HeatDeCam: Detecting Hidden Spy Cameras via Thermal Emissions

Zhiyuan Yu
Washington University in St. Louis
St. Louis, USA
yu.zhiyuan@wustl.edu

Zhuohang Li
University of Tennessee, Knoxville
Knoxville, USA
zli96@vols.utk.edu

Yuanhaur Chang
Washington University in St. Louis
St. Louis, USA
c.yuanhaur@wustl.edu

Skylar Fong
Washington University in St. Louis
St. Louis, USA
skylarfong@wustl.edu

Jian Liu
University of Tennessee, Knoxville
Knoxville, USA
jliu@utk.edu

Ning Zhang
Washington University in St. Louis
St. Louis, USA
zhang.ning@wustl.edu

ABSTRACT

Unlawful video surveillance of unsuspecting individuals using spy cameras has become an increasing concern. To mitigate these threats, there are both commercial products and research prototypes designed to detect hidden spy cameras in household and office environments. However, existing work often relies heavily on user expertise and only applies to wireless cameras. To bridge this gap, we propose HeatDeCam, a thermal-imagery-based spy camera detector, capable of detecting hidden spy cameras with or without built-in wireless connectivity. To reduce the reliance on user expertise, HeatDeCam leverages a compact neural network deployed on a smartphone to recognize unique heat dissipation patterns of spy cameras. To evaluate the proposed system, we have collected and open-sourced a dataset of a total of 22506 thermal and visual images. These images consist of 11 spy cameras collected from 6 rooms across different environmental conditions. Using this dataset, we found HeatDeCam can achieve over 95% accuracy in detecting hidden cameras. We have also conducted a usability evaluation involving a total of 416 participants using both an online survey and an in-person usability test to validate HeatDeCam.

CCS CONCEPTS

• **Security and privacy → Human and societal aspects of security and privacy**.

KEYWORDS

Spy Camera Detection; Privacy; Cyber-physical Security; Thermal

ACM Reference Format:
Zhiyuan Yu, Zhuohang Li, Yuanhaur Chang, Skylar Fong, Jian Liu, and Ning Zhang. 2022. HeatDeCam: Detecting Hidden Spy Cameras via Thermal Emissions. In *Proceedings of the 2022 ACM SIGSAC Conference on Computer and Communications Security (CCS '22), November 7–11, 2022, Los Angeles, CA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3548606.3560669

CCS '22, November 7–11, 2022, Los Angeles, CA, USA
© 2022 Copyright held by the owner/author(s).
ACM ISBN 978-1-4503-9450-5/22/11.
https://doi.org/10.1145/3548606.3560669

1 INTRODUCTION

Hidden surveillance cameras, also known as "spy cameras", are video cameras hidden or disguised as other common objects, generally deployed with the goal of recording people without their knowledge. While surveillance cameras may have legitimate uses for home security, the presence of such cameras in private areas such as dressing rooms raises significant privacy concerns. In 2017, more than 6400 cases of illicit filming were reported in South Korea, many of which took place in hotels, and subsequently led to widespread protests [25, 44]. It is reported that incidents involving hidden cameras in Airbnb accommodations are prevalent where 1 in 132 listings have indicated cameras are installed and more than 17% did not specify where these cameras were placed [24].

Existing Detection Methods. To tackle the emerging threat of unlawful recording with spy cameras, existing approaches generally fall into two categories based on the physical channel on which the detection module operates, i.e., radio frequency (RF) signals and optical reflections. In the line of RF-based detection methods, both commercial detectors and research prototypes aim to capture RF emissions due to network communications of wireless cameras. Existing commercial products often provide an alert when RF signals are detected and can lead to high false positives. Therefore, recent research [12, 21, 29, 41] focuses on finer-grained analysis of network communications. For instance, DeWiCam [12] leveraged supervised learning to discover traffic flows of wireless cameras, and [21, 29, 41] achieved detection by identifying causality between scenario changes (e.g., user motion, ambient light on/off) and network traffic variations. On the other hand, optical-based methods detect cameras using reflections of the lens. Commercial optical detectors emanate red light from built-in LEDs to assist users' subjective judgment with the reflections, and a recent work by Sami et al. [37] leveraged the laser time-of-flight (ToF) depth sensors in newer generations of mobile devices to locate cameras hidden inside pre-identified suspicious objects. When the user scans carefully at an appropriate distance (0.5m), it is possible to leverage higher reflection intensity to reveal hidden cameras.

Usability Limitations of Existing Techniques. Despite recent success, the usability and generalization of existing methods are constrained by the physical signals used to detect spy cameras. Existing approaches that use the emission of RF signals from spy cameras are only applicable to those that are wirelessly connected, which accounts for 60% of the market [4]. However, the remaining

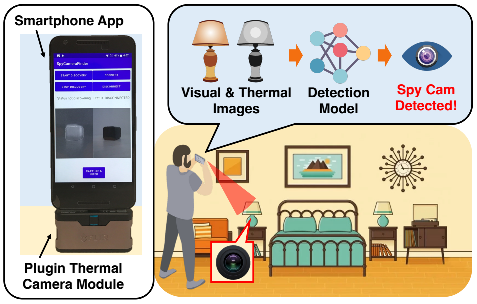


Figure 1: HeatDeCam overview.

40% of spy cameras simply do not have any network connection due to considerations on stealthiness, battery, form factor, and cost. Existing optical-based tools are also limited by the physical vector since the quality of the reflections depends heavily on the relative position of the user. Thus, the user will often first need to know where to check [37]. This can somewhat diminish the usability of the detectors for those without prior knowledge of spy cameras.

Our Detection Method. Recognizing limitations inherent to these physical channels, we turn to thermal imagery as the detection vector, since it is significantly less sensitive to the manner users operate on the equipment. The general idea of using thermal images to manually find hidden objects [15, 31, 45], particularly spy cameras [1, 13, 32], is well established due to its effectiveness in discovering heated objects. However, existing approaches generally rely on users to interpret the thermal images, which requires experience and expertise. In this project, we bridge the gap in usability by developing a neural-network-based automatic detection system, HeatDeCam, to make the technology accessible to users with different levels of prior knowledge.

As shown in Figure 1, HeatDeCam captures both thermal and visual images simultaneously from a thermal camera attachment and presents both the detection result (whether any spy cameras exist in the scene) and the visualized hidden places. Compared to existing approaches [12, 21, 29, 41], the proposed method can be applied to both networked and network-less spy cameras. It also offers detection over a wider region instead of a pre-identified suspicious object [37], making it easier to use for non-experts. Using a thermal camera dongle attachment (often available for around $100 [3, 14]), HeatDeCam can be deployed on most mobile phones using an app to house the lightweight machine learning components.

Challenges and Solutions. There are several technical challenges. One of the primary objectives of HeatDeCam is to make it usable for users without expertise. Instead of asking users to scan from a certain angle or scan pre-identified suspicious objects, HeatDeCam has to work on a wide inspection area that contains both regular electronic devices and spy cameras, both of which emit a non-negligible amount of heat. To facilitate better recognition of the unique heat signature of spy cameras, we incorporated a CBAM attention module [50] and an adaptive soft mask to hint the algorithm to learn high-dimensional heat dissipation features. However, from the user's perspective, knowing the existence of cameras within a broad viewing angle is helpful yet insufficient, since it can still be difficult to locate small spy cameras. To tackle this challenge, HeatDeCam leverages a gradient-based visualization mechanism [40] to explain the decision by highlighting suspicious areas, allowing the users to gradually narrow down to small hidden places. Lastly, HeatDeCam needs to be accessible to a diverse population of users, since some may carry phones with less computational resources. As such, we customized and optimized a lightweight ResNet-based neural network to improve computational efficiency.

Experiments and Findings. We implemented a prototype of HeatDeCam as an Android app displaying a live view of both thermal and visual images. The outputs from the machine learning model are overlaid on top of the images. To validate our machine learning algorithm, we have also collected multiple datasets, which contain 22506 thermal and visual images consisting of 11 cameras collected from 6 rooms across different environmental conditions (such as temperature). We found that our lightweight machine learning model has over 95% accuracy. To evaluate the usability, we have conducted both an online survey and an in-person usability test. In the online survey, we invited 380 participants and asked them to identify spy cameras in the output images of HeatDeCam. We then evaluated the efficacy based on the correctness of responses, and the usability based on the system usability scale (SUS) questionnaire. In the in-person usability test, 36 participants were given two minutes to find spy cameras within an office room, in which five different spy cameras were hidden. The performance of HeatDeCam is then compared with three commercial state-of-the-art detection methods in terms of detection rate, false positives, and usability in the user study. Our contributions are outlined as follows:

- We systemize existing commercially available spy cameras and detectors to understand the usability challenges.
- We propose HeatDeCam, a usable spy camera detection method leveraging unique heat dissipation and heat signature of spy cameras.
- We collect the first spy-camera dataset consisting of 22506 thermal and visual pictures, covering various scenarios including Airbnb, hotel, and office settings. All the data is open-sourced to the community[1].
- We develop a prototype of our approach as a mobile app, with which we study the effectiveness and usability by evaluating it against eleven spy cameras and three state-of-the-art detection methods.

2 BACKGROUND

In order to develop an effective detection method, we conducted a market survey to understand the types of commercially available spy cameras and their deployed environment. Through this process, we also attempted to understand the characteristics of the heat emanation of spy cameras compared to other electronics.

2.1 Spy Camera Market

Spy cameras are widely available for purchase online through marketplaces such as Amazon and Alibaba for prices as low as $10 [4]. We surveyed the top 50 best-selling spy cameras on Amazon, and summarized the results in Table 1. The full table can be found in the

[1] Data and the extended technical report are available at https://heatdecam.github.io/.

Table 1: Summary of top 50 cameras on Amazon.

