PrivacyLens: On-Device PII Removal from RGB Images using Thermally-Enhanced Sensing

Yasha Iravantchi, Thomas Krolikowski, William Wang, Kang G. Shin, Alanson Sample
University of Michigan
Ann Arbor, Michigan, USA
{yiravan,tkrolikowski,willruiz,kgshin,apsample}@umich.edu

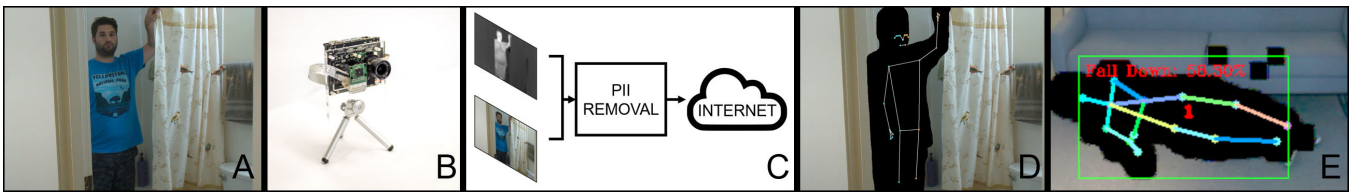


Figure 1: PrivacyLens can be used in sensitive areas (A) by leveraging thermal imaging and an onboard GPU (B) to perform on-device PII removal (C) for privacy-preserved images (D) that support existing CV/ML algorithms, such as fall detection (E).

ABSTRACT

Internet-connected cameras support many useful home monitoring and health applications. However, these same cameras indiscriminately capture sensitive and Personally Identifiable Information (PII), limiting their acceptance in certain settings, such as the home. Prior works removed Region of Interest (ROI) to secure images and improve privacy. However, the methods that rely solely on RGB information to find persons are susceptible to environmental and lighting conditions, causing them to fail and leak PII. From our deployment study, nearly half of the images containing persons had a PII leakage when using RGB-only methods. Furthermore, ROI removal is often performed off-device, requiring the server performing these operations to be trustworthy. This work presents the PrivacyLens system, where with the addition of thermal sensing, our system has a significantly enhanced ability to find persons in RGB images and video and efficiently remove them on the device before any data is stored or transmitted, all while staying under typical IoT power constraints. From our aforementioned deployment study in an office-building atrium, family home, and outdoor park environment, the PrivacyLens prototype effectively removes PII with a sanitization rate of 99.1%. Additionally, PrivacyLens can use its embedded GPU to generate on-device features for downstream CV/ML tasks, as shown in three illustrative applications, further reducing the collection and storage of PII.

KEYWORDS

privacy, computer vision, thermal, camera, embedded, sensing, PII

Proceedings on Privacy Enhancing Technologies 2024(4), 872–891

https://doi.org/10.56553/popets-2024-0146

1 INTRODUCTION

Cameras are one of the most information-rich and ubiquitous sensors in everyday "smart" devices such as phones, doorbells, displays, and thermostats. They power useful applications including public road and infrastructure monitoring [23, 32], in-home health and vital sign tracking [4, 70], fall detection [15, 59, 77], and activity monitoring [74]. As cameras have proliferated in our lives, they have raised privacy concerns regarding the data they collect, store, and send to the cloud. Unfortunately, users often have little insight into what is being done with their data, who can access it, and for what purposes once the data leaves the device. For example, users who purchase consumer devices (e.g., smart doorbells, IoT cameras) often believe their "encrypted" camera feeds are for their eyes only [35]. However, it has been shown that these feeds can be accessed by others, such as employees of the device manufacturer [18, 24, 50], data brokers and third parties [41], hackers [66], and law enforcement agencies without warrants [17]. Furthermore, these cameras indiscriminately capture Personally Identifiable Information (PII), which may be irrelevant to a specific task, such as a robot vacuum mapping a home [31]. These incidents create mistrust, hampering the adoption of these devices in the home and in applications where they are needed most, such as fall detection in the bathroom—the main cause of death for those over 65 [10].

While prior approaches have looked towards Region of Interest (ROI) removal to sanitize sensitive information in images [48, 67, 69], the aforementioned privacy incidents highlight the importance of where and when the sanitization happens. Transmitting raw images to a server for processing still holds potential for abuse. One additional challenge is that, in real-world environments, RGB-only approaches for detecting persons have been shown to be susceptible to environmental and lighting effects [6, 38], which can cause sanitization failure and leakage of PII. These issues are particularly well-known in the Advanced Driving and Assistance (ADAS) community, which has looked towards thermal sensing approaches to robustly find persons in images in the ADAS domain [37]. These factors highlight the importance of both on-device and environmentally robust removal of PII from images.

Thus, we propose **PrivacyLens**, a battery-powered RGB and thermal imaging system with an embedded GPU capable of robustly and efficiently removing five forms of PII (face, skin color, hair color, gender, and body shape) before any data (i.e., images or ML features) is stored or transmitted off-device. As depicted in Figure 1, this privacy-preserving camera can be deployed in sensitive home areas where traditional cameras would be deemed unacceptable. The core idea behind PrivacyLens is that the onboard RGB and thermal cameras can be used together to robustly detect persons and their thermal silhouettes, which are used to "subtract" them from images. This process can be accelerated by a Jetson Nano's power-efficient embedded GPU, which makes PrivacyLens capable of 8 FPS sanitization (FPS limits due to International Traffic in Arms Regulations (ITAR) restrictions on thermal cameras [40]) under typical IoT power limits. PrivacyLens's real-world evaluation shows comprehensive PII removal in 99.1% of images, compared to RGB-only methods that achieve ≤ 57.6% even when given the advantage of best-in-class CV algorithms running on desktop GPUs. These results show that the combination of a thermal camera and embedded GPU is the key to enabling robust, efficient, and real-time PII removal on the edge for IoT camera devices.

Furthermore, the sanitized images and on-device generated features can be used with downstream, cloud-based machine learning models to enable critical applications, such as fall detection, while protecting users' privacy. PrivacyLens supports six operational modes that sanitize varying amounts of PII in images, from swapping faces with a generic face (for maximum downstream compatibility) to removing persons entirely (for maximum privacy). The utility of these modes is demonstrated through three applications: exercise tracking, in-home activity tracking with objects, and posture and fall detection. Evaluated with ten participants, these applications demonstrate PrivacyLens is not only useful but also maintains the compatibility and performance of downstream CV/ML applications while reducing the amount of PII collected.

This paper makes the following contributions:

(1) a threat model analysis detailing three classes of camera-based IoT privacy issues addressed by this work;
(2) the PrivacyLens system that uses an RGB and thermal camera pair with an embedded GPU to efficiently accelerate robust real-time on-device image and video PII removal;
(3) an extensive evaluation of the system across multiple environments showing that thermal imaging can significantly improve PII removal rates (99.1% vs. ≤ 57.6% for RGB-only approaches) and how PrivacyLens addresses the threat model;
(4) applications demonstrating PrivacyLens's utility and its compatibility with downstream CV & ML approaches.

Ultimately, low-power, privacy-preserving cameras that can sanitize images before they are transmitted off the device have the potential to benefit end users wishing to have greater device privacy through reduced PII collection, corporate entities that wish to limit their exposure to liability to consumer privacy laws, and researchers that seek a robust approach to removing PII from images.

2 THREAT MODEL AND RELATED WORK

This section details the threat model PrivacyLens aims to address and a summary of prior works related to PrivacyLens.

2.1 Threat Model

Camera-enabled IoT systems' unbounded data collection and the often ambiguous nature of who can access personal cloud storage (e.g., support staff, IT personnel, etc) mean an adversary has many ways to capture private user information. We identify three main types of threats: *unauthorized access*, *authorized access with unauthorized sharing*, and *data over-collection*. Per our threat model, we assume physical security, as we cannot ensure effective operation if the device is physically tampered with (e.g., the sensors are altered). We additionally assume software security, where reasonable safeguards such as good passwords, patched vulnerabilities, and restricted root access are implemented to prevent an attacker from gaining access to the system and altering its behavior. These safeguards should not incur additional computing demands or affect the performance of the system. We also highlight that PII sanitization is contradictory to surveillance and public safety tasks, where personal identity is necessary and defeats the purpose of the camera. As such, we do not consider these situations part of our threat model.

In *unauthorized access*, an attacker gains access to images containing sensitive information through a tool or attack. This is a well-understood threat, given the frequency of data breaches through compromising image streams directly or the server that stores the image data. For example, attackers can gain access directly from compromised in-home cameras [35]. Historically, investing in greater data security (e.g., firewalls, securing servers, encrypting data) has been the approach to counter this threat.

In *authorized access with unauthorized sharing*, an attacker may have legitimate access to the images (e.g., was given the user's login credentials by the user, is an employee with proper access), but uses that access inappropriately to transfer the data elsewhere. For example, the Massachusetts Bay Transportation Authority (MBTA) has CCTV cameras near subway entrances for maintenance and public safety purposes. However, someone at the MBTA shared an embarrassing video of a woman falling down an escalator, which included identifying features such as hair color, skin color, and face [65]. While the video was removed from YouTube, a search returns many copies. A particular challenge with this threat is that once shared inappropriately, the ease of sharing again makes it nearly impossible to know who can access it and what they will do with it. As noted earlier, removing PII is contradictory to public safety. However, for an employee tasked with doing public maintenance, such as identifying leaks or escalator status, sanitizing these public recordings of PII would not hinder them from performing their duties but would actively prevent them from inappropriately sharing embarrassing content containing PII.

In *data over-collection*, an attacker may have access to information that the user is unaware is being collected, irrespective of the means by which the attacker gained access to this information. In this case, an attacker takes advantage of a user agreeing to share some information but gains access to information outside of what the user agreed to. For example, Roomba users understood their vacuum has an obstacle-avoidance camera, but did not expect it to collect sensitive images of them, such as while using the toilet [31].

PrivacyLens addresses these threats by removing unwanted PII before generating the image. In the case of *unauthorized access*, if

an attacker were to gain access to a repository of PrivacyLens images, the exposure would be significantly reduced since the images would be heavily sanitized and no "raw" versions exist. *Authorized access with unauthorized sharing* is similarly addressed, as images shared without authorization would not have PII. PrivacyLens design choices, however, are most tailored to address the *data over-collection* threat. Since PrivacyLens is selective of the content in the image (either by removing PII, only generating features, or only including relevant parts of an image), by design, it limits the total amount of recorded information. For the robot vacuum, people would be removed from images, but obstacles would remain in a way that could be safely used to train obstacle avoidance models. Ultimately, PrivacyLens presents a proactive intervention to improve the privacy of people in camera view, including those who own the device, bystanders, and the public.

2.2 Related Work

Notions of Privacy & PII. Merriam-Webster defines the right to privacy as "the right of a person to be free from intrusion into or publicity concerning matters of a personal nature" [53]. However, defining the contours of this right is a perpetual challenge: Karachalios [45] describes privacy as a "wicked problem" with many valid definitions that can co-exist but no one-size-fits-all solution. Thus, we require quantifiable definitions that can be addressed to improve privacy surrounding cameras. Relevant to privacy is the handling of Personally Identifiable Information (PII). The National Institute of Standards and Technology (NIST) defines PII as "any representation of information that permits the identity of an individual to whom the information applies to be reasonably inferred by either direct or indirect means" [57]. However, NIST's overly broad definition is hard to apply as a privacy standard. Recently, the California Privacy Rights Act [39] defined what constitutes specific identifiers, including face, skin and hair color, gender, and body shape. PrivacyLens removes these personal identifiers as a quantifiable way to improve user privacy.

Cameras and Privacy in Smart Homes. In-home applications, such as health monitoring, have been a significant factor that drove the adoption of ubiquitous sensing technologies in the home, especially among senior citizens [9]. However, this population prefers simpler single-purpose sensors (e.g., heart rate monitor) over more privacy-invasive general-purpose sensors, such as microphones [42] or cameras [25]. How people appear in images matters, as both Caine et al. [9] and Jacelon et al. [43] found that in smart environments, users are amenable to some representations, such as dots representing a silhouette of a person. However, they do not accept raw images or having their face or sensitive body parts recorded. Griffiths et al. [29] avoided RGB cameras and explored using only thermal imaging to track the in-home movement of individuals. However, this approach limits identifiable activities and discards rich non-PII-containing RGB pixels that could contain valuable information, such as what objects are in the environment. As shown later in one of PrivacyLens's applications, these non-PII-containing RGB pixels help identify what objects a person interacts with, even if no PII-containing RGB pixels are stored.

The kind of activities captured by cameras influences acceptance, as Choe et al. [14] found that reading a book or watching TV did not evoke privacy concerns but sexual or hygienic activities did. Similarly, Hoyle et al. [36], in the context of wearable first-person "lifelogging" cameras, identified numerous contexts where people expressed privacy concerns, such as situations where private or sensitive content could be captured. Additionally, they found participants disliked the burden of reviewing and deleting private information from the lifelog and would rather disable the camera entirely. Viewing these works through the lens of Nissenbaum's theory of privacy as contextual integrity [56], cameras are not incompatible with continuous in-home operation but must meet user privacy needs within the context in which they are deployed.

While the previous works have focused on addressing the privacy concerns of the user benefiting from the camera's services, others have looked at the privacy concerns created for bystanders. As Marky et al. [49] found, a collateral effect of data over-collection is that, even if the host of an IoT-equipped household consented to data collection, their guests are forced into the perilous and sometimes awkward task of finding ways to protect their own privacy from unconsented IoT data collection. Dimiccoli et al. [21] addressed this issue in images taken by wearable cameras by utilizing deep learning to detect bystanders and selectively degrading those portions of the image. Alharbi et al. [2] built upon this work by proposing an approach that masks all but the pixels required to capture a wearer's hand-related activities. PrivacyLens provides a level of privacy to bystanders by removing all PII in an image, as it does not discriminate whether the PII belongs to the camera's owner or a bystander. Given device constraints, however, we opted for a less computationally intensive thermal subtraction approach to achieve on-device sanitization. Future implementations could incorporate these more advanced approaches on-device.

Hybrid Computer Vision using Thermal Imaging. The most prominent use of long wave IR (LWIR) thermal cameras is in Advanced Driver Assistance Systems (ADAS) that have sought to improve collision avoidance with pedestrians [12]. RGB-only methods are insufficiently robust for pedestrian avoidance, especially in conditions where traditional optical methods fail, such as nighttime [6, 38]. As a result, radar, LIDAR, and thermal camera-based methods have emerged as a more consistent approach in detecting pedestrians, finding 90% pedestrian detection accuracy [13]. Non-vehicle person detection uses of thermal cameras include automated temperature measurement for COVID-19 [3, 55] and uncrewed aerial search and rescue [33], demonstrating effective person detection in both near- and far-range applications. Most similar to PrivacyLens is work by Zhang et al. [75] that utilized an IR camera paired with a cold mirror to identify and mask faces in images. However, this approach is limited to removing only facial PII, and the use of a mirror makes the system physically large and increases cost and complexity, unsuitable for typical IoT deployment. PrivacyLens builds upon these prior works to leverage thermal sensing as a robust, environment- and condition-invariant approach to find the entire bounds of a person and remove them entirely from images.

