Advanced Techniques for Preventing Thermal Imaging Attacks

NORAH ALOTAIBI, MD SHAFIQUL ISLAM, KAROLA MARKY, and MOHAMED KHAMIS,

University of Glasgow, UK

Thermal cameras can be used to detect user input on interfaces, such as touchscreens, keyboards, and PIN pads, by recording the heat traces left by the users' fingers after interaction (e.g., typing a message or entering a PIN) and using them to reconstruct the input. While previous work mitigated the thermal attacks by complicating input or distorting heat traces, our research is the first to propose *preventing* thermal attack using deep learning (DL) techniques to prevent malicious use of thermal cameras. Our DL models detect interfaces in the thermal camera feed and then obfuscate heat traces on them. Our preliminary findings show that the proposed framework can detect interfaces and eliminate authentication information from thermal images. At the same time, our methods still reveal if an interface has been interacted with. Thus, our approach improves security without impacting the utility of the thermal camera.

CCS Concepts: • **Computer systems organization** → **Human Computer Interaction**; *IoT*; Security and Privacy; side channel attack.

Additional Key Words and Phrases: Thermal Imaging Datasets; Deep Learning; Image Obfuscation; Prevent Thermal Attacks

ACM Reference Format:
Norah Alotaibi, Md Shafiqul Islam, Karola Marky, and Mohamed Khamis. 2022. Advanced Techniques for Preventing Thermal Imaging Attacks. In *27th International Conference on Intelligent User Interfaces (IUI '22 Companion), March 22–25, 2022, Helsinki, Finland.* ACM, New York, NY, USA, 5 pages. https://doi.org/10.1145/3490100.3516472

1 INTRODUCTION

Thermal imaging cameras are becoming cheaper and more easily accessible to the public. They have many applications, such as building inspection, animal tracking, and more recently for identifying people with abnormal body temperatures in public spaces, which could help reduce COVID-19 cases. However, the ubiquity of this technology also brings a threat to security and privacy. For example, a low-cost thermal camera (<£200 [13]) can be used to infer input provided on touchscreens and keyboards in what is called a thermal attack [2, 3].This creates a new front of less-explored attacks that put user privacy and security at risk as attackers can use thermal cameras to infer the user input, which could involve login credentials, private messages, or sensitive information. Therefore, there is a need to develop frameworks to prevent thermal attacks.

Recent studies on thermal attacks focused on understanding the viability of the threat or developing basic solutions, e.g., touching the screen after PIN entry [1]. This demo presents the first implementation of an approach that prevents malicious use of thermal cameras. This is done by detecting interfaces, such as keyboards and touchscreens, in the thermal camera feed, and then obfuscating any heat traces using filters.

To achieve this, we developed a deep learning (DL) framework to detect user interfaces in the feed of thermal cameras and prevent users from viewing heat traces on these interfaces by obfuscating them. For this, we implemented four obfuscation methods: 1) blurring, 2) masking, 3) pixelation and 4) differential privacy. This demonstration presents: 1)

Manuscript submitted to ACM

Our implementation of DL models that identify user interfaces in thermal images. 2) Our implementation of image obfuscation methods that remove heat traces from the detected interfaces. 3) Our plan for a user study to verify that our approach ensures security by preventing thermal attacks without negatively impacting the utility of the camera.

2 RELATED WORK

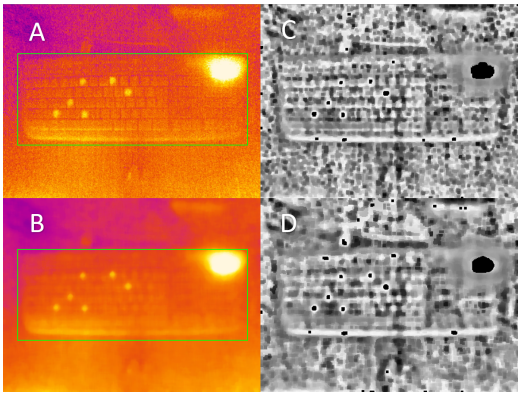
Previous work is rather scarce despite the relevance of thermal attacks and their significant impact on user privacy and security. Based on the threat model, work in this area can be divided into two categories: 1) thermal attacks in which the attacker used an automated approach to analyze the thermal images, and 2) attacks in which the attacker visually inspected the thermal images to determine the input.