Category	Network Connectivity		Market Share		Avg. Price
	Wireless	None	Number	Percentage	
Inconspicuous	16	0	16	32%	$36.91
Non-electrical	2	6	8	16%	$32.87
Electrical	19	7	26	52%	$66.39

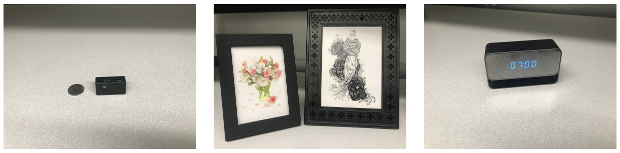

(a) Inconspicuous (b) Non-electrical camouflaged (c) Electrical camouflaged

Figure 2: Examples of the three types of spy cameras.

extended technical report available on the project website. In this work, we categorize existing spy cameras into three types based on heat emissions and deployment environment, which are the most important factors for detection. Spy cameras can either be inconspicuous and hidden in stealthy locations, or camouflaged as regular objects with or without electrical components.

Inconspicuous Cameras: They are generally small in size and are not concealed as regular objects. The low power consumption and small size provide them with unique advantages in stealthiness, enabling them to be hidden in locations such as drywall, bookshelves, or plants. Based on our survey, this type of camera consists of 32% of the top 50 most popular commercial spy cameras.

Non-electrical Camouflaged Cameras: These cameras are concealed as other regular non-electrical objects such as picture frames. Many of these cameras can be identified via thermal vector due to their abnormal heat emanation that mismatches the non-electrical disguises. This category accounts for 16% of the surveyed market.

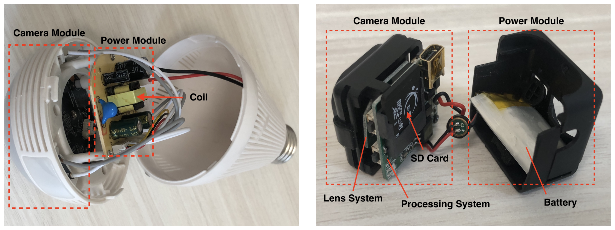
Electrical Camouflaged Cameras: These cameras are concealed inside other household electrical devices, such as a smoke detector, an alarm clock, or a USB charger. Compared to non-electrical camouflaged cameras, detecting these cameras via thermal is less straightforward, as the disguised electrical components emanate heat as well. These cameras share 52% of the surveyed market.

Figure 2 shows examples of an inconspicuous camera, two non-electrical camouflaged cameras concealed as picture frames, and an electrical camouflaged camera concealed as a clock, respectively. Each of these categories can be further split into spy cameras that are wirelessly connected and those that are not. Among the surveyed top 10 best-sellers, only 60% have Wi-Fi capability and are generally more expensive.

2.2 Structure of Spy Cameras

Figure 3 shows the typical structure of spy cameras. From our survey and dissection of common spy cameras, we conclude the structure to two modules, the *power module* and the *camera module*.

Power Module: All spy cameras need to be powered, typically via batteries or power lines. Batteries provide low-voltage DC power in which a large amount will be converted into heat. Cameras powered by wires are provided with relatively high (e.g., $110V$) alternating voltage, and they generally include a power circuit incorporating coils to transform input power to low direct voltage (e.g., $6V$). For electrical camouflaged cameras consisting of additional components, the disguised functionalities are supported by the same power module with cameras and therefore consume more power. Power modules in spy cameras generate significant heat according to Joule's law, which is critical for thermal-based detection.



(a) Camera disguised as a bulb powered by wire (b) Inconspicuous camera powered by battery

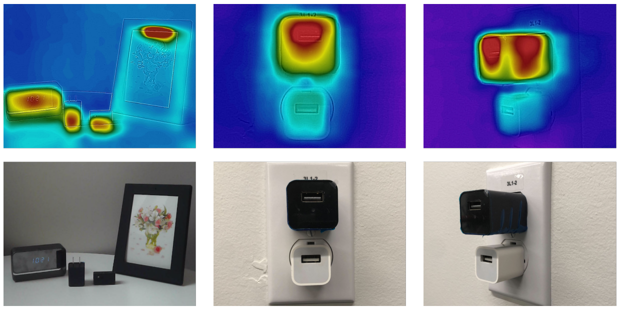
Figure 3: Structure of two typical spy cameras.

Camera Module: While almost all electronic devices contain power modules as well, the camera module is the key component that characterizes spy cameras. The camera module mainly consists of two parts: the *lens system* and the *processing system*. The lens system is the most typical and necessary component in all spy cameras. The lenses generally comprise convex and concave lenses to capture light, and the light beams are processed by image sensors, which convert the light radiance falling on a pixel sensor into a pixel intensity. The processing system provides several functionalities, and the most common yet important one is video processing where the video is compressed based on a coding standard such as *H.264* [49] for further transmission or storage. Additionally, it can provide more features such as Wi-Fi, audio noise reduction, motion detection, etc. These functionalities are incorporated on a board, on which each unit provides a different service.

2.3 Heat Patterns of Spy Cameras

The distinct components and internal layout of spy cameras lead to unique heat emissions that separate them from other objects.

Formation of Heat Patterns: The heat patterns of spy cameras originate from two aspects. First, regardless of how the cameras are concealed, camera-specific components always generate additional heat that leads to higher energy in thermal images. These patterns can be particularly helpful for detection when cameras are embedded in non-typical heat sources (such as picture frames). Second, the need for stealthiness inevitably alters the internal layout of spy cameras, which affects heat dissipation and displays as heat distribution patterns. Heat is a well-known problem for electronics, therefore the internal layout of the electrical components is generally optimized for efficient heat flow [11]. However, in order to add stealthy recording capabilities to existing objects, additional electrical components have to be added without altering the original form factor. This design choice of spy cameras often causes considerable changes to the internal layout and the corresponding



(a) Four spy cameras (b) A camera and a charger (c) A camera and a charger

Figure 4: Thermal view of four cameras and a charger plug.

thermal dynamics, leading to a uniquely different heat distribution pattern under thermal imaging.

Examples of Heat Patterns: The above insights are further demonstrated in Figure 4, which shows some examples of heat patterns. Figure 4(a) shows four spy cameras observed via thermal camera, including a camera concealed as an electronic clock, a camera concealed as a charger, an inconspicuous camera, and a camera concealed as a picture frame. They all display heat that is visible via thermal cameras, and compared to the charger and electronic clock, the abnormal heat emissions from the picture frame are more likely to raise alarm by humans. Figure 4(b) and 4(c) show a regular charger and a camera concealed as charger placed side-by-side. While they appear visually similar, their heat patterns are significantly different and can be attributed to two major reasons. Firstly, the heat on the camera charger is concentrated in two regions where there lies the power module and camera module, which are camera-specific components that produce additional heat. Secondly, the regular charger shows a uniform heat distribution due to the well-studied and commercially-applied layout optimization for heat dissipation [23, 42]. However, the camera components have to be deployed inside without changing the form factor for stealthiness, which interferes with the airflow that leads to the heat patterns.

Summary: These observations motivate us to adopt thermal as the detection vector due to its wide applicability and spy-camera-specific characteristics. On the other hand, it can be difficult for users without expertise to detect all spy cameras (especially those camouflaged as electrical devices). Therefore, usability for non-expert users is a key challenge.

3 RELATED WORK

The growing surveillance threats have stimulated the development of several commercial products and research prototypes for hidden camera detection. From the perspective of physical vectors used for detection, there are three categories, RF signal, optical reflection, and thermal emission, as shown in Table 2.

3.1 Detection via RF Signals

The key idea of RF-based detection lies in that wireless monitoring devices rely on Wi-Fi transmission, which leaks device information via signals in the radio spectrum [51]. Commercial detectors on the current market often have coarse granularity and will trigger an alarm if the received RF power in a particular frequency range is above the configured threshold [43]. However, as these devices generally do not involve signal analysis adapted for hidden cameras, they are likely to trigger a high false-positive rate from surrounding IoTs or Wi-Fi routers. To address the above limitation, prior studies proposed to improve with fine-grained analysis of the captured wireless communications [12, 21, 29, 41]. Cheng et al. [12] proposed to leverage supervised learning to extract and learn network traffic features of wireless hidden cameras, i.e., consisting of both video and audio streams, to distinguish them from other network applications. The following work proposed to identify the causality between physical activities and network traffic patterns. These activities include adjusting the lighting [29, 41], and motion stimuli [21]. While these methods can detect spy cameras in certain cases, they are fundamentally limited to detecting wireless monitoring devices only. However, our survey results show that only 6 of the top 10 (i.e., 60%) most popular spy cameras sold on Amazon support Wi-Fi connection.

Table 2: Comparison of HeatDeCam and existing work.

System	Channel			Detactable Cameras		Human Efforts	Reference
	RF	Optical	Thermal	w/o Wireless	Wireless		
DeWiCam	✓				✓	○	[12]
Blink & Flicker	✓	✓			✓	◐	[29]
SnoopDog	✓				✓	◐	[41]
MotionCompass	✓				✓	◐	[21]
LAPD		✓		✓	✓	●	[37]
Popcultural Thermal			✓	✓	✓	●	[1, 13, 32]
HeatDeCam		✓	✓	✓	✓	◐	This work

●=Maximum, ○=Minimum

3.2 Detection via Optical Reflections

All commercially available spy cameras rely on lenses to capture and record surroundings. Based on this observation, existing optical-based approaches rely on optical reflections from the lenses. A popular consumer detection comes with LEDs of a certain color (typically red) and a colored-glass viewfinder [26], which are designed to assist users in identifying reflections from the lens. However, this tool has a narrow range of inspection angles, and it heavily relies on the experience and expertise of the user to make a judgment call on the surroundings. To overcome this challenge, a very recent approach, LAPD [37], proposed to leverage the laser time-of-flight (ToF) depth sensors in the latest generation of smartphones to automate this process. The working principle is based on the *lens-sensor retro-reflection* effect, where embedded cameras reflect incoming laser pulses at a higher intensity, thereby revealing their hidden locations. However, even though the detection is automatic, users need to stand in an ideal approximation (distance of $0.5m$) to the spy camera and carefully scan the object with a specific speed ($0.05m/s$). Therefore, this approach takes a significant step forward but is limited to common drawbacks of the optical vector which requires the user to pre-identify suspicious objects and have a good distance and angle to the lens.