Embedded GPU Use in Privacy-Sensitive Applications. While Single Board Computers (SBCs), such as Raspberry Pis, have made it easier for researchers to build novel applications, they have only recently become capable of performing significant computational tasks. The introduction of embedded GPUs (e.g., Jetson Nano) and embedded Tensor Processing Units (TPUs) (e.g., Google Coral) have

replaced the need for remote computing resources in several applications, such as portable sign language detection [76] and emotion recognition [5]. More relevant to PrivacyLens are applications where embedded GPUs must locally process sensitive data. In these situations, only sanitized results can be transmitted, such as patient health data for heart attack recognition [54], baby facial expression monitoring for autism detection [62], and anomaly detection in EEG signals to identify substance abuse [20]. PrivacyLens similarly utilizes its embedded GPU as a prototype Privacy Accelerator to process sensitive data and remove PII on the device.

Image Masking, Blurring, and Anonymization Approaches. As discussed above, selectively masking and blurring portions of images are powerful approaches to improve camera privacy. One robust approach is to remove the entire person from the image. This can be achieved by using RGB-based object and person detection models like YOLO [44] or Detectron2 [71]. YOLO is a single-stage detector (SSD) where only a single shot determines whether or not an object (such as a person or their face) exists. This results in fast detection speeds and lower computational requirements, which are particularly useful for low-resource embedded systems. Dectectron2 is a region-based convolutional neural net (R-CNN) detector, where an initial stage finds candidate objects, and a second stage determines their class, removes false positives, and refines bounding boxes. While R-CNNs yield greater mean average precision over SSD-based approaches, they come at a significant cost to computational and memory requirements, making them incompatible with embedded devices. We use these two state-of-the-art RGB approaches to perform RGB-only whole-person PII removal, which we describe in greater detail in Section 3 and Appendix A.

Another set of approaches, including blurring faces [48], adding noise [69], and inpainting [67] have been a consistent practice in removing facial PII, but in some situations they can be reversible [52]. Lopez et al. [61] provided a comprehensive review of various visual privacy protection methods. However, many object detectors, such as the abovementioned YOLO and Detectron2, utilize the face as a positional anchor [28]; without the face in the image, the remaining pose may not be accurately detected. Swapping the face retains the anchor points, whereas removing the face can entirely break the downstream system. Deep fake approaches, where the face is replaced with an anonymous and generated face, are effective in concealing the face of the original person while maintaining compatibility with face recognition systems [16]. However, deep fake and other De-ID approaches [73] are computationally intensive and currently cannot be done in real time on SBCs. Our *Face Swap* approach similarly replaces faces in images using a more computationally lightweight approach (described in Section 3) to achieve a similar goal and maintain downstream compatibility with CV/ML approaches. Future implementations could employ deep fake approaches that would make it more difficult to identify that the face has been swapped. Lastly, there are situations where access to private information should be restricted, but the remaining content can be accessed. For example, CamShield [68] encrypts sensitive ROIs, such as those corresponding to faces, but leaves the rest of the image intact. Similarly, PrivacyLens can replace sensitive ROIs with features (such as with a stick figure representation), encrypt them, or remove them entirely, but does all these operations on-device.

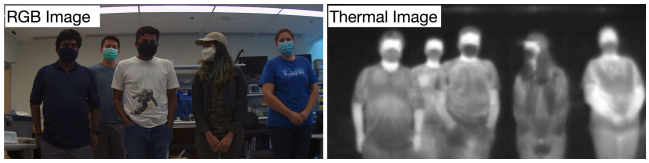


Figure 2: The Lepton 3.5 thermal camera measures black body radiation (long wavelength IR) to provide an absolute measure of surface temperature. This permits robust operation across various skin tones, body shapes, and genders.

3 HARDWARE IMPLEMENTATION

This section details the hardware implementation, design considerations, and technical benchmarks of the PrivacyLens prototype. Before developing the hardware prototype, we conducted a pilot study (Appendix A) highlighting thermal information's utility in removing persons from images over RGB-only methods. Thus, to perform on-device PII removal, PrivacyLens consists of four major components: a Single Board Computer (SBC) with an embedded GPU (Section 3.1), an RGB camera and thermal camera (Section 3.2), and a software pipeline that captures raw sensor data from the two cameras and efficiently processes them on the GPU (Section 3.4).

3.1 Embedded Devices Evaluation

We evaluated our thermal subtraction approach (see Algorithm 1 in Appendix A.1) on low-power embedded devices to create a privacy-preserving camera that removes PII on-device. We narrowed our search criteria to SBCs that are under USD $100, run under typical IoT power constraints (e.g., USB, Power over Ethernet), and have low-level capability sensor interfaces (e.g., Serial Peripheral Interface (SPI), Camera Serial Interface (CSI)). Our search yielded the Raspberry Pi 3 (USD $35) [1] and the Jetson Nano (USD $100) [58]; during development, the Raspberry Pi 4 had SPI library incompatibilities [64]. We designed two benchmarks, a thermal subtraction benchmark and a lightweight face detector benchmark, to evaluate the total performance (frames per second, FPS) and efficiency (FPS per watt) of two PII removal-related tasks on these embedded devices. For each platform, the benchmarks are optimized for their architecture. The Jetson Nano has two power modes (Max and 5 W) evaluated as separate entries. An Intel CPU and Titan RTX GPU are provided as desktop references. The power consumption of each platform is measured with a Kill-A-Watt meter [60]. An increase over idle consumption is reported, providing a direct measure of the task's power consumption and not the effects of attached peripherals (e.g., idle hard drives).

The thermal subtraction benchmark is implemented in C++ and OpenCV 4.5.2. On CUDA platforms, the OpenCV CUDA functions are used, allowing the subtraction to run entirely on the GPU. To avoid differences in I/O performance, ten image pairs (thermal and RGB) were selected randomly from the KAIST ADAS dataset [38] and are preloaded into RAM. Then, 10,000 thermal subtraction operations are applied on each pair, resulting in 100,000 subtractions. The RGB-based facial landmark evaluation utilized platform-optimized versions of MediaPipe, loaded a test image (SciKit-Image

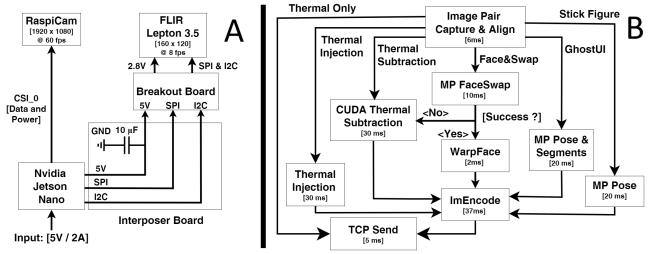


Figure 3: (A) the hardware block diagram of the PrivacyLens prototype showing connections between the RGB camera, thermal camera, Jetson Nano, and a custom interposer board; (B) the software block diagram data flow for different PII sanitization operations and their execution time.

astronaut.jpg) into RAM, and performed a facial landmark detection (458 key points) task 100,000 times. The average FPS, power consumption, and efficiency are reported for both benchmarks.

Overall, we find these two benchmarks perform significantly better on a GPU, not only in total performance but also in efficiency. In particular, the Jetson Nano's GPU was 13x faster while being 16x more efficient than the Raspberry Pi. While the Jetson Nano did not have the highest total performance, bested by the desktop-class Titan RTX, it did have the highest power efficiency across all devices. We note that when the benchmarks use the GPU, the CPU utilization is relatively low, freeing the CPU to perform other tasks. The complete results are in Table 5 in Appendix A.1.

3.2 RGB and Thermal Cameras

While USB webcams offer an easy interface to capture RGB images, they largely hide access to low-level adjustment (e.g., exposure, aperture) and are often incompatible with interchangeable lenses. The Raspberry Pi HQ, an RGB camera for SBCs, connects via CSI for low-level control and accommodates a 6 mm lens to closely match the thermal camera's 57° field of view (FoV). A custom driver interfaces the raw camera output to Jetson's GPU-accelerated GStreamer pipeline at 1080p / 60 FPS with all "auto" settings disabled.

Ideally, a thermal camera could operate at the exact resolution and frame rate as the RGB camera so that each frame can be easily aligned and time-synchronized. However, high-resolution thermal cameras with > 9 FPS are subject to strict ITAR export restrictions and are not easily purchasable [40]. Thus, we selected the FLIR Lepton 3.5, which is reasonably available and takes 160×120 pixel thermal images up to 8 FPS. The Lepton 3.5 is a radiometric sensor, meaning that each pixel value corresponds to a calibrated absolute temperature and is not relative to the thermal content of the environment. Importantly, these thermal cameras measure the black-body radiation emitted by a person and are not affected by illumination conditions or the temperature of the environment. Studies have shown no significant differences in emissivity between people of different skin types, meaning that computer vision (CV) algorithms based on thermal cameras are unlikely to inherit negative racial biases based on skin color [11, 51]. For example, Figure 2 shows consistent thermal imaging across skin tones and genders. We mount both cameras on the same Y and Z plane and as close as possible side-to-side, improving the ability to make a close alignment and

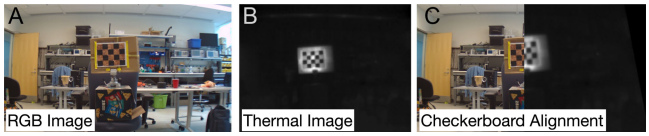


Figure 4: Raw image from the 1980×1080 pixel RGB camera (A), raw image from the 160×120 pixel thermal camera (B), a split view of the calibrated, flattened and aligned RGB and thermal images (C). The copper/black checkerboard colors are inverted in the thermal image since the copper dissipates heat and is cooler than the black squares.

avoiding the necessity of a cold mirror, which introduces additional costs and increases the prototype's overall size and complexity. We route the signals through a custom-designed interposer printed circuit board (PCB) and ribbon cable to improve the stability of the Lepton's 20 MHz-clock Video over SPI transmission. The hardware connections of the PrivacyLens prototype can be found in Figure 3.

3.3 Camera Image Alignment

To align the images, we fabricated a copper and paper checkerboard, which creates a matching checkerboard pattern in the thermal image when heated with a heat gun. We use this checkerboard to find corresponding points and perform a perspective transform to align the thermal image to RGB. Figure 4 demonstrates the alignment. Since the 8 FPS limit of the Lepton is not an even multiple of the 60 FPS RGB camera, the latest RGB frame available from the GStreamer pipeline is used before requesting a frame from the thermal camera, which produces a worst-case synchronization offset of 16.67 ms. This means that fast-moving persons will have an offset between the RGB and thermal images, potentially causing PII leakage. To address this issue, when using the thermal silhouette of a person to subtract corresponding pixels in the RGB image, the subtraction mask is dilated as a buffer for these situations. All results in our work include this dilation with the system operating in real-time. Future implementations with time-synchronized RGB and thermal cameras would also address this limitation.

3.4 Operational Modes & Prototype Evaluation

PrivacyLens uses both RGB and thermal information for PII removal in its hybrid *Thermal Subtraction* mode by using YOLO [44]—an object detection approach suitable for embedded devices—to identify bounding boxes corresponding to persons in the RGB image and by using the thermal image to create a segmentation mask, where the thermal silhouette of a person is identified by identifying pixels that are within a specified temperature range corresponding to human skin temperature, as described in Algorithm 1. The union of these two regions defines the pixels to be sanitized from the RGB image. This approach complements YOLO, which can mask portions of images where thermal alone may have missed (e.g., cold extremities), and thermal can remove PII from images where RGB-only methods struggle to find persons accurately, such as individuals with awkward postures or facing away from the camera.

In addition to the *Thermal Subtraction*, representing a more comprehensive PII removal method, we envisioned five additional PrivacyLens modes that include varying amounts of information in

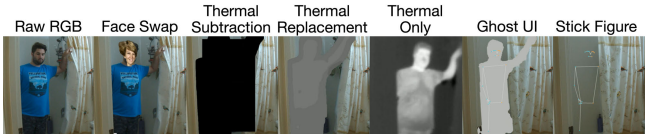


Figure 5: Raw RGB image and six modes of removing PII

an image. These modes offer flexibility for different priorities while reducing the PII encoded in an image, as shown in Figure 5.

In *Face Swap*, PrivacyLens replaces faces in images but keeps the remainder of the image intact, ensuring compatibility with downstream CV/ML approaches. *Face Swap* identifies facial landmarks for each face in an incoming RGB image and performs a perspective transform to place a template face onto the face(s) in the image. This approach removes facial PII while addressing a weakness of reversible face-blur interventions [52]. The face-swapped images are confirmed to be compatible with MediaPipe's Pose and Facemesh, OpenPose, MoveNet, and iOS's built-in face detector. If *Face Swap* is unsuccessful for a given frame, PrivacyLens can default to *Thermal Subtraction* mode or drop the frame entirely.

In *Thermal Replacement*, a person in the RGB image is replaced by a low-resolution thermal silhouette of that person's temperature. In *Thermal Only*, only the thermal image is generated. These two represent situations where a person's temperature, not their PII, is valuable for a task. For example, these representations could be useful in a public health task to gauge the prevalence of fever with reduced PII collection. The low-resolution nature of the thermal camera precludes it from being used for facial recognition.

The remaining two, *Stick Figure* and *Ghost UI*, inspired by Caine's point-light and activity blob representations [9], use the RGB image to find pose landmarks and annotate an optional background image with a stick figure representation. The *Ghost UI* mode adds a person's body shape as a silhouette. Compared to *Thermal Subtraction*, these modes are useful for situations where skeletal keypoint information is needed but specific PII-containing pixels are not, such as detecting when a person is interacting with objects in the environment. These modes can be overlaid on top of the previous modes; adding *Stick Figure* to *Thermal Subtraction* can selectively include additional information and increase the utility of the generated image without reintroducing a large amount of PII. These additional modes and interventions build upon the baseline *Thermal Subtraction* PII removal approach, which is evaluated in Section 4, and inherit its PII removal performance (e.g., Stick Figure), unless they are explicitly permissive of a subset of PII (e.g., Face Swap).

Figure 3 presents PrivacyLens's software block diagram with the average completion time for each operation at 1080p resolution. At 8 FPS, all modes' operations consumed < 5 W, permitting easy integration with various power sources. Finally, PrivacyLens requires ≈13.5 Mbps bandwidth for 8 FPS image transfer at 90% JPEG quality, enabling wireless deployments. The system is designed to run in real time but is limited by the thermal camera (8 FPS), thus resulting in a latency of 125 ms plus network latency. These sequential images are streamed off the device as 8 FPS sanitized video. PrivacyLens's mobile housing shown in Figure 6 includes a 7.4 V 5,200 mAH battery, a fan, and a 5 V step-down regulator. A single charge sustained 5 hours of continuous operation.