Mowery et al. published the earliest work on thermal attacks [12]. They investigated the effectiveness of thermal attacks on ATM keypads using both an automated and a manual approach. According to the study results, while both strategies successfully retrieved passwords, the automated one was more effective. Visual inspection recovered 20-30% of codes after a minute, whereas the automated technique recovered around 50%. Further research by Li et al. investigated thermal attacks on ATM keypads [6]. They determined the entries and their order using a method based on frame-by-frame comparison at different time intervals, which allowed cracking 6-digit PINs with 26.7% accuracy. Another stream of research investigated thermal attacks on common computer keyboards [5]. The system *Thermanator* used a blob detection technique to separate the thermal traces from the background. Entire sets of keypresses could be recovered as late as 30 seconds after initial password entry, while partial sets can still be viable to recover after one minute. A study by Abdelrahman et al. investigated the effectiveness of thermal attacks on user authentication on mobile devices [1]. They developed *ThermalAnalyzer*, which featured a recognition pipeline that reconstructs PINs and Android Lock Patterns by analyzing thermal images.The approach used blob detection and the mean temperatures of the heat traces in the regions of interest to determine the input and order of entry. The correct order of PIN/pattern was obtained by calculating the mean temperature of the detected region of interest (i.e., thermal traces) then sorting them based on their weights. They achieved an overall accuracy of 78% when attacking PINs within 30 seconds of entry and 38.89% when attacking patterns. Attack accuracy was 100% against patterns that do not contain any overlapping input.

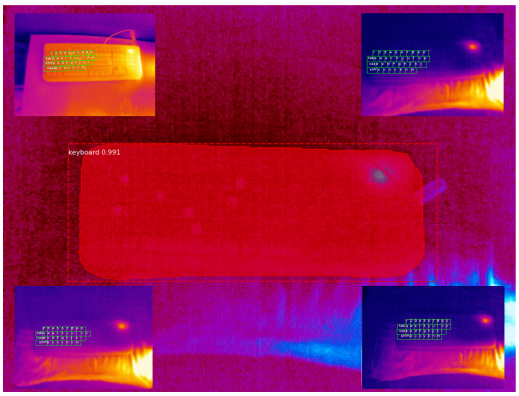
While previous work discussed automated methods for analysing thermal images, other works investigated how well visual inspection of thermal images can reveal input. Wodo and Hanzlik's research presented several scenarios that simulated thermal attacks on computer keyboards, cash machines, digital door locks, and payment terminals [14]. In these scenarios, the majority of user passwords were successfully retrieved within the first 40 seconds. Abdrabou et al. looked into thermal attacks on touch gestures and taps on smartphone touchscreens and laptop touchpads [2]. Study participants were asked to recover graphical passwords from smartphones and laptop touchpads using taps and gestures. The results revealed that touch gestures are more vulnerable than tapping on touchscreens/touchpads (60.65% vs. 23.61%) and that touchscreens are more vulnerable than touchpads (87.04% vs. 56.02%).

3 PREVENTING THERMAL ATTACKS - CONCEPT AND IMPLEMENTATION

This section summarizes the concept and implementation of the proposed techniques for preventing thermal attacks. We designed a DL model to detect user interfaces, such as keyboards, followed by obfuscating the heat traces. In this work, we intend to analyze four major obfuscation approaches: 1) blurring, 2) masking, 3) pixelation, and 4) differential privacy. The first three approaches are extensively used and proven to effectively preserve privacy in RGB images [4, 7]. While these algorithms appear to be effective to the naked eye, they can be reversed [11]. Motivated by this challenge, we implemented a robust algorithm (i.e., differential privacy) which is based on the pixelization concept.



(a) Input Reconstruction



(b) Detection Results

Fig. 1. (a) Example of heat traces located using the Mask-RCNN model: A is the original thermal image with no obfuscation methods applied, B is the blurred image, C is the thermal traces obtained using A, and D is the detected thermal traces using the blurred image B. We can see that A and B both revealed the exact information that can be used to reconstruct the input. (b) The Mask-RCNN detection results where the detected keyboard's coroner coordinates are utilized to deduce the keys used to construct the input.

This technique takes the color held within a pixel, and adds noise to the dominant RGB value acquired in that pixel. This noise is computed as a random value drawn from a Laplace distribution. Since blurring is the most commonly used obfuscation technique to limit information content disclosure in both research and practice [8], it was our first technique to experiment with. According to our preliminary findings, blurring may not provide adequate privacy protection. Even when the thermal traces are blurred, as shown in Figure 1a, we can still obtain the exact information needed to obtain the password from the thermal image.

The implementation of the proposed methodology is summarized as shown in Fig. 2. We recorded 1500 thermal images of keyboards using an Optris pi 450 thermal camera to develop the DL model. The dataset contained images of keyboards with heat traces. Because DL models are extensively used for object detection from images in computer vision tasks, we used two state-of-the-art families of object detection algorithms: 1) region-based convolution neural network (Mask R-CNN), and 2) single-shot detection (SSD). We adopted both algorithms due to their advantages of less computation and improved accuracy [10]. The developed Mask R-CNN utilizes Resnet101 network as backbone and pre-trained weights from the common objects in context (COCO) dataset [9], while the SSD uses VGG-16 as a backbone architecture.