As a result, the optical approaches are more generally applicable to most of the cameras, with and without wireless connectivity, compared to the RF-based detection mechanisms. It relies on the user having more expertise in the technology.

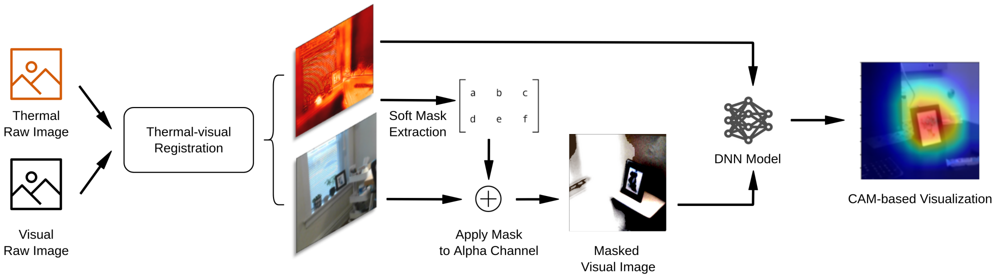


Figure 5: Overview of the HeatDeCam design.

3.3 Pop-cultural Efforts via Thermal Emission

Thermal imaging is a powerful mechanism that allows for inspection of object temperatures that are otherwise not available via traditional cameras. Building on top of this capability, its application can be seen in the industry, agriculture, medical, and public security fields [17]. For instance, its application scenarios range from detecting defects in mechanical parts [2, 47], illegal grow operations [15, 31], peripheral vascular disorder [5], to concealed weapons [45]. Recently, there are also attempts to leverage the unique capability of thermal cameras to detect hidden spy cameras under the pop culture setting [1, 13, 32]. These efforts share a similar insight as HeatDeCam, that all spy cameras will generate heat emissions, which can be captured by thermal imagery and serve as indicators that cameras are present. While these efforts demonstrate the feasibility of thermal imagery as a viable vector for detection, existing work primarily relies on user expertise and experience for interpreting the images and scenarios. Unfortunately, distinguishing subtle differences in heat dissipation patterns is not always easy for new users. To make this promising approach more user-friendly, we lean on machine learning to complement the manual effort in detection and identification.

4 ATTACK MODEL & SYSTEM OVERVIEW

4.1 Attack Model

In this paper, we consider a strong attack model where the attacker can install spy cameras in a target environment (e.g., hotel room and office) inconspicuously and has full control over the spy cameras, including the camera model selection and where/how the camera is installed in the environment. More specifically, the attacker is assumed to have the following capabilities:

- **Concealed Placement.** The attacker will try his/her best to hide spy cameras for stealthiness, and the hiding locations will depend on the type and the disguise of the cameras (Section 2.1).
- **Continuous Monitoring.** We assume the attacker will keep the cameras on to continuously monitor the victim's activities.
- **Heterogeneous Cameras.** For a higher coverage of the monitored space and to avoid being discovered easily, the attacker can employ multiple spy cameras of different types. Such multiform heterogeneity makes it more challenging for detection.
- **Limited Control.** The attacker may modify the configurations and functionalities of spy cameras such as disguising appearance and connectivity, but is not capable of hacking into the user's smartphone to disrupt the detection process of HeatDeCam.

4.2 System Goals

Design Objectives of HeatDeCam. The proposed HeatDeCam needs to meet the following design objectives.

- **Robustness.** It should have high detection accuracy (i.e., true positives) and acceptable false positives, with robustness against environmental variations such as scenarios and light conditions. It should be able to identify spy cameras in the area of interest, regardless of camera types and placement positions.
- **Usability.** It should be able to provide visual information to help users to discover spy cameras. The user operating HeatDeCam does not need to know the working principles of spy cameras.
- **Efficiency.** HeatDeCam is envisioned to be compatible with ubiquitous smartphones, and is capable of detecting spy cameras in real-time. Therefore, the technique needs to be efficient in terms of computing resources and time consumption.

4.3 System Overview

The overall design is shown in Figure 5. There are four technical challenges in design. (1) First, the thermal camera module involves a displacement between the thermal and visual lens, resulting in misalignment between thermal and visual images. To solve this issue, the thermal and corresponding visual images will be processed by thermal-visual registration to eliminate the misalignment. (2) Second, spy cameras are often surrounded by various objects in different environments that will inevitably interfere with the detection process. To filter out some irrelevant information associated with non-heating objects, we extract soft masks based on the thermal images and apply them to the alpha channel of the visual images. (3) Third, the processed thermal and visual images will be fed into a Deep Neural Network (DNN) model for camera detection. In order to handle complicated practical scenarios with efficient computations, we employ a lightweight ResNet-based convolutional neural network with attention modules to focus on important features from the thermal and visual input pairs. (4) To provide better usability and assist users to locate hidden cameras, HeatDeCam adopts a Class Activation Map (CAM) [40] to visualize the decision-making process and highlight the potential hiding placements.

5 SYSTEM DESIGN

Thermal cameras incorporate specialized filters to capture infrared emanation of objects, where its intensity has a positive correlation with the temperature according to the Stefan-Boltzmann Law [18]. Based on this principle, HeatDeCam incorporates algorithms to

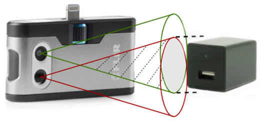


Figure 6: Displacement of lens causes misalignment between thermal and visual images.

analyze thermal imagery obtained from a portable thermal camera, which captures heat signatures. In this section, we will describe our algorithm design, detection model, and visualization scheme.

5.1 Thermal-visual Correlation

While thermal imagery alone shows the heated source, it lacks high dimensional information such as color and edge that are crucial for modern computer vision algorithms to enable inference of higher-level semantics. For instance, picture frames should not have unique heat signatures on the top of the frame. However, even though it is beneficial to have correlations between thermal and visual information, there are two technical challenges. First, the raw thermal and visual images directly captured from cameras do not align due to displacements of the two lenses. Second, heated areas are highlighted in the thermal images, while such focus does not present in visual images. To solve these issues, we incorporated image registration and soft masking mechanisms.

Thermal-visual Registration: As shown in Figure 6, the lenses that capture thermal and visual images are separated due to their distinct hardware structure. The existence of such displacement between the lenses causes the misalignment between the thermal and visual images. For example, when capturing the same spy camera with two lenses simultaneously, the camera may appear closer to the bottom half in the visual image (taken by the upper lens) but appear closer to the upper half in the thermal image (taken by the lower lens). Such misalignment varies based on the distance and angle of the object with respect to the camera. To handle this challenge, we used image registration based on discrete Fourier transform [34]. Given a thermal image and its corresponding visual image, the registration algorithm will calculate the differences between scale, rotation, and position of image features, and the visual image will be transformed to align with the thermal image.

Automatic Soft Masking: A naive approach to filter unnecessary information is to use the thermal information as a mask to remove all other areas to force the attention of the neural network on heated objects. However, this also inevitably eliminates the necessary deployment context information that is important for spy camera detection. Therefore, we propose to narrow down the search space by applying a soft mask. By applying the extracted soft masks from thermal images to the alpha channel of visual images, it is possible to essentially "bleach" the less-important areas. Specifically, the thermal image is first converted to a single-channel grayscale image and smoothed using a Gaussian kernel. Then we apply Otsu's method [33] to find the optimal threshold T from the pixel intensity histogram that maximizes the inter-class variance after binarization. Next, a soft mask M is generated by comparing each pixel value with the threshold. For each pixel value $x_{i,j}$, its corresponding mask value is computed as:

$$M_{i,j} = \begin{cases} x_{i,j} & \text{, if } x_{i,j} \leq T \\ 1 & \text{, otherwise} \end{cases} \tag{1}$$

meaning that the pixels highlighted in thermal images will be emphasized by a given alpha value of 1, and other areas will be given smaller alpha values depending on their heat in thermal images. The extracted masks will then be applied to the corresponding visual images by serving as the alpha channel.

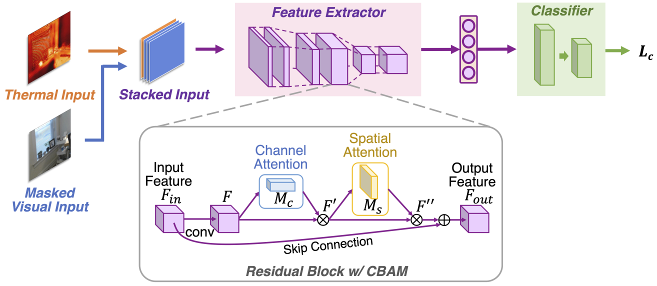


Figure 7: Architecture of the ResNet-based detection model.

5.2 Detection Model

With the processed thermal and visual image pairs, we designed a DNN model for feature extraction and camera detection. As shown in Figure 7, the proposed model consists of two sub-networks.

Attention-based Convolutional Feature Extractor: Following existing studies in thermal image analysis [7, 35], we employ a convolutional neural network based structure to extract useful features for spy camera detection. For computational and storage efficiency on mobile devices, the proposed feature extractor adopts a lightweight ResNet-18 [20] structure, which consists of 18 convolutional layers of 3 × 3 kernels with skip connections. We also evaluated several alternative neural network structures during preliminary experiments, such as MobileNet [22], Inception [46], and dual feature extractors. The experimental comparison is available in Section 7.4.