Figure 6: Image of the PrivacyLens prototype in a tripod-mounted 3D-printed enclosure (A), images from the deployment study showing the office atrium (B), home living room (C), and public park square (C). The thermal camera's images have been contrast-enhanced for ease of viewing.

4 DEPLOYMENT EVALUATION

While the results from our pilot study (see Appendix A) show promise that thermal information can aid in removing persons, and thus PII, from images, we acknowledge that the pilot study's dataset reflects different environments than where we expect PrivacyLens to be deployed and most useful. Thus, we now corroborate this promise in real-world environments. This section details how the PrivacyLens prototype was deployed to collect and sanitize images in three environments (office, home, and public park) and how the images were evaluated for PII content in two studies: Study 1 tasked annotators to find PII with the assistance of the unsanitized image as a ground truth reference and Study 2 tasked annotators to find PII without a ground truth reference.

4.1 Deployment Environments

In human-centric environments, such as the workplace and home, people can be very close to the camera and are more likely to create challenging conditions for removing PII. These situations include having a person partially in the frame, frequently coming in and out of frame, persons in postures that are more challenging for detection and removal (such as sitting cross-legged), or have objects partially obscure them (such as sitting at a table). Indoor air conditioning can also have a different effect on skin temperatures compared to outdoor environments. This variety, which makes PII removal more challenging, is not often captured in prior image sanitization works or existing public ADAS datasets, where persons are often fully in the frame, upright, and standing. Thus, to determine PrivacyLens's ability to remove PII in real-world environments under real hardware constraints, it was evaluated in a workplace and home in addition to a public park, as shown in Figure 6.

In the first environment, the system was deployed in a 40×60 foot multi-story office atrium (Figure 6 – panel 'B') at two locations: on the ground floor and the second floor looking into the atrium mimicking an overhead fixed IoT camera. During deployment, people walked individually or in groups across the atrium, worked at tables, and lounged on couches. In the second environment, the system was deployed in a home living room (Figure 6 – panel 'C'). At this location, people are particularly close to the camera and it was common for people to come in and out of the frame, face away from the camera, and relax on furniture with odd postures. This family consisted of two adults, three children, and a cat. During deployment, the parents tended to the children, and the children

ate, played, and lounged. In the third environment, the system was deployed in a public square, shown in Figure 6 – panel 'D', with people at a wide range of distances. There is no control over factors such as lighting and weather. The environment was generally well-lit, and the ambient temperature outside was ~90°F. During deployment, people walked across the park square, lounged on steps and benches, and sat on the ground. This environment also included several people riding bicycles, skateboards, and scooters.

Ethical Considerations in Deployment. Beyond gaining Institutional Review Board (IRB) approval for all data collection activities and discussions with relevant authorities, we took into account additional ethical considerations. Regarding bystander privacy, in the office building, we posted signs with our contact information denoting that images were being captured (no audio) and discussed and received consent from the building management. For the public park, we had a similar sign on our device and stood next to it during deployment so that anyone could come by if they had questions or concerns (nobody came). Regarding the children in the living room setting, before deployment, we discussed the study and received parental consent to record and use the images, and parents reviewed the images. Furthermore, the parents were in the living room during the deployment. In all three areas, the prototype was placed in an obviously visible location (e.g., near the center of the park square), well within a bystander's view. Lastly, we selected camera views that would minimize intrusion yet would yield a representative dataset. For example, in the atrium, the camera was pointed away from individual offices, and in the living room, the camera was pointed away from the bedrooms.

4.2 Deployment Procedure

In each environment, PrivacyLens was placed on a tripod and battery powered as shown in Figure 6 – panel 'A'. The PII-removed (*Thermal Subtraction*) frames were sent over Ethernet to a laptop for storage. Additionally, the raw RGB and thermal images were recorded for ground truth and a baseline RGB-only sanitization evaluation. In an actual deployment, only the sanitized image would leave the device. The frame rate was set to 1 FPS to maximize experiment runtime and restrict dataset size, though the system can perform 8 FPS sanitization per ITAR restrictions on thermal FPS. Each image trio is ≈9 MB when stored to disk, ≈0.5 GB/hr at 1 FPS.

The deployment aimed to capture roughly one hour of data per location where people were in the environment. However, there were often long stretches in the building and park where no person was in the environment. Thus, these deployments were extended to effectively capture one hour of "people" time. In total, 22,780 image trios (3,261 containing persons) were collected at these locations:
Building Atrium: From the first floor, 6,432 image trios (98 containing persons) in two sessions (morning and afternoon) spanning 3.5 hours. From the second floor, looking down towards the first floor, 6,441 image trios (590 containing persons) in two sessions (morning and afternoon) spanning 3.5 hours. Images from these two spots are evaluated and reported together.
Home Living Room: 3,694 image trios (2,144 containing persons) in one continuous session spanning one hour.
Public Park Square: 6,213 image trios (429 containing persons) in one continuous session spanning 1.7 hours.

Figure 7: An example unsanitized image (A) collected from the home environment and sanitized images using PrivacyLens's hybrid *Thermal Subtraction* (B), RGB-only YOLO (C) and RGB-only Detectron2 (D). PrivacyLens robustly removed both persons in the image, but also removed a portion of the sofa warmed by sunlight. YOLO failed entirely to remove the person sitting on the sofa and leaked the skin color of the person at the bottom of the frame. Detectron2 completely removed the person sitting on the sofa, but similar to YOLO, leaked skin color at the bottom of the frame.

Annotator Recruitment. To quantify PrivacyLens's ability to remove PII alongside two RGB-only approaches for comparison, a total of 48 annotators (undergraduate and graduate students) were recruited to perform various image-related tasks in a multi-step process. For Study 1, 32 annotators were recruited. Two identified ranges of frames containing people, 16 annotated PrivacyLens-sanitized images, and 14 annotated images sanitized by RGB-only approaches. For Study 2, 10 additional annotators were recruited, where 4 annotated PrivacyLens-sanitized images and 6 annotated RGB-only sanitized images. Finally, for the evaluation in Section 4.8, an additional 6 annotators were recruited to annotate more aggressive RGB-only sanitization approaches. They were compensated with food for their participation. The following subsections provide greater detail of the annotation tasks in each study.

4.3 Study 1 (PII Removal with Ground Truth)

Given the large number of captured images and significant time periods without people in view, it was necessary to pare down the dataset to images likely to contain people. Two annotators used a photo gallery view user interface to scroll through the RGB images sequentially and identified ranges of frames containing people. This resulted in 6,274 image sets (sanitized and raw RGB) that would be used for all further annotation tasks. These two annotators did not take part in the later annotation tasks to identify PII in the images.

In this study, annotators were given the raw RGB and sanitized output images for comparison and have the most context when searching for leaked PII in the sanitized image, offering a lower bound on system performance. As a baseline for comparison, the images were sanitized using two RGB-only approaches: YOLO [44] presents an RGB-only baseline for what current embedded devices can do in real-time, and Detectron2 [71] provides an RGB-only baseline for the state-of-the-art, which requires a desktop GPU for real-time sanitization. While many other sanitization approaches exist, as mentioned in Section 2, we utilize these RGB person-detection approaches because they have shown effectiveness in finding and removing the entire bounds of a person, rather than just blurring or masking only the face (see Appendix A for more details). A codebook was not used in annotation, but annotators were shown at least ten example images that contained each of the five forms of

Table 1: Study 1 PrivacyLens PII removal success rates on the 3,261 collected images.

PrivacyLens (Hybrid, Embedded)	Atrium	Home	Park	All Env.
Face	99.9%	100%	100%	99.9%
Skin Color	97.4%	98.5%	98.8%	98.3%
Hair Color	99.1%	99.1%	100%	99.2%
Gender	99.9%	99.9%	100%	99.9%
Body Shape	99.4%	99.9%	99.5%	99.8%
All PII Removal w/ Ground Truth	**97.4%**	**97.8%**	**98.4%**	**97.8%**

PII leakage (or a combination thereof) and were instructed to be as critical as possible and lean on the side of flagging an image if they were unsure about whether the image contained PII, such that the results provided a conservative estimate of scores. Each image was annotated once per study: once by an annotator in the first study with ground truth and once again in a second study by another without ground truth, detailed in Study 2 below.

For the first study, using this subset of images that likely contained persons, annotators were presented with both the raw RGB image and the PII-removed image through a Web UI, which could only be accessed through a private network. The annotators were assigned images from a randomly shuffled list and could not receive an image they had seen before as it would be removed from the list after annotation. The annotators were also prevented from being assigned sequential images, which could be similar to an image they had seen before. The annotators were instructed to determine if the PII in the ground truth raw RGB image was also present in the PII-removed image. If PII was present, annotators were asked to identify it as one of the following classes: Face, Skin Color, Hair Color, Gender, and Body Shape. Here, Body Shape refers to features such as height or weight that could be used to identify an individual. If no PII was visible in the PII-removed image, they were instructed to select "No PII". The annotators were asked to be especially thorough, with no time limit set for each image pair, and to lean on the side of flagging an image if they were on the fence about the PII content. The annotators were also permitted to report issues with the image, such as if there were no persons in the raw RGB image or technical issues (e.g., image quality issues, image not loading). This process was conducted for all three conditions: PrivacyLens's proposed approach and the two baseline approaches, YOLO (embedded) and Detectron2 (desktop). A group of 16 annotators annotated the PrivacyLens-sanitized images and a separate set of 14 annotators annotated the images sanitized by the baseline RGB-only approaches. Overall, annotators annotated 9,783 sanitized images (3,261 collected images, sanitized in three ways).

4.4 Study 1 Results

PrivacyLens's Hybrid Approach. In total, 73 images out of 3,261 collected images containing persons were flagged as containing some form of PII, resulting in an across-environment complete PII sanitization rate of 97.8%. For individual environments, PII sanitization rates are: office atrium = 97.4%, home = 97.8%, and public park = 98.4%. Examining the types of PII leakage, the majority of PII issues came from skin color and hair color exposed, often at the extremities (e.g., top of the head, fingertips), which affected 86.3% of the 73 images containing PII leakages. Only one image in the entire dataset was flagged as containing a face after sanitization. No images had "total failures," which we define as when all five types of PII were exposed. Table 1 provides summary statistics of the PII sanitization rates. The individual PII removal rates (e.g., face, gender) are generally higher than the "All PII Removal" (APR) rates. This is due to 15.1% of images with PII leakage having more than one PII issue, meaning it is more difficult to remove all 5 types of PII than any one particular type of PII.

Table 2: Study 1 YOLO and Detectron2 PII removal success rates on the 3,261 collected images (each). Note: Detectron2 required a desktop GPU.

YOLO (RGB Only, Embedded)	Atrium	Home	Park	All Env.
Face	87.9%	83.5%	70.6%	82.5%
Skin Color	62.9%	43.0%	33.4%	45.3%
Hair Color	66.9%	57.0%	49.2%	57.7%
Gender	83.9%	77.4%	61.8%	76.4%
Body Shape	82.8%	78.5%	62.8%	77.2%
All PII Removal w/ Ground Truth	**52.4%**	**33.7%**	**28.6%**	**36.4%**
Detectron2 (RGB Only, Desktop)	**Atrium**	**Home**	**Park**	**All Env.**
Face	97.7%	95.8%	94.7%	96.0%
Skin Color	60.1%	54.7%	62.7%	56.8%
Hair Color	67.9%	56.9%	60.4%	59.4%
Gender	94.5%	93.0%	92.3%	93.2%
Body Shape	89.9%	92.5%	89.9%	91.7%
All PII Removal w/ Ground Truth	**44.0%**	**37.4%**	**40.8%**	**39.1%**

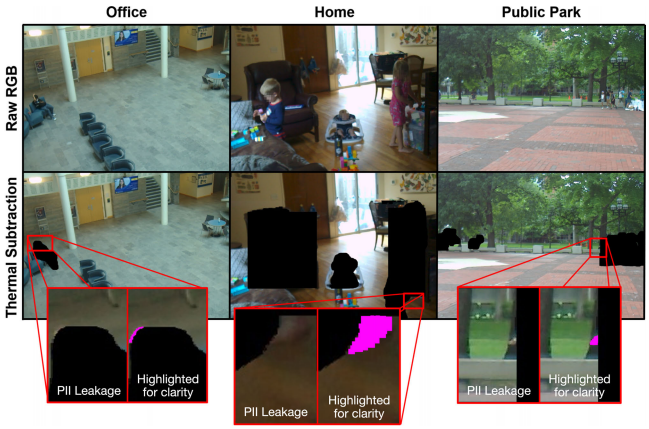


Figure 8: Examples of edge conditions of PII leaks highlighted in pink. (A) Hair is revealed from a person sitting in a chair. (B) Skin color is revealed indoors from the foot. (C) Skin color is revealed outdoors from the hand.

Additional PII leakage occurred at the edges of the masking operation. For example, the image in Figure 8 (A) shows a person sitting in the office atrium, which was flagged as containing hair color, with the PII highlighted in pink for clarity of presentation. Likewise, Figure 8 (B) shows PII leakage in the home setting, where the foot of the child is visible and thus represents a skin color PII leakage, which is highlighted in pink. Finally, Figure 8 (C) shows PII leakage in the park, where a person's hand is visible. Although the root causes of these PII leakages are less clear, we believe that a lack of pixel-to-pixel correspondence between the RGB and thermal cameras, along with poor time synchronization, has led to these

Table 3: Study 2 PrivacyLens PII removal success rates.

PrivacyLens (Hybrid, Embedded)	Atrium	Home	Park	All Env.
Face	100%	100%	100%	100%
Skin Color	99.4%	99.5%	98.8%	99.4%
Hair Color	99.9%	99.6%	100%	99.8%
Gender	100%	100%	100%	100%
Body Shape	100%	100%	100%	100%
All PII Removal w/o Ground Truth	**99.3%**	**99.1%**	**98.8%**	**99.1%**

edge failures. Section 6 details how future versions of PrivacyLens could address these edge cases.

RGB-only Baseline Conditions. The baseline RGB-only conditions exhibited significantly more pronounced PII leakages. A substantial contributor to the low APR rates are "total failures" where the RGB-only approaches missed a person entirely and thus exposed all five types of PII. Common causes of these failures were due to a person facing away from the camera or having an odd posture while sitting or lying down. Annotators flagged 14.6% of YOLO-sanitized images as containing all five types of PII. Detectron2, compared to YOLO, reduced the "total failure" rate to 3.1%, primarily due to its significantly improved face and body detection models, resulting in more people being redacted from images. However, Detectron2's APR rate of 39.1% is only marginally better than YOLO's of 36.4%, since both have poor performance in drawing accurate bounding boxes, resulting in many skin color and hair color PII leakage events. For reference, PrivacyLens's hybrid approach had zero "total failures" and achieved a 97.8% APR rate. Summary statistics for both baseline conditions can be found in Table 2.

