takuya Funatomi, Tsuyoshi Takatani, and Yasuhiro Mukaigawa. Time-resolved light transport decomposition for thermal photometric stereo. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, Salt Lake City, UT, US, June 2018. IEEE.

[47] M.R. Tay, Y.L. Low, X. Zhao, A.R. Cook, and V.J. Lee. Comparison of infrared thermal detection systems for mass fever screening in a tropical healthcare setting. *Public Health*, 129(11):1471–1478, November 2015.

[48] Teledyne FLIR. *Lepton 3 Longwave Infrared (LWIR) Datasheet*, 6 2018. Rev. 204.

[49] Alexander Vilesov, Pradyumna Chari, Adnan Armouti, Anirudh Bindiganavale Harish, Kimaya Kulkarni, Ananya Deoghare, Laleh Jalilian, and Achuta Kadambi. Blending camera and 77 ghz radar sensing for equitable, robust plethysmography. *ACM Transactions on Graphics (TOG)*, 41(4):1–14, 2022.

[50] 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, New Orleans, LA, USA, 2022. IEEE.

[51] Welch. Care temp touch-free thermometer - directions for use, jan 2015.

[52] Zhihua Xie, Guodong Liu, Shiqian Wu, Zhijun Fang, and Yun Gan. Pennes equation based blood perfusion model and its application in face recognition. In *The 2010 IEEE International Conference on Information and Automation*, pages 2443–2446, Harbin, China, June 2010. IEEE.

[53] Xiaojiang Xu, Anthony J. Karis, Mark J. Buller, and William R. Santee. Relationship between core temperature, skin temperature, and heat flux during exercise in heat. *European Journal of Applied Physiology*, 113(9):2381–2389, June 2013.

[54] Mengkai Yan, Jianjun Qian, Renke Wang, Shangbing Gao, and Jian Yang. Dynamic group difference coding based on thermal infrared face image for fever screening. *IEEE Transactions on Instrumentation and Measurement*, 72:1–13, 2023.