To handle the varying environmental factors (e.g., light conditions, viewing angles) and further enforce the model to focus on critical features, we adapt the Convolutional Block Attention Module (CBAM) [50] to increase the representation power. CBAM achieves this by sequentially applying a channel attention sub-module and a spatial attention sub-module, which learns to emphasize representative features along the channel axes and the spatial axes. Specifically, the network is built on residual blocks with CBAM embedded, each of which implements the following operations: given an input feature map $F_{in} \in \mathbb{R}^{C \times H \times W}$ and a set of weights W, the block first computes an intermediate feature map F via convolution $F = conv(F_{in}, W)$. Next, the attention module sequentially infers a 1D channel attention map $M_c \in \mathbb{R}^{C \times 1 \times 1}$ and a 2D spatial attention map $M_s \in \mathbb{R}^{1 \times H \times W}$. The inferred attention maps are then applied to the feature map sequentially:

$$\begin{aligned} F' &= M_c(F) \otimes F, \\ F'' &= M_s(F') \otimes F', \end{aligned} \tag{2}$$

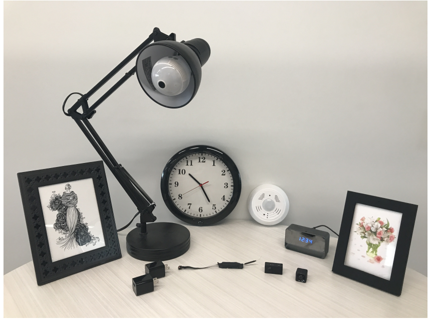


Figure 8: Selected spy cameras included in the dataset.

where ⊗ denotes element-wise multiplication. Finally, the output feature map is the sum of the refined feature map F'' and the input feature map passed along the skip connection, $F_{out} = F'' + F_{in}$. Built on such residual blocks, the feature extractor will produce a set of meaningful feature representations for each pair of thermal and masked visual image inputs.

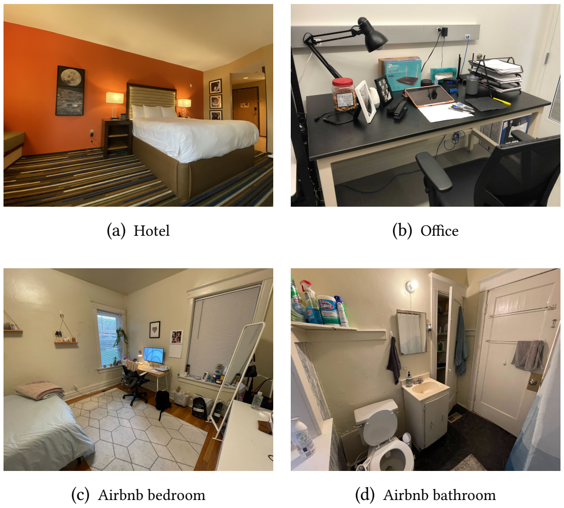
Classifier: The classifier contains a fully-connected layer that takes the extracted latent representations from the feature extractors and outputs the predicted probabilities. The shape of the output probability vector is adjusted according to the specific task, i.e., spy camera detection (binary classification) or spy camera recognition (multi-class classification). In either case, the classification loss L_c can be measured using the cross-entropy function.

5.3 Grad-CAM Visualization

To achieve the design goal of usability and improve model explainability, we further employ Gradient-weighted Class Activation Mapping (Grad-CAM) [40], a gradient-based approach for visualizing the decision-making process of the detection model by highlighting class discriminative regions. Using the gradients of the target concept (i.e., spy camera) flowing back to the final convolutional layer, Grad-CAM produces a localization map that highlights the regions that are important for prediction. The resulting visualization can be used to inform the user about highly suspicious regions and further assist the user to find the potential spy cameras.

6 THERMALVIEW DATA COLLECTION

The collected dataset incorporates four unique characteristics. First, it was collected in six rooms across four types of scenarios. These rooms include two rooms in a *hotel*, two rooms in an *Airbnb*, and two rooms in an *office*. Second, it was collected by 5 unique individuals, accounting for the differences in how people deploy objects and how users hold their phones when inspecting. Third, our dataset was collected at different times of the year, accounting for temperature and lighting variations. To summarize, this dataset was collected in uncontrolled settings, and more details will be discussed in the rest of the section. To further evaluate our system under an adverse environment where the entire space is heated, an additional dataset ThermalView-Adv was collected in a high ambient temperature for evaluation on adverse temperature environments, which will be described in Section 7.6 with experiments.



(a) Hotel

(b) Office

(c) Airbnb bedroom

(d) Airbnb bathroom

Figure 9: The environments contained in ThermalView.

6.1 Target Spy Cameras

The spy cameras included in the dataset are important, as they need to be representative of the market and real-world scenarios. Therefore, we surveyed the most popular spy cameras sold on Amazon and eBay, and selected a collection of 11 spy cameras including 3 inconspicuous cameras, 6 electrical camouflaged cameras, and 2 non-electrical camouflaged cameras. As shown in Figure 8, these cameras differ in a variety of attributes, including camera types, manufacturers, connectivity, disguises, and costs. The wide range of varieties and their popularity help us gain a set of heterogeneous spy cameras that are most likely to be deployed in real life.

6.2 Deployment Environment

Selecting deployment environment also plays an important role and is a non-trivial task. The hiding placements should fit the disguise of each spy camera, for example, a camera disguised as a charger is generally plugged into an outlet, and a camera concealed as a picture frame is often placed on the table. Thus, such environments should cover large diversity in the real world. We define an environment with three factors: scenario, room, and corner.

Scenario: A scenario is the type of space. In our dataset, we chose the three scenarios that are reported to contain spy cameras frequently, i.e., hotel, Airbnb, and office.

Room: It specifies the room in which the images are collected, as a scenario may host several different rooms. There are six rooms included in total, the office scenario hosts two rooms, and the hotel and Airbnb scenarios each includes a bedroom and a bathroom. Figure 9 shows four rooms that are included in the hotel, office, and Airbnb scenarios.

Corner: Even within a single room, objects can vary depending on the specific areas. For instance, a bedside table generally supports small IoTs like electronic clocks, while large appliances such as televisions are less likely to occur in this corner. There are thirteen corners in total, with four corners in the office scenario, four corners in the Airbnb scenario, and five corners in the hotel scenario.

6.3 Data Collection Strategy

The procedure of data collection can be summarized into two stages, scenario setup and image collection. In the first stage, we prepared various regular objects that are normally used in real life, such as voice assistants, iPad, laptops, smart phones, alarm clocks, headsets, routers, mouse & mouse pad, file organizers with books, pens & pen case, bottles, cups, snacks, tissue boxes, staplers, etc. Then we asked 5 participants to place these regular objects and change deployments based on their preferences. After each participant finished setting up the scenario, he/she conducted the second stage of image collection. In a single *corner* environment, the trials consisted of visual and thermal images being captured with the variation of four factors: (1) objects deployment in the room, (2) distance from the spy camera, (3) angle with respect to the spy camera's field of view (FOV), and (4) the ambient light condition. The participant was guided to hide the target spy cameras in the corners and repeat trials of taking pictures. After a participant finished collecting data in the scenario, we recycled deployed objects and guided the next participant to repeat the above steps.

To protect the participants' privacy, they were informed of the existence of hidden cameras in the room, and instructed to only take pictures when they were alone with no private items (e.g., driver license, family portrait) present. They were required to hold their phones as they would in a photography position. To avoid desirability bias, the participants were not made aware of the purpose of the data collection, which would have otherwise motivated them to take pictures with significant efforts that could best benefit the research (e.g., take clearer pictures at close distances).

By repeating these steps, the trials resulted in a total of 20474 thermal and visual images, of which 10738 (52.45%) images include spy cameras, while the rest 9736 (47.55%) images do not. A summary of the collected dataset is present in Table 3.

6.4 Data Annotation

The acquired thermal and visual images were labeled manually. For effective pairing of thermal images and the corresponding visual images, the two images in a pair share the same name based on the time (year-month-date-hour-minute-second) that the pictures were taken, while visual images have additional "*-orig*" at the end for distinction. We also manually checked and discarded the blurred pictures due to significant motion when capturing. We also created a separate csv file to record the attributes of each image. The file will specify the image name, image type (whether it is a thermal or visual image), numbered label of the spy camera inside the image, scenario label, and room label. We use 0 to represent no camera inside, and label 1 to 11 to represent each camera model.

7 EVALUATION VIA THERMALVIEW

7.1 Measurement Results

We randomly selected 80% data samples from the ThermalView dataset for HeatDeCam model training and used the rest for evaluation. During preprocessing, the image data was first normalized to $(0, 1)$ scale and then resized to 256×256 pixels resolution. As for data augmentation, we randomly cropped the image to 224×224 and applied a random horizontal flip. The model is trained on the cross entropy loss for a total number of 40 epochs, using the Adam optimizer [27] with $\beta_1 = 0.9$, $\beta_2 = 0.999$, and an initial learning rate of 0.001 (decayed by 0.1 every 10 epochs).

Table 3: Dataset statistics.

Image Types	Scenarios		
	Office	Hotel	Airbnb
# of Benign Images	3440	378	5918
# of Spy Camera Images	4816	1552	4370
# of Total Images	8256	1930	10288

Binary Classification: Figure 10(a) presents the binary classification results (i.e., whether the image contains spy camera) under various environments (i.e., office, hotel, apartment, and the overall dataset). We observe that our model is able to achieve a high detection accuracy (> 0.95) and recall (> 0.97) across all environment settings, which demonstrates the effectiveness of HeatDeCam.