4.5 Study 2 (PII Removal w/o Ground Truth)

While the previous study used the raw RGB images as a reference to help annotators complete the most stringent evaluation of PII leakage possible, it does not represent the privacy threat model of real-world deployment since PrivacyLens would not store or transmit the raw RGB image for comparison with the PII-sanitized output image. Thus, without the context of the raw RGB image to help "fill in the blanks", it may be challenging to identify the few strands of hair exposed, the child's foot, or the person's hand in identified Figure 8. Therefore, we conducted a second study to evaluate what types of PII leakage are detectable using only the sanitized images. In this second study, a separate set of 10 total annotators was recruited.

In the case of the PrivacyLens-sanitized images, a new dataset containing all 73 images flagged as containing PII and 73 randomly selected images containing no PII (creating a 50–50 balance to not bias the results) was prepared for annotation. Then, 4 new annotators were asked to review and label a total of 146 PrivacyLens-sanitized images for PII. For the baseline comparisons using RGB-only techniques, more than 50% of the 3,261 sanitized images were flagged as containing at least one type of PII. Thus, to reannotate every image that was flagged as having PII, we could not maintain a 50–50 balance (i.e., there was an insufficient number of non-PII images to balance), and the entire dataset was reannotated (3,261 images) for both YOLO and Detectron2, and 6 new annotators annotated without ground truth.

Table 4: Study 2 YOLO and Detectron2 PII removal success rates. Note: Detectron2 required a desktop GPU.

YOLO (RGB Only, Embedded)	Atrium	Home	Park	All Env.
Face	96.9%	90.1%	79.5%	89.9%
Skin Color	85.0%	55.8%	49.9%	60.2%
Hair Color	84.4%	67.9%	55.1%	69.1%
Gender	94.2%	85.5%	69.5%	84.9%
Body Shape	93.1%	85.7%	69.9%	84.9%
All PII Removal w/o Ground Truth	**79.0%**	**49.7%**	**45.6%**	**54.4%**
Detectron2 (RGB Only, Desktop)	**Atrium**	**Home**	**Park**	**All Env.**
Face	95.9%	97.3%	93.5%	96.5%
Skin Color	72.6%	73.8%	74.0%	73.6%
Hair Color	74.4%	69.1%	63.9%	69.4%
Gender	95.0%	95.1%	90.5%	94.4%
Body Shape	92.7%	94.0%	88.8%	93.0%
All PII Removal w/o Ground Truth	**58.4%**	**57.8%**	**55.6%**	**57.6%**

4.6 Study 2 Results

PrivacyLens's Hybrid Approach. Of the 73 images previously flagged as having PII, only 30 were confirmed to have PII in this second study, and none of the randomly selected "clean" images were identified as having PII. Of the 30 images marked as having leaked PII, only skin color and hair color were exposed. No other PII exposure types were identified. It should be noted that the one image previously marked as exposing a *face* in Study 1 was relabeled by new annotators as *skin color* since only part of the side of the face was visible. This resulted in a "Sanitized Images Only" rate of 99.1% for all PII across all environments. Per environment results are shown in Table 3.

RGB-only Baseline Conditions. Even without ground truth, annotators still identified a significant number of "total failures": 8.2% for YOLO (14.6% with ground truth) and 2.9% for Detectron2 (3.1% with ground truth). Furthermore, they flagged nearly half of all images as containing PII for both YOLO and Detectron2 (see Table 4). Skin color and hair color remained a significant weakness, and their sanitization rates marginally improved compared to the ground truth reference condition, suggesting very obvious exposures in many images. For reference, PrivacyLens had zero "total failures" and an APR rate of 99.1% without ground truth.

Overall, these studies demonstrate that PrivacyLens robustly removes PII on the device in real-time and across multiple environments. Additionally, Detectron2 did not improve PII sanitization rates over YOLO, despite being significantly more advanced and consuming 180 W on a desktop GPU. This suggests that even if more advanced RGB-only algorithms were used on embedded devices, an alternative sensing approach is required for robust PII sanitization. Ultimately, these studies show the critical importance of thermal information in performing PII-removal tasks where PrivacyLens presents a strong first-line defense for on-device PII removal.

4.7 Effectiveness Against Threats

Here, we relate PrivacyLens's performance to the threat model described in Section 2.1. Since 0.9% of images retained a form of PII after sanitization, all three threat scenarios potentially benefit from this leakage. However, the context of the threat determines whether that PII is useful for an attack. In the *unauthorized access* case, an attacker harvesting PII will at most find PII in 0.9% of images, but

none would contain face, gender, or body shape. Similarly, for *authorized access with unauthorized sharing*, an attacker who shares images will only share PII in 0.9% of images. In one particular form of the *authorized access with unauthorized sharing* case, where the attacker's goal is to share embarrassing content, they may have more difficulty since 100% of body shapes, genders, and faces were obscured, making it hard to identify who was in the shared image. For *data over-collection*, since PrivacyLens can robustly remove PII, the ability of the attacker to leverage unintended PII leakage is reduced to 0.9%. Furthermore, PrivacyLens's various modes can selectively construct images from individual semantic elements to only explicitly what the user wishes to capture, effectively reducing data over-collection. For example, PrivacyLens can generate a composite image where a person's skeletal keypoints are overlaid on a background. While this may reduce accuracy, it removes all five PII forms in *every* image as failure yields just the background.

4.8 More Aggressive RGB-only PII Removal

Our results above highlight that RGB-only methods struggled to consistently remove all five forms of PII content from images, particularly skin and hair color. These leakages often came as a result of the predicted bounding box being too small to remove all PII completely. To investigate how sensitive CV-only PII removal approaches are to bounding box alignment, we can increase the aggressiveness of PII removal for these RGB-only methods by dilating the size of their predicted bounding boxes. Similarly to our thermal model, which dilates the thermal mask size by 14% (in terms of pixels) on average, we increased the predicted bounding box for YOLO and Detectron2 by 14% (in terms of pixels). With these "more aggressive" RGB-only approaches, we re-sanitized the images and recruited six additional annotators to annotate 6,522 images (3,261 sanitized two ways) with ground truth similar to in Study 1.

Overall, both RGB-only approaches benefited from dilating the bounding box size. Aggressive versions of YOLO and Detectron2 significantly improved skin color and hair color removal rates, which improved to a total sanitization rate of 65.8% and 87.4% APR rates, respectively. These numbers even improved on the Study 2 results, where annotators did not have ground truth, suggesting that the dilation removed many of the more obvious PII leakages that did not require ground truth to identify. However, even with this aggressive dilation, these results show that a significant number of images—34.2% and 12.6% for YOLO and Detectron2, respectively—retained a form of PII leakage. All the images with a "total failure," where persons are missed completely, were flagged again as total failures, matching the results from Study 1. The full results for this study can be found in Table 6 in Appendix B.3.

Thus, even if these images were sanitized again with an even greater dilation factor, this would not address these images where the RGB-only methods failed to remove a person from the image and may ultimately create issues for downstream applications by further removing a greater number of pixels that do not contain PII. These results again highlight the value of thermal information as a secondary information channel for finding persons in situations that are challenging for RGB-only approaches.

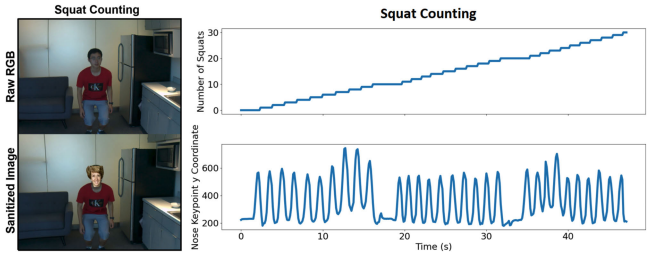


Figure 9: Result timeline of three sessions of PrivacyLens of the squat counting and detection application using squat pose trace with the face swap privacy preserving setting.

5 PII-REDUCED APPLICATIONS

This section demonstrates the effectiveness of PrivacyLens in supporting privacy-preserved versions of three useful proof-of-concept applications: exercise counting, hand-to-object detection, and fall detection. We selected these scenarios because they showcase valuable in-home CV applications but invoke privacy concerns and would benefit from on-device PII removal or feature generation. We evaluated these PII-reduced applications through a 10-participant user study, which highlights that the sanitized images produced by PrivacyLens remain useful and achieve comparable performance to their PII-invasive counterparts, showing that reducing PII collection is not inherently incompatible with supporting downstream CV/ML applications. The applications presented in this section were influenced by a user study we conducted to identify acceptable PII interventions relative to the context they are deployed in (see Appendix C). For example, *Face Swap* would not be an acceptable mode for fall detection in the bathroom.

With IRB approval, all three applications were evaluated through user studies with 10 participants, mainly undergraduate and graduate students (2 female, 8 male, mean age = 23.7, SD = 2.9). Participants were screened for whether they could comfortably perform the following tasks: squat exercises, interact with objects and place them in various locations, and routines involving standing, sitting, lying down, walking, and controlled falls to the ground. The participants were not compensated for the study, which took at most an hour in total to complete all three tasks. All the applications operate in real time at 8 FPS video, but the raw data was saved strictly for evaluation against the baseline PII-invasive counterparts. A lab server was used for various computational tasks, as detailed below.

5.1 Exercise Counting

Several CV-based systems identify exercises, count repetitions, and provide feedback to the user on their technique [4, 26, 27]. In this application, we prototype a squat counter in two ways to highlight the utility of different PrivacyLens modes in comparison to an off-the-shelf counter. The first way directly inputs *Face Swap* images into an existing off-the-shelf Posenet-based squat counter [8] running on the lab server. This demonstrates compatibility without modifying an existing application. The second way composites "Stick Figure" and "Thermal Subtraction" to create a PII-sanitized yet informative image that sends keypoints to the lab server for classification. This highlights PrivacyLens's ability to generate quality features on-device that can achieve comparable performance to

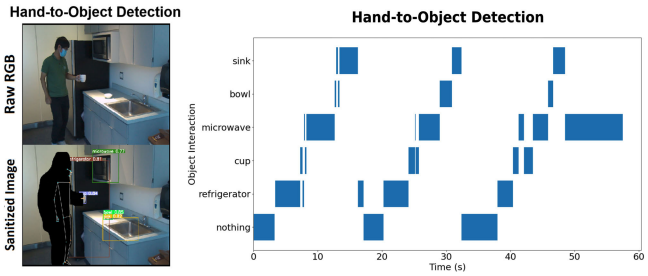


Figure 10: Result timeline of three sessions of the activity detection routine with the five objects.

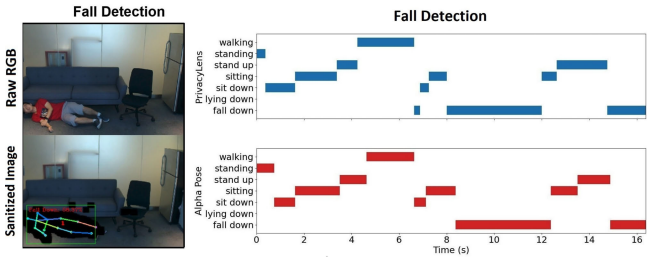


Figure 11: PrivacyLens-generated features (above) and AlphaPose-generated features (below) both using ST-GAN fall detection to track a person's activity timeline.

the off-the-shelf application while removing PII, namely from the face, skin and hair color, gender, and body shape.

Procedure. Participants were asked to perform ten squat repetitions, forming a session. They were then asked to repeat the session twice more, resulting in 30 squats per participant. Figure 9 shows an example of the squat counting application using PrivacyLens features (right) and depicts both the raw RGB and the *Face Swap* image (left). To evaluate this task, the number of counted squats for each method (i.e., *Face Swap*-Posenet-based and PrivacyLens-generated) is reported as a percentage of expected squats.

Results. The existing Posenet-based squat counter, which was provided with both raw and *Face Swap* images as input, correctly identified all squat repetitions at 100% accuracy across all participants for both sets of images. Our PrivacyLens-based approach, which generates its own keypoints to count squats, also had 100% accuracy across all participants. These results show that *Face Swap* remains compatible with existing downstream CV applications with matching performance using entirely on-device generated features.

5.2 Hand-to-Object Detection

Enabling computers to identify people's activities in their living spaces has been an active area of research in the health sensing domain, such as recognizing cooking and hygienic activities [74]. We developed a proof-of-concept hand-to-object detection application to track users in the kitchen. As shown in Figure 10, PrivacyLens first identifies the objects in the background using YOLO [44], applies Thermal Subtraction to remove the person, and then the object bounding boxes and person's keypoints are overlaid on the sanitized image. This allows the image to encode interaction with objects in the environment without including PII. On the lab server, an object interaction algorithm measures the intersection of hand keypoints and objects and records them as events.

Procedure. The 10 participants performed a routine with five objects in the environment: open the *refrigerator*, grab the *cup* from the refrigerator, place the cup in the *microwave*, grab the *bowl* from the counter, place the bowl in the *sink*. We asked each participant to perform this routine thrice. To evaluate this task, hand keypoint interactions with the object bounding boxes were detected as events. If interactions with all five objects were correctly identified, the session would be scored a 5 out of 5. The total number of detected events out of expected events was reported. An example object detection timeline can be seen in Figure 10.

Results. Across all participants, PrivacyLens correctly identified 143 of 150 hand-to-object events (95.3%). The seven missed interactions were caused by occlusion issues, such as when a participant's hand completely blocks the camera's view of the object preventing its proper detection. This performance is identical to the the version of this application that does not perform PII removal, as the object detection and keypoint detection are performed prior to PII sanitization and, thus, have equivalent sets of keypoints and identified bounding boxes. Though preliminary, these results show a promising first step towards privacy-preserved in-home activity recognition with cameras that do not collect PII but can still robustly encode useful information.

5.3 Fall Detection

The CDC report that one-third of adults over 65 have a fall resulting in injury each year [10]. While CV-based approaches are capable of detecting falls [15, 59] without requiring the user to continuously wear a monitoring device [15, 77], many elderly citizens and family members are reluctant to install cameras in their homes without privacy protections [14, 25, 59]. These CV approaches often have a similar pipeline: 1) identify the person in the image, 2) identify their keypoints, 3) track them across frames, and 4) predict the event.

In this application, we utilize an existing fall detector [30] that is pre-trained using the InVia Fall Dataset [77]. As implemented, the application first uses AlphaPose [47] to generate keypoints for each frame and, based on those keypoints, uses a trained ST-GCN model [72] to predict an event. This pipeline has significant computational requirements; AlphaPose, which needs access to the camera to generate keypoints at 10 FPS, requires an Nvidia 2080Ti GPU [46]. It is not economically feasible to deploy an expensive GPU to run the entire pipeline locally in each home. However, this application can be split such that PrivacyLens handles the privacy-sensitive data and generates keypoints for the resource-intensive ST-GCN model to predict events from the cloud.

During initial testing, we discovered that MediaPipe keypoints diverged considerably from AlphaPose keypoints during fast movements, such as falls. Using these keypoints resulted in poor detection performance. MoveNet [26], which omits certain keypoints, such as hands and face, in return for greater accuracy during fast movements, closely matched the AlphaPose ones used by the ST-GCN model. Thus, we use MoveNet to generate keypoints on device, overlay them on a Thermal Subtraction image, and send them to the lab server that runs the pretrained ST-GCN model. On the

Jetson Nano, MoveNet achieved 16 FPS but could not benchmark AlphaPose due to resource constraints.

Procedure. The 10 participants performed the following routine: stand, sit down on a chair, stand, sit on a couch, lie down on a couch, walk forward, and fall to the ground. They repeated this routine three times. Raw RGB frames were provided directly to the unmodified AlphaPose-ST-GCN fall detector pipeline, which outputs an activity label per frame as a "reference". PrivacyLens generates MoveNet keypoints per frame and is input to the same ST-GCN model for synchronized per-frame activity labels. To evaluate this task, for every activity event from the AlphaPose-ST-GCN pipeline, if the PrivacyLens-ST-GCN pipeline had an overlapping event at any point in time with the same label, it was considered a matching event. An example detected event, as well as the timeline of the routine from both approaches, can be found in Figure 11.

Results. Of the 198 reference events across all participants predicted by the full AlphaPose-ST-GCN pipeline, the PrivacyLens-ST-GCN version had 178 matching events (89.9%). The reference events that missed a matching PrivacyLens-ST-GCN event were caused by a significant divergence between the two sets of keypoints, such as when participants laid down facing away. These results show that, without any modification to a pre-trained ST-GCN model, on-device generated features can offer a compelling substitute and show promise for privacy-preserving CV-based fall detection.

5.4 PII Sanitization Results

Similar to Study 1 in the previous section, we evaluated the PII removal performance of the three applications. However, since the framerate was set to 8 FPS to support interactive applications—which generates ≈28k frames per hour—even if the video were cropped to only when the participant was actively performing the task, the resulting 14k frames (42k frames when sanitized three ways) would be a significantly burdensome annotation task. Thus, we randomly selected 20 frames from each participant for each application, resulting in 200 images per application and sanitized with PrivacyLens's hybrid *Thermal Subtraction*, YOLO, and Detectron2 methods, resulting in 600 images per application and 1800 images in total. The Exercise Counting application, where *Face Swap* was also utilized, added an additional 200 images to be annotated, resulting in a final set of 2,000 images annotated with ground truth by a total of 10 new annotators using the same procedure from Study 1. The full results can be found in Table 7 in Appendix B.4

Exercise Counting. In the 200 randomly selected frames, *Face Swap* removed 100% of facial PII in the frame but retained all other forms of PII. *Thermal Subtraction* only leaked skin color or hair color across a total of 7 frames, resulting in an APR rate of 96.5% (vs. 97.8% in Study 1). Apart from one "total failure", in which YOLO missed the person completely, resulting in all five forms of PII leaking, YOLO leaked skin color and hair color across a total of 115 frames, resulting in an APR rate of 42.5% (vs. 36.4% in Study 1). Detectron2 only leaked skin color or hair color across a total of 125 frames, resulting in an APR rate of 37.5%, similar to that in Study 1 (39.1%). For both RGB-only approaches, leakages were often due to the bounding box insufficiently covering the person, such as from hair, elbows, and legs poking out beyond the bounding box.

Hand-to-Object Detection. Compared to exercise counting, hand-to-object detection presented a more challenging PII removal scenario, where objects in the environment (such as a refrigerator door) could occlude the view of the person, potentially resulting in PII leakage. PrivacyLens only leaked skin and hair color, resulting in an APR of 92.0%. YOLO had a significant number of total failures due to occlusions, which caused no bounding box to be drawn and leak all five forms of PII, but otherwise achieved an APR of 42.0%. Detectron2 had no total failures, but still leaked skin color and hair color due to occlusions, which caused the bounding box to only cover a portion of the person, resulting in an APR of 38.0%.

Fall Detection. Fall detection introduced additional challenges, where participants could also appear in the image sideways (e.g., while lying down or falling to the ground). Like the previous applications, PrivacyLens leaked only skin and hair color, resulting in an APR of 95.5%. YOLO, which had a significant number of total failures due to participants appearing sideways, leaked all five forms of PII with an APR of 48.5%. Detectron2 had just 2 total failures, which leaked all five forms of PII, but otherwise only leaked skin and hair color, and performed particularly robustly, as there were no objects to create occlusions in this application, and was able to detect sideways persons, resulting in an APR of 63.5%.

These results show that while CV-only approaches are effective at supporting downstream activity monitoring applications, their relatively low sanitization rates and their propensity for total failures do not make them viable for privacy-sensitive applications.

5.5 Envisioned Future Use Cases & Feasibility

Privacy and economic factors prevent applications from operating entirely on the edge or in the cloud. For example, the aforementioned Nvidia 2080Ti required to support the full AlphaPose-ST-GCN pipeline originally retailed for USD $999 [63]—this cost does not include the remaining components to build a computer (e.g., CPU, RAM, disks) or the nontrivial electricity costs (Nvidia recommends a 650-watt power supply). Cloud resources leverage economies of scale to reduce per-device computational costs as they get deployed at a large scale (vs. the costs of implementing personal infrastructure that increases linearly per GPU). However, sending unsanitized image content to the cloud to perform all computational tasks can introduce privacy concerns.

PrivacyLens shows that useful yet privacy-sensitive applications can be distributed by generating PII-free features for applications with a significant computational component that must run on servers. Furthermore, all three applications demonstrate that even though PII has been removed, PrivacyLens's output remains *useful*: All three applications match the performance of their privacy-invasive counterparts. These results show that privacy does not have to be an all-or-nothing proposition; it is possible to leverage pre-built applications with enhanced privacy and develop compelling applications with high-quality features without capturing PII. In this manner, our approach reduces costs by enabling the opportunity to leverage cloud computing resources while still offering a baseline level of privacy. We hope these examples motivate further exploration of privacy-preserved CV/ML applications using on-device PII removal and features.

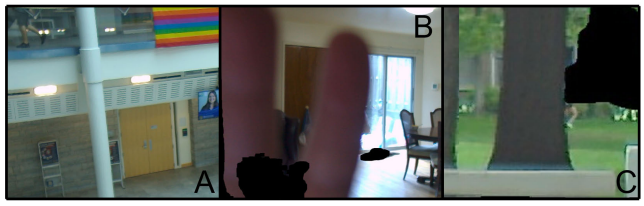


Figure 12: Common failure situations. (A) Thermal masking is prevented by glass blocking the transmission of infrared waves. (B) Imperfect camera alignment reveals fingers that are directly touching the RGB camera. (C) Pixel resolution masking limitations resulting in PII leakage.

6 DISCUSSION

This section discusses the limitations of the PrivacyLens prototype, alternative sanitization approaches, and future design concepts.

Limitations. Our real-world deployment study revealed limitations of the current implementation of the PrivacyLens prototype within certain situations. Illustrative examples are shown in Figure 12. One situation is when a person is partially in the frame, such as just their legs, but behind a glass structure. Since the glass blocks the person's blackbody wavelengths, the thermal camera cannot detect the person, and YOLO cannot detect just their legs, resulting in a PII leakage. More advanced RGB models that can detect partial persons should address this issue. Another set of "edge condition" issues is caused by the limited resolution and FPS used by the thermal camera in the prototype. Export restrictions by the International Traffic in Arms Regulations (ITAR) governing body make it relatively challenging to source high-resolution ($> 640 \times 512$) and high-FPS (> 9) thermal cameras, which is why PrivacyLens uses the readily available FLIR Lepton 3.5. Specifically, the 160×120 resolution is insufficient for a perfect 1-to-1 mapping to the RGB image (which caused fingers to be revealed in Figure 12) or to accurately resolve persons at long distances; at 50 m distance, a person is roughly one thermal pixel wide, which caused a runner's legs to be revealed in Figure 12. The slower FPS prevents frame-by-frame synchronization with the RGB camera. PrivacyLens partially addresses these sensor limitations by increasing the dilation factor of the thermal mask and ensuring a worst-case synchronization error of 16.67 ms. Despite these limitations, PrivacyLens achieved $> 99\%$ sanitization. Future work with higher-resolution thermal cameras and synchronized RGB and thermal cameras could improve the PII removal efficiency and reduce the need to dilate the thermal image. While the sensitivities of most radiometric thermal cameras are similar, however, their prices increase with resolution.

Another possible scenario that reduces the utility of a produced image but not its PII sanitization performance is when other heat elements are in the scene at the same temperature as the human body, such as an animal or hot surface. In this case, the pixels corresponding to these objects would also be removed from the image (e.g., a cat is removed from an image alongside the persons in the image). This would not affect the PII sanitization rate, but rather the non-PII content of the remaining image: Introducing additional heat elements would not "trick" the camera into unmasking the pixels that correspond to persons. Furthermore, since the thermal camera is radiometric and self-calibrating, the value of an individual pixel is absolute and not influenced by the temperature of the scene or environmental conditions (i.e., no min/max scaling), so an attacker can only cause a greater number of pixels to be removed from the image rather than inducing a PII leakage.

Finally, while the participants in the application study skew towards male participants, this resulted from who was available and willing to participate after recruiting through departmental email lists and direct individual recruitment. While performance was robust across genders, future work should explore systematic biases in PII sanitization approaches with a more diverse sample.

Alternative Sanitization Approaches. In the related work section, we highlighted alternative image sanitization approaches to protect bystander privacy that included deep learning (DL) models to blur and mask bystanders. For this, facial PII, blur, blocking, and inpainting approaches have been explored, but more recently, Deepfake approaches have maintained compatibility with systems that rely on facial detection for other tasks, such as pose detection, while removing facial PII. While these approaches demonstrate compelling results, they optimize for maximum performance and were not originally designed with the constraints of embedded hardware in mind. While future SBCs may be capable of employing those approaches for real-time image sanitization, the current capabilities of SBCs under USD $100 are insufficient to support them. However, there may be ways to support these approaches in the near future. For example, quantized DL models may approximate the sanitization performance while reducing the computational overhead to run on an SBC. Given the effectiveness of these alternative approaches using RGB alone, we expect the addition of thermal information to further enhance their capabilities. Additionally, similar to our fall detection application in the previous section, PrivacyLens could generate features (e.g., from Hasan et al. [34]: identifying persons, generating body-pose and facial expression features) and pass them to cloud resources for bystander classification. If those persons are classified as bystanders, the cloud computer can signal to PrivacyLens which persons should be removed while keeping the camera's owner intact. This ensures that raw images never leave the device while leveraging additional computing resources. As SBCs become more capable, we hope to see greater adoption of these approaches.

7 CONCLUSION

Without assurances of privacy, users are unlikely to accept cameras in their private spaces, such as bedrooms and bathrooms. To address this need for greater privacy, we explored the creation of a hybrid RGB and thermal camera to robustly detect and remove PII from images. To evaluate its effectiveness, PrivacyLens was deployed to capture images in three environments and results showed robust removal of PII with a 99.1% sanitization rate (with 100% sanitization of face, body shape, and gender), compared to 54.4% and 57.6% using RGB-only approaches. To showcase that PrivacyLens's PII-sanitized outputs remain useful, we evaluated three applications with 10 participants, finding the PII-sanitized application performance closely matched their privacy-invasive counterparts. Ultimately, PrivacyLens provides strong assurances that PII never leaves the device while maintaining compatibility with downstream applications, enabling a path for greater adoption of camera-based ubiquitous sensing applications.

REFERENCES

[1] 2024. *Raspberry Pi 3 Model B+*. Retrieved February 22, 2024 fromhttps://www.raspberrypi.com/products/raspberry-pi-3-model-b-plus/.

[2] Rawan Alharbi, Mariam Tolba, Lucia C. Petito, Josiah Hester, and Nabil Alshurafa. 2019. To Mask or Not to Mask? Balancing Privacy with Visual Confirmation Utility in Activity-Oriented Wearable Cameras. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 3, 3, Article 72 (sep 2019), 29 pages. https://doi.org/10.1145/3351230

[3] Aisha Fahad Alraeesi, Hanan Fekri Kharbash, Jawaher Saif Alghfeli, Shamma Sultan Alsaedi, and Munkhjargal Gochoo. 2021. Privacy-Preserved Social Distancing System Using Low-Resolution Thermal Sensors and Deep Learning. In *2021 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. 66–71. https://doi.org/10.1109/SMC52423.2021.9659292

[4] Ilktan Ar and Yusuf Sinan Akgul. 2014. A computerized recognition system for the home-based physiotherapy exercises using an RGBD camera. *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 22, 6 (2014), 1160–1171.

[5] Ashish Arora and Juan M Corchado. 2020. Face Detection and Recognition, Face Emotion Recognition Through NVIDIA Jetson Nano. In *Ambient Intelligence-Software and Applications: 11th International Symposium on Ambient Intelligence*, Vol. 1239. Springer Nature, 177.

[6] Jeonghyun Baek, Sungjun Hong, Jisu Kim, and Euntai Kim. 2017. Efficient pedestrian detection at nighttime using a thermal camera. *Sensors* 17, 8 (2017), 1850.

[7] Alison Bell. 2013. Randomized or fixed order for studies of behavioral syndromes? *Behavioral Ecology* 24, 1 (2013), 16–20.

[8] bipinkc19. 2020. *Using pose-net to count the number of squats done.* Retrieved September 8, 2021 from https://github.com/bipinkc19/squat-counter.

[9] Kelly E Caine, Arthur D Fisk, and Wendy A Rogers. 2006. Benefits and privacy concerns of a home equipped with a visual sensing system: A perspective from older adults. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. SAGE Publications Sage CA: Los Angeles, CA, 180–184.

[10] CDC. 2021. *"Keep on Your Feet—Preventing Older Adult Falls"*. Retrieved September 8, 2021 from https://www.cdc.gov/injury/features/older-adult-falls/index.html.

[11] Matthew Charlton, Sophie A. Stanley, ZoÃ« Whitman, Victoria Wenn, Timothy J. Coats, Mark Sims, and Jonathan P. Thompson. 2020. The effect of constitutive pigmentation on the measured emissivity of human skin. *PLOS ONE* 15, 11 (11 2020), 1–9. https://doi.org/10.1371/journal.pone.0241843

[12] Zhilu Chen and Xinming Huang. 2019. Pedestrian detection for autonomous vehicle using multi-spectral cameras. *IEEE Transactions on Intelligent Vehicles* 4, 2 (2019), 211–219.