Multi-class Classification: Figure 10(b) shows the confusion matrix of the multi-class classification result, where class label 0 represents the benign case (no spy camera) and labels 1-9 represent different types of spy cameras. We observe that our model can correctly differentiate different types of spy cameras with a high accuracy of 0.960.

Previously Unseen Cameras: For practical usage, it is important to examine how HeatDeCam performs in detecting cameras that are previously unseen in the training set. We split ThermalView based on the camera models, then randomly selected three cameras and separated the corresponding images, together with 20% of benign images to form the test set. As such, these three cameras are previously unseen by the model during the training process. The experiments were repeated ten times, each with different camera models included in the test dataset. HeatDeCam achieved a mean accuracy of 0.935 in average of ten trials, showing the robustness against previously unseen camera models. Compared to previous results, the slight performance degradation indicates that different camera models carry variations in their heat patterns. As expected, the learning model is not completely generalizable to all new cameras.

False Positives of Electronics: In order to avoid missing spy cameras, false positives are almost inevitable. However, too many false positives can significantly harm the usability of the system. A major challenge in this aspect is to distinguish spy cameras from regular electronics because they both emit heat capturable by thermal cameras. To this end, we evaluated the false positives using 1269 sets of pictures across three scenarios containing a voice assistant, a JBL speaker, an iPad, two laptops, three phones, a power bank, a charger plug, a power hub, an alarm clock, a headset, a pair of earbuds, a wrist band, a router, and a computer mouse. We utilized the previously trained model for binary classification, where 98.2% (n=1246) of the images were correctly classified as non-camera and only 1.8% of the images triggered false positives. One mitigation mechanism for the false positive is to encourage the user to move closer to the suspected object, and generally with fewer objects in view, the false positive rate will also drop.

Model Size & Detection Time: The model size and detection time are also critical for usability. The model is measured at 45.4 MB,

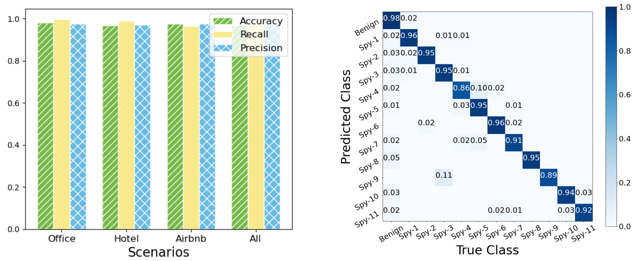


Figure 10: Binary (left) and multi-class (right) results.

which is relatively lightweight to be stored on mobile devices. To evaluate the detection time, we used the well-trained model to sequentially make predictions for 100 thermal-visual image pairs. The measured average inference time on a NVIDIA RTX 2080TI GPU is 12 ms, which is sufficient for supporting real-time detection.

7.2 Ablation Study

We conducted an ablation study across three scenarios to investigate the effectiveness of the components in HeatDeCam. The results are shown in Table 4. We observe that the performance of HeatDeCam degrades with the removal of the registration process and CBAM attention module, indicating that these components indeed improve camera detection. Notably, the removal of the attention module results in a more significant performance decrease (mean accuracy decrease of 2.7%) compared to the registration (mean accuracy decrease of 0.6%). This is because the misalignment between thermal-visual images is generally of a lesser magnitude and therefore has less impact on performance. However, the attention module plays an important role in HeatDeCam by enabling the algorithm to focus on key features, such as the heat pattern, that can best distinguish cameras. The removal of this critical module is likely to result in a significant performance decrease.

Table 4: Ablation study results.

Evaluated System	Office			Hotel			Airbnb		
	Acc.	Rec.	Prec.	Acc.	Rec.	Prec.	Acc.	Rec.	Prec.
HeatDeCam	0.982	0.997	0.975	0.967	0.989	0.971	0.976	0.965	0.976
HeatDeCam w/o Registration	0.978	0.988	0.980	0.961	0.975	0.978	0.969	0.958	0.966
HeatDeCam w/o CBAM	0.963	0.966	0.974	0.953	0.964	0.977	0.928	0.950	0.922

Acc.=Accuracy, Rec.=Recall, Prec.=Precision

7.3 CAM Visualization

The CAM-based visualization is designed to present users with potential locations of spy cameras detected in the captured images. This functionality is crucial for usability, and therefore it is important to validate if it properly highlights the desired areas.

Figure 11 shows visualization results on locating a spy camera disguised as a picture frame detected in two challenging environments, i.e., other heat sources exist and the distance is large. The three figures at the top show the first challenging scenario where the spy camera is deployed in the office and surrounded by other heat sources, such as voice assistants. While heat emanation from a voice assistant is obvious and may mislead users, the visualization highlights only the picture frame and therefore eliminates environmental interference. The three figures at the bottom show another challenging scenario, where the user is away from the spy camera. Therefore the heat emanation from the picture frame is not visible to the naked eye. Our algorithm shows robustness in such scenarios by emphasizing the picture frame areas. In these situations, it could be difficult for a human to locate the spy camera when only presented the thermal images, given the multiple heat sources and invisible heat patterns. However, our algorithm is shown to be robust for localization under such challenging circumstances.

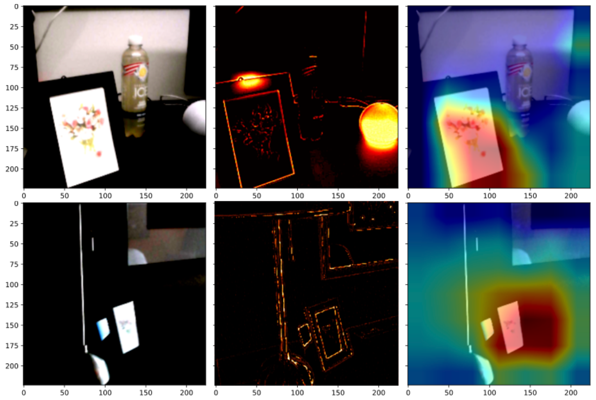


Figure 11: CAM visualization of a spy camera disguised as a picture frame in the office (top) and hotel (bottom).

7.4 Alternative Designs

In this section, we evaluated the performance of HeatDeCam with alternative technical designs mentioned in Section 5.

There are four alternative designs in question. 1) *Thermal Input Only:* As thermal images carry the most critical features - heat patterns, it is possible that algorithms can only rely on thermal images to detect spy cameras. As such, visual input will not be stacked with the thermal images. 2) *Thermal/Visual Dual Feature Extractors:* Besides stacking thermal and visual inputs for detection, it is also possible that they each go through a separate feature extractor. The parameters of feature extractors can be updated independently. 3) *Hard Mask:* Soft mask is designed to manipulate the transparency of image regions to reduce the input of less-important information. In contrast, hard mask is an alternative where those regions will be completely filtered out, i.e, replaced with black or white. 4) *DNN Architecture:* The architectural design of DNN models is a major research direction in the field of machine learning, and numerous architectures have achieved state-of-the-art performance in different application domains. In this study, we evaluated two of the most well-known models, MobileNet [22] and Inception [46].

The evaluation results are shown in Table 5. We observe that modifying DNN architectures and adopting a dual-extractor structure has a relatively small impact on performance compared to the current optimal design. This is because the major functionality (i.e., extract features) of these DNN models is attributed to the

Table 5: Performance of alternative design elements.

Alternative Designs		Accuracy	Recall	Precision
Thermal Input Only		0.919	0.923	0.930
Dual Feature Extractors		0.955	0.951	0.966
Hard Mask		0.824	0.840	0.843
DNN Model	MobileNet	0.970	0.965	0.980
	Inception	0.952	0.940	0.971

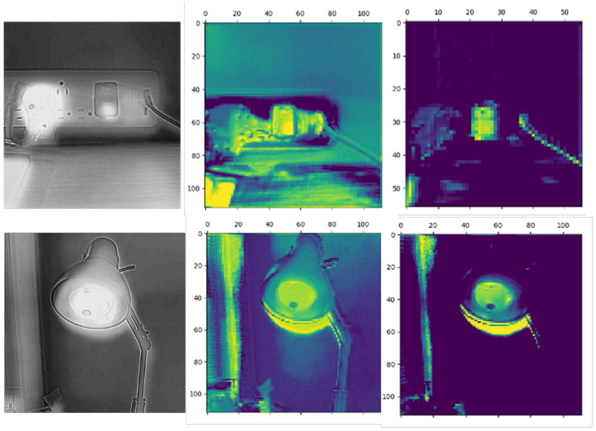


Figure 12: Intermediate results of two cameras disguised as a charger (top) and a bulb (bottom). Highlighted areas contribute most to the detection.

convolution layers, which are the foundations of these CNN-based neural networks. As a result, the variations in their architectures only slightly affect the effectiveness of feature extraction and representation. In contrast, removing visual input and applying hard masks to entirely filter out image regions that display less heat can significantly degrade performance. This can be attributed to the elimination of some critical features used for detection. We further demystify those features through experiments in Section 7.5.

7.5 High Dimensional Feature Space

Understanding the key features extracted from thermal-visual inputs is essential for future explorations. In general, DNN-based machine learning algorithms work by mapping input to a high-dimensional feature space where the data are expected to be more separable [10, 16]. While those features enable classifiers to achieve promising performance, interpreting their exact physical meanings and how they aid the decision-making process remains an open research problem [28, 36]. In this effort, we validated two key features by 1) analyzing intermediate results of the model and 2) experimentation in the format of an ablation study.