[13] Myeon-gyun Cho. 2019. A Study on the Obstacle Recognition for Autonomous Driving RC Car Using LiDAR and Thermal Infrared Camera. In *2019 Eleventh International Conference on Ubiquitous and Future Networks (ICUFN)*. 544–546. https://doi.org/10.1109/ICUFN.2019.8806152

[14] Eun Kyoung Choe, Sunny Consolvo, Jaeyeon Jung, Beverly Harrison, and Julie A. Kientz. 2011. Living in a Glass House: A Survey of Private Moments in the Home. In *Proceedings of the 13th International Conference on Ubiquitous Computing* (Beijing, China) *(UbiComp '11)*. Association for Computing Machinery, New York, NY, USA, 41–44. https://doi.org/10.1145/2030112.2030118

[15] Pongsatorn Chutimawattanakul and Pranchalee Samanpiboon. 2022. Fall Detection for The Elderly using YOLOv4 and LSTM. In *2022 19th International Conference on Electrical Engineering/Electronics, Computer, Telecommunications and Information Technology (ECTI-CON)*. 1–5. https://doi.org/10.1109/ECTI-CON54298.2022.9795534

[16] Umur A. Ciftci, Gokturk Yuksek, and Ilke Demir. 2023. My Face My Choice: Privacy Enhancing Deepfakes for Social Media Anonymization. In *IEEE/CVF Winter Conference on Applications of Computer Vision, WACV 2023, Waikoloa, HI, USA, January 2-7, 2023*. IEEE, 1369–1379. https://doi.org/10.1109/WACV56688.2023.00142

[17] Mitchell Clark. 2022. *"Google, like Amazon, may let police see your video without a warrant"*. Retrieved July 27, 2022 from https://www.theverge.com/2022/7/26/23279562/arlo-apple-wyze-eufy-google-ring-security-camera-foortage-warrant.

[18] Stephanie Condon. 2019. *"Alexa's latest creepy move: recording a couple's private conversation and sharing it"*. Retrieved September 8, 2021 from https://www.zdnet.com/article/alexas-latest-creepy-move-recording-a-couples-private-conversation-and-sharing-it/.

[19] Rosario Delgado and Xavier-Andoni Tibau. 2019. Why Cohen's Kappa should be avoided as performance measure in classification. *PloS one* 14, 9 (2019), e0222916.

[20] Emon Dey and Nirmalya Roy. 2020. OMAD: On-device Mental Anomaly Detection for Substance and Non-Substance Users. In *2020 IEEE 20th International Conference on Bioinformatics and Bioengineering (BIBE)*. 466–471. https://doi.org/10.1109/BIBE50027.2020.00081

[21] Mariella Dimiccoli, Juan Marín, and Edison Thomaz. 2018. Mitigating Bystander Privacy Concerns in Egocentric Activity Recognition with Deep Learning and Intentional Image Degradation. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 132 (jan 2018), 18 pages. https://doi.org/10.1145/3161190

[22] Piotr Dollar, Christian Wojek, Bernt Schiele, and Pietro Perona. 2009. Pedestrian detection: A benchmark. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*. 304–311. https://doi.org/10.1109/CVPR.2009.5206631

[23] Aleksandr Fedorov, Kseniia Nikolskaia, Sergey Ivanov, Vladimir Shepelev, and Alexey Minbaleev. 2019. Traffic flow estimation with data from a video surveillance camera. *Journal of Big Data* 6, 1 (2019), 1–15.

[24] Sarah Frier. 2019. *"Facebook Has Been Paying Contractors to Transcribe Users' Facebook Messenger Voice Chats"*. Retrieved July 27, 2022 from https://time.com/5651395/facebook-contractors-transcribe-conversations-audio-files/.

[25] Roschelle Lynnette Fritz. 2015. *The influence of culture on older adults' adoption of smart home monitoring: A qualitative descriptive study.* Washington State University.

[26] Google. 2021. *"Next-Generation Pose Detection with MoveNet and TensorFlow.js"*. Retrieved September 8, 2021 from https://blog.tensorflow.org/2021/05/next-generation-pose-detection-with-movenet-and-tensorflowjs.html.

[27] Google. 2021. *"Pose classification (extended).ipynb – Google Colab Fitness Demo"*. Retrieved September 8, 2021 from https://colab.research.google.com/drive/19txHpN8exWhstO6WVkfmYYVC6uug_oVR.

[28] Google for Developers. 2023. *Pose detection.* https://developers.google.com/ml-kit/vision/pose-detection Accessed on: August 14, 2023.

[29] Erin Griffiths, Salah Assana, and Kamin Whitehouse. 2018. Privacy-Preserving Image Processing with Binocular Thermal Cameras. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* 1, 4, Article 133 (jan 2018), 25 pages. https://doi.org/10.1145/3161198

[30] Garun Gun. 2021. *Human Falling Detection and Tracking.* Retrieved September 8, 2021 from https://github.com/GajuuzZ/Human-Falling-Detect-Tracks.

[31] Eileen Guo. 2022. *"A Roomba recorded a woman on the toilet. How did screenshots end up on Facebook?"*. Retrieved January 2, 2023 from https://www.technologyreview.com/2022/12/19/1065306/roomba-irobot-robot-vacuums-artificial-intelligence-training-data-privacy/.

[32] Youngjib Ham, Kevin K Han, Jacob J Lin, and Mani Golparvar-Fard. 2016. Visual monitoring of civil infrastructure systems via camera-equipped Unmanned Aerial Vehicles (UAVs): a review of related works. *Visualization in Engineering* 4, 1 (2016), 1–8.

[33] Koji Harada, Ismail Arai, Shigeru Kashihara, and Kazutoshi Fujikawa. 2020. A Performance Investigation of Thermal Infrared Camera and Optical Camera for Searching Victims with an Unmanned Aerial Vehicle: Poster Abstract. In *Proceedings of the 18th Conference on Embedded Networked Sensor Systems* (Virtual Event, Japan) *(SenSys '20)*. Association for Computing Machinery, New York, NY, USA, 647–648. https://doi.org/10.1145/3384419.3430459

[34] Rakibul Hasan, David Crandall, Mario Fritz, and Apu Kapadia. 2020. Automatically Detecting Bystanders in Photos to Reduce Privacy Risks. In *2020 IEEE Symposium on Security and Privacy (SP)*. 318–335. https://doi.org/10.1109/SP40000.2020.00097

[35] Sean Hollister. 2023. *"Anker finally comes clean about its Eufy security cameras"*. Retrieved January 31, 2023 from https://www.theverge.com/23573362/anker-eufy-security-camera-answers-encryption.

[36] Roberto Hoyle, Robert Templeman, Steven Armes, Denise Anthony, David Crandall, and Apu Kapadia. 2014. Privacy Behaviors of Lifeloggers Using Wearable Cameras. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing* (Seattle, Washington) *(UbiComp '14)*. Association for Computing Machinery, New York, NY, USA, 571–582. https://doi.org/10.1145/2632048.2632079

[37] Rasheed Hussain and Sherali Zeadally. 2019. Autonomous Cars: Research Results, Issues, and Future Challenges. *IEEE Communications Surveys Tutorials* 21, 2 (2019), 1275–1313. https://doi.org/10.1109/COMST.2018.2869360

[38] Soonmin Hwang, Jaesik Park, Namil Kim, Yukyung Choi, and In So Kweon. 2015. Multispectral Pedestrian Detection: Benchmark Dataset and Baselines. In *Proceedings of IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*.

[39] IAPP. 2019. *"The California Privacy Rights Act of 2020"*. Retrieved September 8, 2021 from https://iapp.org/resources/article/the-california-privacy-rights-act-of-2020/.

[40] ICI. 2019. *"Thermal Camera Export Restrictions"*. Retrieved September 8, 2021 from https://infraredcameras.com/thermal-camera-export-restrictions/.

[41] Umar Iqbal, Pouneh Nikkhah Bahrami, Rahmadi Trimananda, Hao Cui, Alexander Gamero-Garrido, Daniel Dubois, David Choffnes, Athina Markopoulou, Franziska Roesner, and Zubair Shafiq. 2022. Your Echos are Heard: Tracking, Profiling, and Ad Targeting in the Amazon Smart Speaker Ecosystem. https://doi.org/10.48550/ARXIV.2204.10920

[42] Yasha Iravantchi, Karan Ahuja, Mayank Goel, Chris Harrison, and Alanson Sample. 2021. PrivacyMic: Utilizing Inaudible Frequencies for Privacy Preserving Daily Activity Recognition. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems* (Yokohama, Japan) *(CHI '21)*. Association for Computing Machinery, New York, NY, USA, Article 198, 13 pages. https://doi.org/10.1145/3411764.3445169

[43] Cynthia S Jacelon and Allen Hanson. 2013. Older adults' participation in the development of smart environments: An integrated review of the literature.

Geriatric Nursing 34, 2 (2013), 116–121.
[44] Glenn Jocher, Ayush Chaurasia, Alex Stoken, Jirka Borovec, NanoCode012, Yonghye Kwon, TaoXie, Jiacong Fang, imyhxy, Kalen Michael, Lorna, Abhiram V, Diego Montes, Jebastin Nadar, Laughing, tkianai, yxNONG, Piotr Skalski, Zhiqiang Wang, Adam Hogan, Cristi Fati, Lorenzo Mammana, AlexWang1900, Deep Patel, Ding Yiwei, Felix You, Jan Hajek, Laurentiu Diaconu, and Mai Thanh Minh. 2022. *ultralytics/yolov5: v6.1 - TensorRT, TensorFlow Edge TPU and OpenVINO Export and Inference.* https://doi.org/10.5281/zenodo.6222936
[45] Konstantinos Karachalios. 2014. *"The Wicked Problem of Privacy and Security in the Opportunity Driven Connected Person World".* Retrieved September 17, 2020 from https://connect-world.com/the-wicked-problem-of-privacy-and-security-in-the-opportunity-driven-connected-person-world/.
[46] Jiefeng Li, Chao Xu, Zhicun Chen, Siyuan Bian, Lixin Yang, and Cewu Lu. 2019. *AlphaPose Github Issue Tracker.* Retrieved September 8, 2021 from https://github.com/MVIG-SJTU/AlphaPose/issues/364.
[47] Jiefeng Li, Chao Xu, Zhicun Chen, Siyuan Bian, Lixin Yang, and Cewu Lu. 2021. Hybrik: A hybrid analytical-neural inverse kinematics solution for 3d human pose and shape estimation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 3383–3393.
[48] Yifang Li, Nishant Vishwamitra, Bart P. Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Blur vs. Block: Investigating the Effectiveness of Privacy-Enhancing Obfuscation for Images. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW).* 1343–1351. https://doi.org/10.1109/CVPRW.2017.176
[49] Karola Marky, Nina Gerber, Michelle Gabriela Pelzer, Mohamed Khamis, and Max Mühlhäuser. 2022. "You offer privacy like you offer tea": Investigating mechanisms for improving guest privacy in IoT-equipped households. (2022). https://doi.org/10.56553/popets-2022-0115
[50] Natalia Drozdiak Matt Day, Giles Turner. 2019. *"Thousands of Amazon Workers Listen to Alexa Users' Conversations".* Retrieved July 27, 2022 from https://time.com/5568815/amazon-workers-listen-to-alexa/.
[51] Siavash Mazdeyasna, Pejman Ghassemi, and Quanzeng Wang. 2023. Best Practices for Body Temperature Measurement with Infrared Thermography: External Factors Affecting Accuracy. *Sensors* 23, 18 (2023). https://doi.org/10.3390/s23188011
[52] Sachit Menon, Alex Damian, McCourt Hu, Nikhil Ravi, and Cynthia Rudin. 2020. PULSE: Self-Supervised Photo Upsampling via Latent Space Exploration of Generative Models. In *The IEEE Conference on Computer Vision and Pattern Recognition (CVPR).*
[53] Merriam-Webster. 2022. *"Right of privacy definition".* Retrieved August 15, 2022 from https://www.merriam-webster.com/legal/rightofprivacy.
[54] HM Mohan, S Anitha, Rifai Chai, and Sai Ho Ling. 2021. Edge Artificial Intelligence: Real-Time Noninvasive Technique for Vital Signs of Myocardial Infarction Recognition Using Jetson Nano. *Advances in Human-Computer Interaction* 2021 (2021).
[55] Vu-Anh-Quang Nguyen, Jongoh Park, Kyeongjin Joo, Thi Tra Vinh Tran, Trung Tin Tran, and Joonhyeon Choi. 2020. *Human Face Recognition and Temperature Measurement Based on Deep Learning for Covid-19 Quarantine Checkpoint.* Association for Computing Machinery, New York, NY, USA. https://doi.org/10.1145/3440749.3442654
[56] Helen Nissenbaum. 2004. Privacy as contextual integrity. *Wash. L. Rev.* 79 (2004), 119.
[57] NIST. 2021. *"personally identifiable information (PII)".* Retrieved September 8, 2021 from https://csrc.nist.gov/glossary/term/personally_identifiable_information.
[58] Nvidia. 2021. *"Nvidia Jetson Nano Product Specifications".* Retrieved September 8, 2021 from https://developer.nvidia.com/embedded/jetson-nano-developer-kit.
[59] Guendoul Oumaima, Ait Abdelali Hamd, Tabii Youness, Oulad Haj Thami Rachid, and Bourja Omar. 2021. Vision-based fall detection and prevention for the elderly people: A review & ongoing research. In *2021 Fifth International Conference On Intelligent Computing in Data Sciences (ICDS).* 1–6. https://doi.org/10.1109/ICDS53782.2021.9626736
[60] P3International. 2021. *"Kill A Watt".* Retrieved September 8, 2021 from http://www.p3international.com/products/p4400.html.
[61] José Ramón Padilla-López, Alexandros Andre Chaaraoui, and Francisco Flórez-Revuelta. 2015. Visual privacy protection methods: A survey. *Expert Systems with Applications* 42, 9 (2015), 4177–4195. https://doi.org/10.1016/j.eswa.2015.01.041
[62] Ramendra Pathak and Yaduvir Singh. 2020. Real Time Baby Facial Expression Recognition Using Deep Learning and IoT Edge Computing. In *2020 5th International Conference on Computing, Communication and Security (ICCCS).* 1–6. https://doi.org/10.1109/ICCCS49678.2020.9277428
[63] PCMag. 2024. *NVIDIA GeForce RTX 2080 Ti Founders Edition Review.* Retrieved February 22, 2024 fromhttps://www.pcmag.com/reviews/nvidia-geforce-rtx-2080-ti-founders-edition.
[64] pigpio. 2021. *"Pigpio Raspberry Pi GPIO Library".* Retrieved September 8, 2021 from https://abyz.me.uk/rpi/pigpio/.
[65] PrivacySOS. 2012. *"MBTA gives embarrassing video of woman's fall to the press".* Retrieved January 2, 2023 from https://privacysos.org/blog/today-in-wtf-mbta-gives-embarrassing-video-of-womans-fall-to-the-press/.
[66] Emma Roth. 2022. *"Apple and Meta shared data with hackers pretending to be law enforcement officials".* Retrieved July 27, 2022 from https://www.theverge.com/2022/3/30/23003600/apple-meta-shared-data-hackers-pretending-law-enforcement-officials.
[67] Qianru Sun, Liqian Ma, Seong Joon Oh, Luc Van Gool, Bernt Schiele, and Mario Fritz. 2018. Natural and Effective Obfuscation by Head Inpainting. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition.* 5050–5059. https://doi.org/10.1109/CVPR.2018.00530
[68] Zhiwei Wang, Yihui Yan, Yueli Yan, Huangxun Chen, and Zhice Yang. 2022. CamShield: Securing Smart Cameras through Physical Replication and Isolation. In *31st USENIX Security Symposium (USENIX Security 22).* USENIX Association, Boston, MA, 3467–3484. https://www.usenix.org/conference/usenixsecurity22/presentation/wang-zhiwei
[69] Yunqian Wen, Bo Liu, Ming Ding, Rong Xie, and Li Song. 2022. IdentityDP: Differential Private Identification Protection for Face Images. *Neurocomput.* 501, C (aug 2022), 197–211. https://doi.org/10.1016/j.neucom.2022.06.039
[70] Hao-Yu Wu, Michael Rubinstein, Eugene Shih, John Guttag, Frédo Durand, and William T. Freeman. 2012. Eulerian Video Magnification for Revealing Subtle Changes in the World. *ACM Transactions on Graphics (Proc. SIGGRAPH 2012)* 31, 4 (2012).
[71] Yuxin Wu, Alexander Kirillov, Francisco Massa, Wan-Yen Lo, and Ross Girshick. 2019. *Detectron2.* Retrieved September 8, 2021 from https://github.com/facebookresearch/detectron2.
[72] Sijie Yan and Dahua Lin. 2019. *MMSkeleton.* Retrieved September 8, 2021 from https://github.com/open-mmlab/mmskeleton.
[73] Liming Zhai, Qing Guo, Xiaofei Xie, Lei Ma, Yi Estelle Wang, and Yang Liu. 2022. A3GAN: Attribute-Aware Anonymization Networks for Face De-Identification. In *Proceedings of the 30th ACM International Conference on Multimedia* (Lisboa, Portugal) *(MM '22).* Association for Computing Machinery, New York, NY, USA, 5303–5313. https://doi.org/10.1145/3503161.3547757
[74] Chenyang Zhang and Yingli Tian. 2012. RGB-D camera-based daily living activity recognition. *Journal of computer vision and image processing* 2, 4 (2012), 12.
[75] Yupeng Zhang, Yuheng Lu, Hajime Nagahara, and Rin-ichiro Taniguchi. 2014. Anonymous Camera for Privacy Protection. In *Proceedings of the 2014 22nd International Conference on Pattern Recognition (ICPR '14).* IEEE Computer Society, USA, 4170–4175. https://doi.org/10.1109/ICPR.2014.715
[76] Zhenxing Zhou, Yisiang Neo, King-Shan Lui, Vincent W.L. Tam, Edmund Y. Lam, and Ngai Wong. 2020. A Portable Hong Kong Sign Language Translation Platform with Deep Learning and Jetson Nano. In *The 22nd International ACM SIGACCESS Conference on Computers and Accessibility* (Virtual Event, Greece) *(ASSETS '20).* Association for Computing Machinery, New York, NY, USA, Article 89, 4 pages. https://doi.org/10.1145/3373625.3418046
[77] ImVia Lab UFR INSTITUTS ÉCOLES. 2021. *Fall Detection Dataset.* Retrieved September 8, 2021 from https://imvia.u-bourgogne.fr/en/database/fall-detection-dataset-2.html.