Heat Patterns: The key idea behind HeatDeCam lies in the unique heat patterns of spy cameras, which display on thermal images as spatial distributions and edges. These features are further validated in the intermediate results extracted from the model. Figure 12 shows some examples of intermediate results, where the highlighted areas represent the regions that contribute most to the prediction. These two examples show that the model primarily focuses on the heated regions of cameras even though there are other heat sources. To disrupt this feature in the experiments, we shuffle the spatial distribution of the raw infrared data within the heated regions while maintaining their statistical distribution. As shown in Table 6, the detection accuracy degrades significantly when the heat dissipation pattern is removed, showing that the uniqueness of the spy camera thermal signature is a key feature used by the model.

Table 6: HeatDeCam performance with removed features.

System	Accuracy	Recall	Precision
HeatDeCam	0.968	0.978	0.976
w/o Heat Patterns	0.769	0.746	0.816
w/o RGB Colors	0.932	0.914	0.961

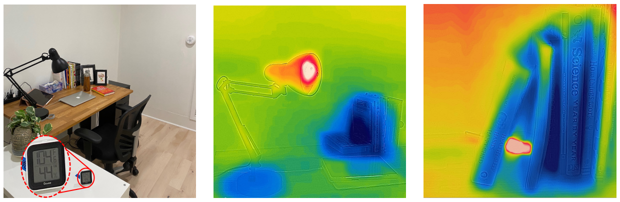
RGB Colors: RGB colors are important features for human visual systems [8] and have proven effective in various image recognition tasks [19, 30]. They are high-dimensional by nature, as they are represented by three channels, each with the same size of the original images. In the specific task of spy camera detection, colors contain rich context information, such as the type of object the spy camera is disguising as. For experiments, we converted visual inputs from RGB to grayscale before they were fed into HeatDeCam. As shown in Table 6, by removing the RGB color, the accuracy decreases from 0.968 to 0.932. It is also interesting that the impact of removing the heat pattern is significantly larger than removing the RGB color, reinforcing the importance of heat dissipation.

7.6 Results in an Adverse Environment

All physical vectors have their own limitations, and can be "saturated" by the same physical signals within the target environment. Thermal is not an exception. Similar to omnipresent wireless signals for RF-based detection, thermal detection approaches, such as HeatDeCam, can be impacted by the ambient temperature, and it is important to understand such limitations.

Data Collection: We collected an additional dataset ThermalView-Adv in a *top floor bedroom* during summer, where the average room temperature reached 104.6°F. ThermalView-Adv is a distinct dataset from ThermalView that contains 1016 sets of images (837 with cameras and 179 benign images), involving 9 cameras, 8 electronics, and 10 non-electronic objects. The room setup and measured temperature are shown in Figure 13(a). Figure 13(b) and 13(c) show examples of heat emissions from a camera concealed as a lamp bulb and an inconspicuous camera hidden behind books. Even with the high ambient temperature, the thermal signature of spy cameras remains visible due to the additional heat from video recording, this pattern can be further sharpened using the automatic calibration in either software or hardware.

Evaluation in the Adverse Environment: We conducted two experiments to validate how such an environment could impact the performance of HeatDeCam. We first utilized the model trained on images in regular temperatures (ThermalView) and used the images in the adverse environment (ThermalView-Adv) for testing only. The resulting detection accuracy is at 0.951 as compared to the original 0.968. This shows that our model is relatively robust in such an adverse environment, while a higher temperature of the ambient



(a) Room setup (b) A camera concealed as bulb (c) A camera behind books

Figure 13: Experiments in a top floor room representing the adverse environment, with a temperature of 104.6°F.

and surrounding objects indeed mislead the recognition algorithm to some extent. Next, we also investigated to incorporate adverse environment in training to further enhance HeatDeCam. We trained a model on a combination of ThermalView and ThermalView-Adv following the same settings, and achieved a detection accuracy of 0.969 and recall of 0.983. It shows that adverse environment can be mitigated by involving corresponding data in the training process.

8 ONLINE SURVEY STUDY

To understand how well the design empowers users to detect spy cameras, we conducted online surveys to study the efficacy. Additional details on the survey (including previous rounds) are available in the technical report on the project website. We obtained formal approval for these experiments from our university's Institutional Review Board (IRB).

8.1 Survey Design

The survey was fully structured and designed prior distribution. All the questions were designed in a multiple-choice manner. The survey consists of four major components as described below.

Task to Find Spy Cameras: The participants were presented with 5 sets of images and asked to finish the task of judging whether there are any spy cameras hidden inside. Each set consists of three images: a visual image as the baseline, and the corresponding thermal and CAM images provided by HeatDeCam for comparison. We further clarified that the thermal and highlighted images were two functionalities provided by a system. To make the survey accessible to the diverse population, we refer to the CAM image as "highlighted image" in the questions. For each test, the participants were first presented with the visual image only and asked if they believe there are any spy cameras hidden inside ("Yes", "No", "I don't know"), and how confident were they about the judgment on a 5-point scale. They were then presented with the thermal and CAM images placed side-by-side and asked the same question for their judgments. One example of a set of such images is shown in Figure 11.

Usability Test: Followed by the five tests, we asked participants to rate the usability of our scheme with a standard system usability scale (SUS) [6, 9], which consists of ten questions with five response options (from strongly agree to strongly disagree).

Cost: We also try to measure users' acceptability of hardware costs. The question is phrased as, "Please rate your willingness to pay $100 for a thermal camera that can be attached to your phone and used to detect spy cameras." The choices rate extremely unwilling to extremely willing on a 5-point Likert scale.

Demographics: In the end, the participants were asked about their age range, gender identity, and previous knowledge of detecting spy cameras. All the answers were collected anonymously.

8.2 Survey Results

A total of 106 people participated in the second-round online survey. The results are summarized from the following perspectives.

Effectiveness: Based on the responses to the tasks, we defined a score metric to calculate the detection results, $Score = D \times \frac{Conf}{5}$, where D represents the decision correctness (1 point for a correct answer, -1 for a wrong answer, and 0 for "I don't know"), and $Conf$ is the rated level of confidence. Such multiplication involves both user judgment and their confidence, in an attempt to reduce the impact of random guessing. The score indicates the correct level of participants that successfully identify spy cameras. The range for the score is $[-1, 1]$, with a higher score indicating better judgment. For the given 5 sets of images, the participants achieved an average score of -0.38 with only visual images, which indicates that they were more likely to fail to identify the existence of spy cameras hidden in the visual images. However, after they were presented with the thermal and CAM images, the average detection score increases to 0.82. We denote such detection score with visual image only as V score and that with images provided by HeatDeCam as T score. Based on these two sets of scores quantifying the detection effectiveness with and without HeatDeCam, we conducted the Welch's t-test [48] to further confirm their statistical significance in difference, under the null hypothesis that "there is no difference between paired V scores and T scores". We obtained a test statistic of 59.82 with 159.84 degrees of freedom ($p < 2^{-16}$) from sample data, and hence we reject the null hypothesis at 1% significance level, implying that such a difference is significantly different from zero. As a result, it indicates that thermal and CAM images are helpful to the participants for spy camera identification.

Usability: We calculated and normalized SUS scores for individual responses, and the mean SUS score for our provided scheme is 92.57 (±5.89). With reference to SUS on a curve with percentile ranks [38, 39], our scheme achieves a SUS score above average of 68 (at 50% percentile), validating the usability.

Costs: In terms of cost, most participants (n=96, 90.6%) rate their willingness to pay as 4 to 5. Specifically, 22.6% (n=24) participants rate 5 and 67.9% (n=72) rate 4, leading to an overall mean willingness to pay of 4.09. The results indicate that such cost is acceptable to the majority of participants.

Demographics: The participants consist of 68.9% male and 31.1% female. Besides, most of them (n=72, 67.9%) aged 18-34, 16% (n=17) aged 35-44, and 16% (n=17) aged 45 and above.

9 IN-PERSON USABILITY TEST

To test the effectiveness and usability of HeatDeCam in practical usage, we also conducted an in-person usability test followed by observational interviews. The experiments and interviews were formally classified as exempt by our university's IRB. In these experiments, participants were separated into four groups, each assigned

Table 7: Deployed spy cameras in the experiments.

Type	Wi-Fi	Disguise	Hiding Place	Denotion
Inconspicuous	N	-	Under the book shelf	I-N
Inconspicuous	Y	-	Behind the books	I-Y
Non-electrical	Y	Frame	On the side table	NE-Y
Electrical	Y	Clock	On the desk	E-Y
Electrical	N	Charger	Plugged to the socket	E-N

one detection method. As is shown in Figure 14(a), the evaluated detectors involve HeatDeCam and three state-of-the-art detection methods as baselines. These baseline methods include a G319 RF-based detector, and a red LED detector in the flashing mode and in constant lighting mode, respectively. After training on how to use the detection tool, participants were asked to use it to find spy cameras hidden in a pre-set scenario within two minutes. During the following interviews, the interviewees were asked about their user experience with the given tools.

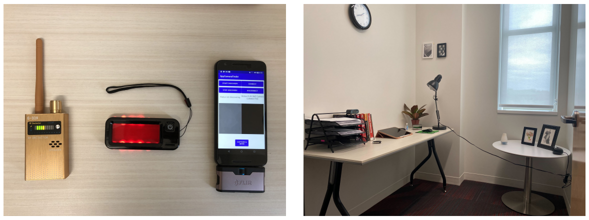
9.1 Recruitment Strategy

Detecting spy cameras is a human-in-the-loop process that involves users' empirical judgments. The participants' prior knowledge and detection experience can affect the detection results. As such, we first asked the participants to rate their prior knowledge and experience in the pre-experiment survey, and the answers were collected in an anonymous manner.

Specifically, we surveyed age range, gender identity, and participants' self-estimated knowledge of detecting spy cameras on a scale from 1 to 10. There were 36 participants in total, consisting of 17 (47.2%) male and 19 (52.8%) female. Additionally, 24 (66.7%) participants aged 18-29, 9 (25%) participants aged 30-50, and 3 participants (8.3%) aged over 50. As for the prior knowledge of detecting spy cameras, many participants (n=12, 33.3%) estimated themselves to have the least knowledge with a score of 1, and 4 participants possessed the highest knowledge level among all participants with a score of 6 out of 10. As such, no participant self-reported advanced expertise or experience in detecting spy cameras. Based on the survey results, we grouped all the participants into four groups by evenly assigning people with similar prior knowledge to different groups. As a result, each group was assigned 9 participants, and all groups held a similar overall knowledge level.

9.2 Experiment Setup

The test site was set up in a separate study room. To simulate practical usage of HeatDeCam, the room was not included in the ThermalView, and the room set-up is shown in Figure 14(b). With the spy cameras and other regular objects deployed in the room, we invited four volunteers to enter the room and asked them to find hidden cameras within two minutes. After they found cameras, we redeployed the selected spy cameras to new hiding places, and repeat the process until they could not find any cameras with raw eyes. This was done to ensure that all the spy cameras were properly hidden (*concealed placement* in the attack model). Note that these four participants were not involved in the following experiments. As a result, 5 spy cameras were deployed as summarized in Table 7, including two small hidden cameras, two electrical camouflaged cameras disguised as a charger and a clock, and one non-electrical camouflaged camera disguised as a picture frame. Among these cameras, three of them support Wi-Fi transmission. These cameras were selected to simulate *heterogeneous cameras* in our attack model. The denotations of these cameras are based on their types and connectivity (whether support Wi-Fi transmission) and will be used as an abbreviation in the illustration of detection results.



(a) RF detector, red LED detector, and ours (b) Experiment room set up

Figure 14: Environment set up and evaluated detectors.

Table 8: In-person experiment results.

Tool	Num of Participants Who Detected Cameras					ADR	FP
	I-N	I-Y	NE-Y	E-Y	E-N		
HeatDeCam	9	7	7	8	8	86.7%	1
Flash	3	1	6	3	5	40%	3
Constant	2	3	5	2	7	42.2%	2
RF	1	4	6	7	3	46.7%	8

ADR=Average detection rate, FP=False positive

9.3 Experiment Process

All the spy cameras have been kept running for an hour prior to the experiments, which was done to simulate *continuous monitoring* in the practical attack model. Before the participants entered the room and began camera hunting, they were informed of the task of "try to find suspicious hidden spy cameras within two minutes." However, they were not informed of the number and types of spy cameras hidden in the room. The experiments began once we ensured that the participant was familiar with the detection process using the given tool. The participants were invited to the set-up room one by one, each was given two minutes to find hidden cameras. After the time limit, each participant was notified to exit the room, and the organizer would enter the room to help verify the spy camera identification results. Then, the user was asked to describe his/her user experience and indicate whether the given tool was helpful as "helpful", "maybe helpful", or "not helpful". All the surveillance recordings were permanently deleted after the experiments.

9.4 Experimental Results

The evaluation metrics include detection rate calculated with the number of identified cameras, false positives, and usability. The detection rate and false positive are measured quantitatively while usability was assessed based on the post-experiment interview. Specifically, the average detection rate is calculated as the average detection rate of participants in a group, and false positive refers to the number of objects that were falsely suspected as spy cameras.

Detected cameras: Although the in-person experiment was conducted in a new room that is not included in ThermalView dataset on which the algorithm was trained, it shows the transferability of our algorithm which can reliably recognize spy cameras in new environments. The number of participants that detected each spy camera is shown in Table 8. Our designed HeatDeCam achieved the highest detection rate of 86.7% among all detection methods due to the following advantages. First, our method was shown effective in detecting spy cameras with and without wireless connections, whereas users with the G319 RF-based detector were less successful in detecting cameras without wireless connections (I-N and E-N). Second, flash and constant methods achieved a similar detection rate, and they were less impacted by connectivity but were less useful when detecting I-N, I-Y, and E-Y cameras. After interviewing users, we found this was because users were not able to see the reflection spot since the cameras were properly hidden under the bookshelf and behind the books, respectively. Besides, the E-Y camera disguised as a clock has a glass in front of the camera lens, which therefore reflected most of the lights and hindered the reflection-based detection. Our method does not exhibit such limitations in practical usage.

False positive: HeatDeCam has one false positive during the experiments. It was because the user mistook a regular alarm clock for a spy camera, although, our system did not indicate so. As a comparison, flash and constant lighting with an LED detector have 3 and 2 false positives, while the G319 RF detector exhibits the worst false positive among all these four compared methods. Based on the post-experiment study, it was because the G319 detector kept beeping when close to some regular objects due to the nearby electronics like wireless mouse. Such coarse-grained indicators misled the participants to take more false positives and downgraded the usability.

Usability: We also surveyed participants' user experience after the experiments. Most participants (n=34, 94.4%) indicated that our HeatDeCam prototype was "helpful" in aiding in detecting spy cameras. Specifically, users mentioned that the thermal view on the app was helpful in quickly finding cameras hidden behind obstacles (e.g., books) and emanating significant heat. The first participant indicated "not helpful" for our tool because the phone frequently went to sleeping mode due to monitor protection settings, which was fixed for the following participants. Two participants indicated "maybe helpful" because they thought the app's UI design could be further simplified to achieve better usability.

10 LIMITATIONS

Device Cost: Compared to existing work that relies on network analysis using mobile devices, the prototype of HeatDeCam requires a thermal camera attachment that costs $100. However, such thermal cameras can be built as low as $26 with a dongle [14]. Besides, our survey results show that most people are willing to pay this amount to protect their privacy (Section 8.2). Therefore, we believe the cost of our method is acceptable.

New Cameras and Environments: Although we have covered as many types of cameras and deployment environments in the ThermalView collection and experiments as possible, it is hard to enumerate all the spy cameras and rooms. However, our evaluation on previously unseen cameras, and the in-person experiments conducted in a new room that was not included in the ThermalView show preliminary evidence of the transferability of our algorithm, where users were able to localize almost all spy cameras equipped with our developed app. We also envision that our method could further incorporate users' input during their usage, supplementing the data set and improving the performance.

Participant Recruitment: During the experiments, we grouped participants into four groups based on their prior knowledge and experience of detecting spy cameras, while disregarding the age distribution. This is because we would consider the impact of age to mainly affect the knowledge of spy camera technology and detection techniques, which is included in the pre-experiment questions. Unfortunately, our recruitment of participants involves some level of sample selection bias due to limited resources. For instance, the age distribution of invited participants is unbalanced, with the majority of participants (n=24, 66.7%) aged 18-29, while only include a few elder participants (n=3, 8.3%) aged over 50. Lastly, the participants have the highest self-estimated prior knowledge of 6 out of 10, therefore our experiments did not involve people with advanced knowledge in detecting spy cameras. It is possible that advanced users may exhibit a smaller performance gap in detecting cameras when using different tools. However, we consider it as a reflection of the knowledge level of the general public, where people scarcely have too much experience detecting spy cameras.

Desirability Bias: We designed the survey and experiments in a way to mitigate ambiguity and communication inefficiency. For instance, we refer to the CAM-based image as "highlighted image" to make the survey questions more accessible to the diverse population. We also attempted to mitigate the desirability bias by including both positive and negative statements to avoid priming, and hiding the research goal. However, they may still be subject to such bias that is unobserved from the researcher's perspective. For instance, the participants might be aware that they are evaluating the efficacy of a newly-developed tool (i.e., HeatDeCam). As such, some may be inclined to over-report positive feedback as they consider it to be beneficial to and desirable to the researcher.

11 CONCLUSION

In this paper, we surveyed and analyzed existing spy cameras and detectors from a perspective of security and usability. Based on identified limitations of existing detection mechanisms and analysis of real-world scenarios, we leveraged thermal imagery and machine learning to design an automatic spy camera detection tool, aiming to offer accurate detection without requiring significant user expertise. We have collected and open-sourced a large dataset of spy cameras in different settings to validate our design and for the community. Our design is further validated using both online and in-person usability tests.

ACKNOWLEDGMENT

We thank the reviewers for their feedback and Matthew Venezia for his help in this project. This work is supported in part by US National Science Foundation under grants CNS-1916926, CNS-2038995, CNS-2154930, CNS-2114161, ECCS-2132106, and CBET-2130643, and by Army Research Office under contract W911NF-20-1-0141.

REFERENCES

[1] 2022. Can a thermal imaging camera detect hidden cameras? (July 2022). https://www.quora.com/Can-a-thermal-imaging-camera-detect-hidden-cameras

[2] I Jorge Aldave, P Venegas Bosom, L Vega González, I López De Santiago, Birgit Vollheim, Lennard Krausz, and Marc Georges. 2013. Review of thermal imaging systems in composite defect detection. *Infrared Physics & Technology* 61 (2013).

[3] Amazon. 2022. PerfectPrime IR202, (IR) Infrared Thermal Imager Camera 4800 Pixels, -40 752°F, 15Hz for Android Type C Mobile Phone. (2022).

[4] Amazon. 2022. Spy Camera. (2022). https://www.amazon.com/s?k=spy+cameras

[5] Subramaniam Bagavathiappan, T Saravanan, John Philip, T Jayakumar, Baldev Raj, R Karunanithi, TMR Panicker, M Paul Korath, and K Jagadeesan. 2009. Infrared thermal imaging for detection of peripheral vascular disorders. *Journal of medical physics/Association of Medical Physicists of India* 34, 1 (2009), 43.

[6] Aaron Bangor, Philip T Kortum, and James T Miller. 2008. An empirical evaluation of the system usability scale. *Intl. Journal of Human–Computer Interaction* 24, 6 (2008), 574–594.

[7] Yukti Bhatia, Rachna Rai, Varun Gupta, Naveen Aggarwal, Aparna Akula, et al. 2019. Convolutional neural networks based potholes detection using thermal imaging. *Journal of King Saud University-Computer and Information Sciences* (2019).