APPENDIX

A PILOT STUDY

As referenced in Section 3, we conducted a pilot study to inform the design of the PrivacyLens prototype, which was formally evaluated in real-world settings in Section 4. This section details those pilot studies performed on the existing Korea Advanced Institute of Science and Technology (KAIST) Advanced Driver Assistance Systems (ADAS) RGB and thermal dataset [38] to develop approaches that can robustly remove persons from images. We additionally evaluated those approaches to determine their efficiency on various embedded devices, as described in Section 3 and further detailed in Appendix B.1.

A.1 Approach

The addition of thermal imaging to RGB cameras has shown significant improvement in ADAS person-detection tasks over RGB-only methods [37]; as an additional sensing approach, it may also enhance the robustness of finding and removing entire persons from images and, thereby, PII. To present RGB-only baselines, we utilized two RGB-only approaches: YOLOv5 [44], a lightweight model that can run in real time on an embedded device, and Detectron2 [71], a state-of-the-art Region-based Convolutional Neural

Network (R-CNN) approach using desktop-class GPUs. We then evaluated a thermal-only and an RGB and thermal hybrid approach to determine how effectively thermal information can aid in person-removal tasks. Our approach, described in Algorithm 1, uses the thermal image to create a segmentation mask that can "subtract" the pixels from the RGB image, which can be performed efficiently on a GPU. For a direct comparison to the RGB-only approaches, a bounding box was calculated from the span of the thermal mask, and the entirety of that bounding box was removed. The hybrid approach pairs YOLO with thermal subtraction by finding the union of the two bounding boxes.

Algorithm 1 PrivacyLens's thermal subtraction algorithm, where a subtraction mask is made from pixels that are within a range for typical human skin temperature and is used to perform a bitwise operation on the RGB image to remove those specified pixels.

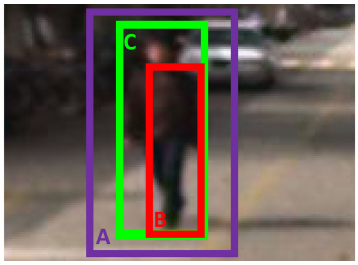
Inputs:
`aligned_thermal_im`, `aligned_rgb_im`
Create binary mask (`lower_mask`) from thermal image lower temperature threshold
Create binary mask (`upper_mask`) from thermal image upper temperature threshold
Invert (`Bitwise_NOT`) `upper_mask`
Create binary mask (`final_mask`) by `Bitwise_AND(lower_mask, upper_mask)`
Invert (`Bitwise_NOT`) `final_mask`
Erode `final_mask`
Dilate `final_mask`
Convert `final_mask` to 8-bit RGB image
`Bitwise_AND(final_mask, aligned_rgb_im)` and return `thermal_sub_im`
Outputs:
`thermal_sub_im`

A.2 Procedure

To evaluate these four approaches, we used the KAIST Multi-spectral Advanced Driver Assistance Systems (ADAS) dataset [38], as there are very few publicly available RGB and thermal datasets that contain a wide variety of persons with curated human-labeled annotations outside of the ADAS community. While the environment in the ADAS dataset differs from typical human-centric environments, it is a very large dataset of 95k RGB-thermal image pairs that contains individual bounding boxes for persons in a variety of different poses (e.g., sitting, standing, walking), orientations (e.g., facing directly, sideways, and away from the camera), and clothing styles that cover exposed skin (e.g., hats, long sleeves, pants). This variety is similar to the poses, orientations, and clothing encountered in human-centric environments, which we formally validate in Section 4. Thus, using the ADAS dataset should be sufficiently representative for a pilot study to determine the relative value of thermal information for person removal in images. To prepare the dataset for our evaluation, we used the training half of the dataset, which has 50,200 total images, of which 41,500 contain persons. There may be more than one person in each image, resulting in 67,991 annotated bounding boxes corresponding to a person.

ADAS traditionally uses the Intersection over Union (IoU) metric, as shown in Figure 14, to evaluate pedestrian detection success rates, where > 50% IoU is tallied as a correct detection [22]. However, IoU is not an ideal metric for privacy; since IoU optimizes for box alignment rather than coverage, it does not assess whether an entire person was removed from an image and thus sanitized of PII. For example, in Figure 14, there are three bounding boxes: Box A is a predicted bounding box larger than the ground truth Box C; Box B is a predicted bounding box smaller than the ground truth Box

Figure 13: A sample image pair from the KAIST ADAS dataset with the RGB image (left) and the thermal image (right).



$$IoU_{AC} = \frac{A \cap C}{A \cup C} = \frac{1}{2} = .50$$

$$IoU_{BC} = \frac{B \cap C}{B \cup C} = \frac{.5}{1} = .50$$

Figure 14: An illustration where two predicted bounding boxes, A and B, can have the same IoU score relative to the ground truth annotated bounding box, C. If bounding box A is removed from the image, the person is entirely removed. However, if bounding box B is removed, identifiers such as skin and hair color remain in the image.

C. Box A, which has twice the area of ground truth but completely covers it has an equivalent IoU to Box B, which is half the area of ground truth but only partially covers the person. In this situation, Box A correctly removes the person entirely, but Box B exposes PII, demonstrating that IoU is not an adequate metric to determine the entire person's removal. Thus, we define a metric that quantifies the success rate based on how much of the ground truth bounding box is removed. First, the percentage of pixels removed is calculated for each annotated bounding box. Using a pixel-percentage threshold, that removal is counted as either a "success" or a "failure". The total proportion of successes is reported as a success rate relative to a set pixel-percentage threshold. For example, if the pixel-percentage threshold is set to 95%, and 99/100 bounding boxes had ≥95% of pixels removed, the success rate for that threshold would be 0.99. A privacy-success curve is generated as the threshold is increased from 1% to 100% (increments of 1%) for each person-removal approach, representing the relative robustness of an approach to an increasing privacy standard.

A.3 Results

For each approach, the privacy-success curve was computed by increasing the threshold starting from 1% to 100%, with a step size of 1%, and can be found in Figure 15. There is no value computed for a 0% threshold. While both RGB-only approaches provide robust person detection, they struggle to correctly identify the entire bounds of a person, especially when a person has their limbs extended or is

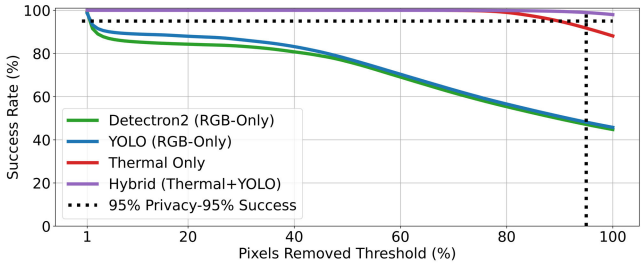


Figure 15: The privacy-success curves for the four approaches. The top right corner represents the ideal case; 100% of images have 100% of pixels removed. The bottom left corner represents the worst case, where 0% of images had at least 1% of pixels removed (as the threshold starts at 1%, there is no 0% pixel removal threshold). Based on these results, hybrid models are needed to surpass >95% success with a >95% pixel removal threshold.

not facing the camera. YOLO often had more generously predicted bounding boxes, thus slightly outperforming Detectron2 in this task. Overall, both RGB-only approaches are insufficient, as less than 50% of bounding boxes were successfully sanitized with a 95% pixel-level privacy threshold.

The thermal-only approach significantly improves over RGB-only but only achieves 91.7% performance at a 95% pixel-level privacy threshold. We observed that thermal-only has difficulties when extremities are lower than the threshold (e.g., cold hands) or are covered (e.g., with a hat). We found the hybrid approach addresses the weaknesses of each standalone approach: YOLO can extend the bounding box to cold extremities, and the thermal component can capture situations where persons have awkward postures or face away from the camera, a challenge for RGB-only methods. The hybrid approach achieves a 98.9% success rate at the 95% pixel-level privacy threshold, demonstrating the possibility of robust detection and removal of people from images.

B ADDITIONAL EVALUATION DETAILS

In this section, we include greater details for the efficiency benchmarks referenced in Section 3, annotator statistics for the deployment evaluation in Section 4, further details regarding the more aggressive RGB-only evaluation in Section 4, and the full PII sanitization results of the applications presented in Section 5.

B.1 Efficiency Benchmarks

Per the efficiency evaluation detailed in Section 3, Table 5 shows the full results for the thermal subtraction and facial landmark benchmarks. Overall, these tasks were especially challenging for the embedded devices' CPUs. Neither could exceed 13 FPS in both tasks, making them insufficient for real-time sanitization of images while leaving little headroom for other CPU-bound tasks. However, the Jetson Nano's GPU achieved up to 260 FPS on the thermal subtraction benchmark and up to 50 FPS on the facial landmark benchmark with significantly higher per-watt efficiency than the CPU alone. Offloading these tasks to the GPU frees CPU resources to perform additional operations. Additionally, the embedded GPU supports certain lightweight ML-based tasks, which are helpful for other PII removal tasks or generating features as part of a more extensive ML-driven application, as was shown in Section 5.

Table 5: The results of our thermal subtraction (top) and facial landmark (bottom) benchmarks, with the best performance in bold. The Jetson Nano GPU has the highest FPS/W for both tasks, making it ideal for mobile applications.

Thermal Subtraction Benchmark	Watts	FPS	FPS/W
Raspberry Pi 3	2.5	16.9	6.8
Jetson CPU (Max)	1.6	78.1	48.8
Jetson CPU (5W)	**0.9**	45.3	50.3
Jetson GPU (Max)	2.5	259.7	103.8
Jetson GPU (5W)	2.0	221.3	**110.7**
Intel 9900K	38.0	434.0	11.4
Titan RTX	205.0	**18883.5**	92.1
Facial Landmark Benchmark	**Watts**	**FPS**	**FPS/W**
Raspberry Pi 3	4.6	4.8	1.0
Jetson CPU (Max)	2.7	12.4	4.6
Jetson CPU (5W)	**0.6**	3.1	5.2
Jetson GPU (Max)	2.9	50.0	17.2
Jetson GPU (5W)	0.9	17.5	**19.4**
Intel 9900K	95.8	226.3	2.4
Titan RTX	59.0	**229.3**	3.9

B.2 Across-Study Annotator Statistics

Cohen's kappa is a traditional statistic to model annotator reliability [19]. We computed kappa to compare the reliability between annotators from Study 1 to Study 2 in Section 4 for all images sanitized by the same PII sanitization methods. For PrivacyLens, YOLO, and Detectron2, we found a score of 0.577, 0.648, and 0.641, respectively. While this suggests only moderate agreement among annotators from Study 1 to Study 2, Cohen's kappa considers class imbalance (e.g., the proportion of images annotators agree contain PII versus images annotators agree contain no PII) and thus can lower the score even if there is significant agreement among annotators [19]. In the PrivacyLens case, the ratio of images that annotators agreed contained PII to images that annotators agreed contained no PII was 0.0094, effectively further reducing the kappa score despite annotators agreeing on whether the image contained PII in 98.7% of images. However, in the YOLO and Detectron2 cases with higher kappa values, the ratios of images were much closer to 1 (1.253 and 1.084) despite having a significantly higher prevalence in differing PII content ratings (18.0% and 18.5%) across images.