[8] Inês Bramão, Luís Faísca, Karl Magnus Petersson, and Alexandra Reis. 2012. The contribution of color to object recognition. In *Advances in object recognition systems*. InTech, 73–88.

[9] John Brooke et al. 1996. SUS-A quick and dirty usability scale. *Usability evaluation in industry* 189, 194 (1996), 4–7.

[10] Dong Chen, Xudong Cao, Fang Wen, and Jian Sun. 2013. Blessing of dimensionality: High-dimensional feature and its efficient compression for face verification. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 3025–3032.

[11] Kai Chen, Jianwei Xing, Shuangfeng Wang, and Mengxuan Song. 2017. Heat source layout optimization in two-dimensional heat conduction using simulated annealing method. *International Journal of Heat and Mass Transfer* 108 (2017).

[12] Yushi Cheng, Xiaoyu Ji, Tianyang Lu, and Wenyuan Xu. 2018. Dewicam: Detecting hidden wireless cameras via smartphones. In *Proceedings of the 2018 on Asia Conference on Computer and Communications Security*. 1–13.

[13] Ahad Cove. 2020. Can a Thermal Camera Find Hidden Spy Cams? (June 2020). https://www.youtube.com/watch?v=Z1mc9dVguzo

[14] Electromaker. 2022. Grove - Thermal Imaging Camera - Mlx90614 Dcc Ir Array With 35 Fov. (2022).

[15] Lita Epstein. 2004. *The complete idiot's guide to the Supreme Court*. Penguin.

[16] Jianqing Fan and Jinchi Lv. 2010. A selective overview of variable selection in high dimensional feature space. *Statistica Sinica* 20, 1 (2010), 101.

[17] Rikke Gade and Thomas B Moeslund. 2014. Thermal cameras and applications: a survey. *Machine vision and applications* 25, 1 (2014), 245–262.

[18] Steven B Giddings. 2016. Hawking radiation, the Stefan–Boltzmann law, and unitarization. *Physics Letters B* 754 (2016), 39–42.

[19] Ali Güneş, Habil Kalkan, and Efkan Durmuş. 2016. Optimizing the color-to-grayscale conversion for image classification. *Signal, Image and Video Processing* 10, 5 (2016), 853–860.

[20] Kaiming He, Xiangyu Zhang, Shaoqing Ren, and Jian Sun. 2016. Deep residual learning for image recognition. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 770–778.

[21] Yan He, Qiuye He, Song Fang, and Yao Liu. 2021. MotionCompass: pinpointing wireless camera via motion-activated traffic. In *Proceedings of the 19th Annual International Conference on Mobile Systems, Applications, and Services*. 215–227.

[22] Andrew G Howard, Menglong Zhu, Bo Chen, Dmitry Kalenichenko, Weijun Wang, Tobias Weyand, Marco Andreetto, and Hartwig Adam. 2017. Mobilenets: Efficient convolutional neural networks for mobile vision applications. *arXiv preprint arXiv:1704.04861* (2017).

[23] CUI Inc. 2020. Design Considerations for Thermal Management of Power Supplies. (July 2020). https://www.cui.com/catalog/resource/design-considerations-for-thermal-management-of-power-supplies

[24] David Janssen. 2020. Many Airbnbs have cameras installed, especially in the US, Canada and Singapore. *VPN Overview* (Nov 2020). https://vpnoverview.com/news/camera-presence-airbnb-accommodations/

[25] Sophie Jeong and James Griffiths. 2019. Hundreds of motel guests were secretly filmed and live-streamed online. *CNN* (Mar 2019). https://edition.cnn.com/2019/03/20/asia/south-korea-hotel-spy-cam-intl/

[26] Adam Juniper. 2022. Best hidden camera detector in 2022. (March 2022). https://www.digitalcameraworld.com/buying-guides/best-hidden-camera-detector

[27] Diederik P. Kingma and Jimmy Ba. 2015. Adam: A Method for Stochastic Optimization. In *3rd International Conference on Learning Representations, ICLR 2015, San Diego, CA, USA, May 7-9, 2015, Conference Track Proceedings*, Yoshua Bengio and Yann LeCun (Eds.). http://arxiv.org/abs/1412.6980

[28] Pantelis Linardatos, Vasilis Papastefanopoulos, and Sotiris Kotsiantis. 2020. Explainable ai: A review of machine learning interpretability methods. *Entropy* 23, 1 (2020), 18.

[29] Tian Liu, Ziyu Liu, Jun Huang, Rui Tan, and Zhen Tan. 2018. Detecting wireless spy cameras via stimulating and probing. In *Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services*. 243–255.

[30] Martin Loesdau, Sébastien Chabrier, and Alban Gabillon. 2014. Hue and saturation in the RGB color space. In *International conference on image and signal processing*. Springer, 203–212.

[31] John Metcalfe. 2013. This Heat-Sensing Helicopter Can Find Your Illicit Marijuana-Grow House. (Oct 2013). https://www.bloomberg.com/news/articles/2013-10-16/this-heat-sensing-helicopter-can-find-your-illicit-marijuana-grow-house

[32] Hacker News. 2019. How to increase your chances of finding a hidden camera? (April 2019). https://news.ycombinator.com/item?id=19602342

[33] Nobuyuki Otsu. 1979. A threshold selection method from gray-level histograms. *IEEE transactions on systems, man, and cybernetics* 9, 1 (1979), 62–66.

[34] B Srinivasa Reddy and Biswanath N Chatterji. 1996. An FFT-based technique for translation, rotation, and scale-invariant image registration. *IEEE transactions on image processing* 5, 8 (1996), 1266–1271.

[35] Christopher Dahlin Rodin, Luciano Netto de Lima, Fabio Augusto de Alcantara Andrade, Diego Barreto Haddad, Tor Arne Johansen, and Rune Storvold. 2018. Object classification in thermal images using convolutional neural networks for search and rescue missions with unmanned aerial systems. In *2018 International Joint Conference on Neural Networks (IJCNN)*. IEEE, 1–8.

[36] Wojciech Samek, Grégoire Montavon, Andrea Vedaldi, Lars Kai Hansen, and Klaus-Robert Müller. 2019. *Explainable AI: interpreting, explaining and visualizing deep learning*. Vol. 11700. Springer Nature.

[37] Sriram Sami, Sean Rui Xiang Tan, Bangjie Sun, and Jun Han. 2021. LAPD: Hidden Spy Camera Detection using Smartphone Time-of-Flight Sensors. In *Proceedings of the 19th ACM Conference on Embedded Networked Sensor Systems*. 288–301.

[38] Jeff Sauro. 2011. Measuring Usability with the System Usability Scale (SUS). (Feb 2011). https://measuringu.com/sus/

[39] Jeff Sauro and James R Lewis. 2016. *Quantifying the user experience: Practical statistics for user research*. Morgan Kaufmann.

[40] Ramprasaath R Selvaraju, Michael Cogswell, Abhishek Das, Ramakrishna Vedantam, Devi Parikh, and Dhruv Batra. 2017. Grad-cam: Visual explanations from deep networks via gradient-based localization. In *Proceedings of the IEEE international conference on computer vision*. 618–626.

[41] Akash Deep Singh, Luis Garcia, Joseph Noor, and Mani Srivastava. 2021. I Always Feel Like Somebody's Sensing Me! A Framework to Detect, Identify, and Localize Clandestine Wireless Sensors. In *30th {USENIX} Security Symposium ({USENIX} Security 21)*.

[42] ETS Solution. 2020. Optimize Thermal Performance of Power Adapters with Simcenter Flotherm. (December 2020). https://www.etssolution-asia.com/blog/optimize-thermal-performance-of-power-adapters-with-simcenter-flotherm

[43] Spygadgets. 2022. Spy Gadgets - Counter Surveillance. (2022). https://www.spygadgets.com/counter-surveillance/

[44] Staff. 2018. South Korean women turn out in their thousands to protest against widespread spycam porn crimes. *The Telegraph* (Jul 2018). https://www.telegraph.co.uk/news/2018/07/07/south-korean-women-turn-thousands-protest-against-widespread/

[45] SSI Staff. 2021. The company's AI thermal cameras can be used to pre-screen individuals for elevated temperatures and weapons prior to entering facilities. (Jan 2021). https://www.securitysales.com/surveillance/athena-security-occupancy-tracking-concealed-gun-detection/

[46] Christian Szegedy, Vincent Vanhoucke, Sergey Ioffe, Jon Shlens, and Zbigniew Wojna. 2016. Rethinking the inception architecture for computer vision. In *Proceedings of the IEEE conference on computer vision and pattern recognition*. 2818–2826.

[47] R Vadivambal and Digvir S Jayas. 2011. Applications of thermal imaging in agriculture and food industry—a review. *Food and bioprocess technology* 4, 2 (2011), 186–199.

[48] Bernard L Welch. 1947. The generalization of 'STUDENT'S'problem when several different population varlances are involved. *Biometrika* 34, 1-2 (1947), 28–35.

[49] Thomas Wiegand, Gary J Sullivan, Gisle Bjontegaard, and Ajay Luthra. 2003. Overview of the H. 264/AVC video coding standard. *IEEE Transactions on circuits and systems for video technology* 13, 7 (2003), 560–576.

[50] Sanghyun Woo, Jongchan Park, Joon-Young Lee, and In So Kweon. 2018. Cbam: Convolutional block attention module. In *Proceedings of the European conference on computer vision (ECCV)*. 3–19.

[51] Zhiyuan Yu, Zack Kaplan, Qiben Yan, and Ning Zhang. 2021. Security and privacy in the emerging cyber-physical world: A survey. *IEEE Communications Surveys & Tutorials* 23, 3 (2021), 1879–1919.