Thus, we directly analyzed annotator disagreements by manually identifying individual images where the annotator in Study 1 flagged an image as having PII and another in Study 2 flagged it as not having PII (or vice versa, if that case occurred). The authors discussed potential reasons and identified that disagreement between the annotator from Study 1 and Study 2 on whether the image contained PII was highly dependent on how the sanitization system leaked PII, suggesting that Cohen's kappa presents an unideal statistic to determine the reliability of annotators across the two studies. For the PrivacyLens approach, 58.9% of the images flagged as containing at least one form of PII by an annotator in the first study had a disagreement, whereas the annotator in the

second study flagged it as containing no PII at all. For YOLO and Detectron2, the disagreement percentage was 28.3% and 30.3%, respectively. We attribute the difference in disagreement percentages to how the approaches leak PII. The RGB approaches have more obvious PII leakage behavior; their high rate of "total failure" flagged images often corresponded to whole persons being missed. With PrivacyLens, the leakage was often more subtle, such as unmasked pixels located on the edge of where the thermal subtraction mask fell short. Thus, a high proportion of the images flagged as having PII were the "on the fence" type, which is why many were not flagged as having PII by a different annotator, leading to a higher disagreement percentage. For images flagged as containing no PII, the disagreement among all approaches (PrivacyLens, YOLO, Detectron2) was 0%, as they were all confirmed to have no PII. This suggests our annotators were very careful when actively flagging an image as having no PII.

B.3 More Aggressive RGB-only PII Removal

In Section 4, we conducted an additional evaluation where the RGB-only methods were made to remove PII from images more aggressively compared to their default settings. The predicted bounding boxes for both YOLO and Detectron2 were increased in area by 14%, thus removing a greater portion of the image for each predicted bounding box. Overall, sanitization rates for all environments and all forms of PII improved, most notably hair and skin color, which were often leaked when hair or exposed extremities (e.g., fingers, feet) extended past the predicted bounding boxes. However, increasing the predicted bounding box size still did not remedy situations where no bounding box was predicted, leaving that portion of an image completely unsanitized. Thus, these results further demonstrated that information provided by the thermal camera is critical for robust PII removal across a multitude of situations.

Table 6: The aggressive RGB-only PII removal success rates.

Aggressive YOLO (RGB Only, Embedded)	**Atrium**	**Home**	**Park**	**All Env.**
Face	82.8%	79.9%	67.2%	78.7%
Skin Color	74.8%	67.0%	56.0%	66.9%
Hair Color	76.8%	72.4%	62.0%	71.8%
Gender	79.9%	75.8%	62.9%	74.8%
Body Shape	79.5%	75.4%	63.2%	74.4%
All PII Removal w/ Ground Truth	**73.4%**	**65.8%**	**55.5%**	**65.8%**
Aggressive Detectron2 (RGB Only, Desktop)	**Atrium**	**Home**	**Park**	**All Env.**
Face	97.2%	95.5%	92.9%	95.5%
Skin Color	86.6%	89.5%	86.4%	88.6%
Hair Color	88.9%	93.4%	89.3%	92.0%
Gender	95.4%	93.9%	91.1%	93.8%
Body Shape	94.0%	94.6%	91.1%	94.0%
All PII Removal w/ Ground Truth	**85.7%**	**88.5%**	**84.0%**	**87.4%**

B.4 PII Removal in Applications

In Section 5, we evaluated the PII removal rates for the exercise counting, hand-to-object activity recognition, and fall detection applications. Overall, PrivacyLens's hybrid approach outperformed the two RGB-only approaches, similar to the results presented in our deployment evaluation in Section 4. However, one particular advantage PrivacyLens had in these applications was when the participants were in awkward poses (e.g., while laying on the couch or falling down) or when an object came between the participant and the camera (e.g., refrigerator and microwave doors). In these situations, the RGB-only approaches had difficulty drawing bounding boxes appropriately. In the case of Detectron2, the bounding boxes may have exposed additional skin and hair, but rarely the face. Additionally, since Detectron2 could draw bounding boxes when a person appeared sideways in an image (such as when lying down), it missed drawing a bounding box in only two images. For YOLO, in addition to exposing skin and hair in a similar manner to Detectron2, it struggled to identify persons in images when they were sideways or had an object occlusion, which resulted in many unsanitized images exposing all five forms of PII. PrivacyLens's thermal subtraction remained robust to persons appearing sideways or occluded, which resulted in no total failures. Table 7 presents the full results per application and PII removal type.

Table 7: The PII removal success rates for each application per each PII removal approach.

PrivacyLens (Hybrid, Embedded)	**Exercise**	**Object**	**Fall**	**All Apps**
Face	100%	100%	100%	100%
Skin Color	98.0%	96.0%	99.0%	97.7%
Hair Color	98.0%	95.0%	95.5%	96.2%
Gender	100%	100%	100%	100%
Body Shape	100%	100%	100%	100%
All PII Removal w/ Ground Truth	**96.5%**	**92.0%**	**95.5%**	**94.7%**
YOLO (RGB Only, Embedded)	**Exercise**	**Object**	**Fall**	**All Apps**
Face	99.5%	94.5%	93.0%	95.7%
Skin Color	57.5%	68.5%	79.0%	68.3%
Hair Color	64.0%	60.0%	56.0%	60.0%
Gender	99.5%	94.0%	92.5%	95.3%
Body Shape	99.5%	92.5%	92.5%	94.8%
All PII Removal w/ Ground Truth	**42.5%**	**42.0%**	**48.5%**	**44.3%**
Detectron2 (RGB Only, Desktop)	**Exercise**	**Object**	**Fall**	**All Apps**
Face	100%	100%	99.0%	99.7%
Skin Color	63.0%	68.5%	92.0%	74.5%
Hair Color	59.0%	61.0%	67.5%	62.5%
Gender	100%	100%	99.0%	99.6%
Body Shape	100%	100%	99.0%	99.6%
All PII Removal w/ Ground Truth	**37.5%**	**38.0%**	**63.5%**	**46.3%**

C PRIVACYLENS SANITIZATION MODES AND INTERVIEWS

We evaluated PrivacyLens's ability to remove PII robustly in Sections 4 and 5. While some end-user applications do not require any information about people (such as monitoring the level of trash in a public garbage bin), other applications are human-centric (such as the ones presented in Section 5 like activity monitoring and fall detection) and do require a level of information on detected humans, leading to a tradeoff between user privacy and the utility of the end application. In a user study, we explore several sanitization modes with varying degrees of PII removal for different end-user applications. To evaluate the user perception and acceptance of PrivacyLens's sanitization modes, a user study was conducted by interviewing 15 participants from 20 to 74 years of age.

From our literature search and various legal definitions of PII, we defined six PrivacyLens sanitization modes, as seen in Figure 5. Each

of these sanitization modes presented a tradeoff between privacy and recorded information; having multiple options empowers users to define the amount of PII they wish to permit to be captured and stored. Additionally, for some human-centric applications, an image that contains PII is needed to power downstream ML and AI tasks. In this case, a limited number of privacy-preserving features can be generated on the device (such as the user's body pose) and can be used to augment the PII-sanitized images or sent as as additional metadata. This section informed the design of the PrivacyLens applications in Section 5 that demonstrates that these sanitization modes are compatible with existing, off-the-shelf applications.

Table 8: Demographic information and statistics for the smart device questions of the 15 interview participants.

Demographic Information	Mean	SD	Min	Max
Age	38.3	21.9	20	74
How familiar and how often do you use your smart devices? (1 = Not at all, infrequent use 7 = Very familiar, daily use)	6.3	1.1	4	7
How much do you trust your smart devices to protect your privacy? (1 = Not at all 7 = Completely trusting)	3.7	1.7	1	6

C.1 Study Procedure

We recruited 15 participants (5 female, 9 male, 1 non-binary) through snowball sampling between the ages of 20 and 74 with IRB approval for this study. Four participants had an undergraduate-level education; the remaining 11 had graduate-level education. We additionally asked two questions to guage how familiar our participants were with smart devices and how trusting they were of these devices to protect their privacy. Table 8 provides demographic statistics about the participants. To start, each participant was shown a pre-recorded video clip from the raw RGB stream of the PrivacyLens prototype that demonstrated a person walking in a park, walking in their home, and exiting the shower (clothed). Then, for each sanitization mode (as seen in Figure 5), the participant was shown a pre-recorded video clip of that sanitization mode in operation. No questions were presented through videos. After watching the video, a researcher verbally presented each of the following scenarios surrounding three proposed environments sequentially as part of a semi-structured interview:

(1) *Imagine a city employee needs a way to monitor and maintain public spaces (e.g., trash on the ground, public recycle bins full, sidewalk repairs), which would allow the city to more quickly address issues.*
(2) *Imagine a tele-health application that can keep track of your daily self-care routines (e.g., doing dishes, exercise, etc.) in general home spaces, such as the kitchen or living room. These routines would be used for the early detection of chronic health conditions, such as emerging heart failure, Alzheimer's, and Multiple Sclerosis.*
(3) *Imagine a home safety application that can detect slip and fall events, the leading cause of death in those over 64, in private home spaces, such as the bathroom, and alert emergency medical services.*

Per Bell [7], to guard against carryover effects given the number of interviewees, the scenarios were presented in the order shown above starting with the least sensitive context (public) and ending with the most sensitive context (sensitive home). After each scenario was presented and discussed, for each of the six proposed sanitization modes they were asked to answer "How much would using this sanitization mode improve your sense of privacy compared to using a regular camera?" on a 7-point Likert scale (where 1 denotes "Significantly Worse", 4 denotes "Same as a regular camera", 7 denotes "Significantly Better"). The participants were allowed to ask any clarifying questions and asked only to consider the privacy aspects of the sanitization mode and to assume the sanitization mode has no effects on the performance of the application in the scenario with general comments recorded for each sanitization mode.

After each sanitization mode had been rated on a 7-point Likert scale, we asked our participants questions regarding which modes they would be willing to accept. For each scenario, we asked the participants to state which modes they would feel comfortable with, which acceptable mode they felt most comfortable with, which acceptable mode they felt least comfortable with, and which unacceptable mode they felt least comfortable with. We also recorded general thoughts and preferences from participants, comparing the modes across the three scenarios. Demographic information (age, gender, education level) about the participant was collected at the end of the interview and can be found in Table 8. Interviews took roughly 30 minutes on average.

C.2 Study Results

Overall, all of the sanitization modes averaged above 5.3 across all environments. Only Face Swap and Thermal Only had participants provide a score of 4 (the same as a regular RGB camera); all other sanitization modes had a minimum score of 5 or better. Thermal Subtraction and Stick Figure performed particularly well, averaging above 6.53 for all environments. Table 9 provides summary statistics for each mode across each environment. The scores for the modes cluster into three "tiers" that correlate to their semantic representation. Tier 1 sanitization modes have a visual representation of the body (Face Swap, Thermal Only), Tier 2 sanitization modes include the body's silhouette (Thermal Replacement, Ghost UI), and Tier 3 sanitization modes entirely remove the body's shape (Thermal Subtraction, Stick Figure). Those in Tier 1 have a decline in scores from public to home environments, those in Tier 2 have a decline in scores from general home to sensitive home environments, and those in Tier 3 score consistently high across all three environments.

Though these scores show each sanitization mode improves the perception of privacy over raw RGB cameras, whether users are willing to accept them presents a more challenging task. Overall, all participants reported at least one sanitization mode they were comfortable with across all environments. For public spaces, nearly half accepted all modes and all participants reported at least three acceptable modes. For general home spaces, all participants reported at least two acceptable modes. In private home spaces, all participants reported at least one acceptable mode, with two-thirds reporting at least two acceptable modes. The environment influenced the acceptability of the three "tiers" observed earlier. Tier 1 was generally accepted in public spaces, Tier 2 was generally accepted in public and general home spaces, and Tier 3 was generally accepted in all spaces. Thermal Subtraction or Stick Figure were consistently the preferred sanitization mode across all environments.

Table 9: The average scores for each sanitization mode and each environment (public, general home, and sensitive home). All situations scored above 5.3, meaning all sanitization modes were perceived to improve privacy over a regular RGB camera.

Sanitization Mode in Public	Mean	SD	Min	Max
Face Swap	5.7	0.9	4	7
Thermal Only	5.5	0.9	4	7
Thermal Replacement	6.5	0.6	5	7
Ghost UI	6.5	0.6	5	7
Stick Figure	6.8	0.4	5	7
Thermal Subtraction	6.9	0.4	6	7

Sanitization Mode in General Home	Mean	SD	Min	Max
Face Swap	5.3	1.2	4	7
Thermal Only	5.4	1.0	4	7
Thermal Replacement	6.3	0.7	5	7
Ghost UI	6.3	0.7	5	7
Stick Figure	6.7	0.6	5	7
Thermal Subtraction	6.7	0.6	5	7

Sanitization Mode in Sensitive Home	Mean	SD	Min	Max
Face Swap	5.3	1.0	4	7
Thermal Only	5.3	1.0	4	7
Thermal Replacement	6.3	0.6	5	7
Ghost UI	6.1	0.8	5	7
Stick Figure	6.5	0.6	5	7
Thermal Subtraction	6.7	0.6	5	7

C.3 Qualitative Results

We observed some overlap in the commentary provided by participants. In public spaces, many participants reported having a diminished expectation of privacy, making them more willing to accept the less aggressive PII-removal modes. However, in the home, multiple participants had significant concerns regarding modes that could capture whether the person was wearing clothes. Regarding Face Swap, one participant reported: *"I think this will make me comfortable in public spaces, but probably not enough in private spaces (although it's better than no intervention). It will also annoy me because I'll probably feel that I have to wear clothes any time at home."* However, if the sanitization mode could mask the person's body, such as with Ghost UI, the mode becomes much more acceptable: *"I may be nude in the bathroom and as long as I don't expose more information than body shape it will be fine."* Sanitization modes that prevent capturing nude bodies are critical for in-home cameras.

We also observed that personal preferences might lead them to trust one accepted sanitization mode more than others. One participant who favored Thermal Subtraction stated: *"It removes the body shape, and this does not cause privacy concerns. If the whole body is removed, there is no privacy concern for me."* Another participant who preferred Stick Figure stated: *"The stick figure makes the images seem way less humanoid than the other interventions. I like that."* While the PII removed from the image is similar in both cases, the participants strongly preferred one over the other. Presenting users with multiple options will allow them to choose a more comfortable sanitization mode and increase their likelihood of long-term adoption in their homes.

Finally, while modes such as Face Swap and Thermal Only did not receive widespread acceptance when used in sensitive areas, the consensus among all participants was that *any* sanitization mode provided an improvement over a regular camera. One participant reported: *"I would prefer Face Swap over a regular camera in all instances."* While no single sanitization mode is perfect, they collectively raise the minimum privacy standard as an alternative to raw RGB camera images.

ACKNOWLEDGMENTS

This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.