Our preliminary results, as shown in Fig. 1a and Fig.1b, reveal that both detection models are able to detect interfaces, and the Mask R-CNN model achieved the lowest train and validation losses of 0.109 and 0.146 in classifying interfaces, respectively. When integrating the mentioned obfuscation methods our model has the potential to correctly detect interfaces and eliminate heat traces from the thermal images. The advantage of our approach is that it obfuscates only the detected heat traces rather than the whole thermal image. Therefore, the utility of the thermal camera will be maintained while ensuring the prevention of malicious uses.

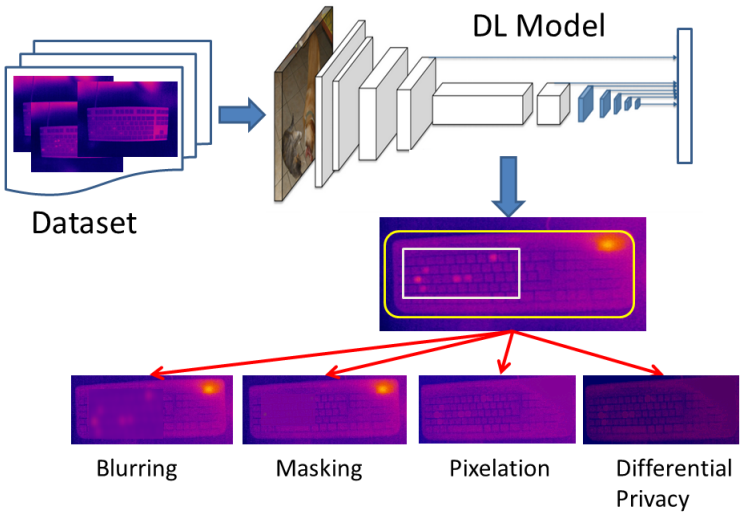


Fig. 2. Proposed DL model-based framework for detecting interfaces and eliminating heat traces from thermal images.

4 CONCLUSION AND OUTLOOK

In this research work, we implemented deep learning models to detect interfaces and heat traces from thermal images. Four image obfuscation methods have been utilized to eliminate heat traces from the detected interfaces. Our preliminary results show that the proposed models can detect interfaces in thermal images and obfuscate the heat traces on them. Thus, the framework has the potential to prevent thermal attacks without greatly impacting the camera's utility.

In the future, we plan to conduct an online user study to evaluate our approach for preventing thermal attacks. We will recruit 20–30 participants using an online platform. After giving their consent, the participants will see thermal images or videos and will be asked to guess the user input. The thermal images or videos will be recorded using an Optris pi 450 (764px × 480px, 80 Hz, 40mK NETD, -20°C to 2450°C), and will be positioned to record a standard computer keyboard. Each participant must complete a fixed set of tasks. In each task, a different obfuscation technique will be applied. Participants will have a field in which they can provide their guesses.

The proposed approach will be evaluated by measuring the similarity between the participants' guesses and the actual user input to assess the effectiveness of obfuscation. The success rate of the thermal attack is determined by how close the guesses by the participants are to the original user input. We hypothesize that after interface and heat trace detection through applying DL models followed by image obfuscation, the participants will be less successful in identifying the user input in the thermal images. The following step will be to integrate our approach into thermal cameras.

ACKNOWLEDGMENTS

This work has been supported by the PETRAS National Centre of Excellence for IoT Systems Cybersecurity, which has been funded by the UK EPSRC under grant number EP/S035362/1, an EPSRC New Investigator award (EP/V008870/1), Taif University and Royal Embassy of Saudi Arabia Cultural Bureau in London, and the Royal Society of Edinburgh (Award number 65040).

REFERENCES

[1] Yomna Abdelrahman, Mohamed Khamis, Stefan Schneegass, and Florian Alt. 2017. Stay cool! understanding thermal attacks on mobile-based user authentication. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 3751–3763.

[2] Yasmeen Abdrabou, Yomna Abdelrahman, Ahmed Ayman, Amr Elmougy, and Mohamed Khamis. 2020. Are Thermal Attacks Ubiquitous? When Non-Expert Attackers Use Off the Shelf Thermal Cameras. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Salerno, Italy) *(AVI '20)*. Association for Computing Machinery, New York, NY, USA, Article 47, 5 pages. https://doi.org/10.1145/3399715.3399819

[3] Yasmeen Abdrabou, Reem Hatem, Yomna Abdelrahman, Amr Elmougy, and Mohamed Khamis. 2021. Passphrases Beat Thermal Attacks: Evaluating Text Input Characteristics Against Thermal Attacks on Laptops and Smartphones. In *IFIP Conference on Human-Computer Interaction*. Springer, 712–721.

[4] Rakibul Hasan, Eman Hassan, Yifang Li, Kelly Caine, David J Crandall, Roberto Hoyle, and Apu Kapadia. 2018. Viewer experience of obscuring scene elements in photos to enhance privacy. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*. 1–13.

[5] Tyler Kaczmarek, Ercan Ozturk, and Gene Tsudik. 2019. Thermanator: Thermal Residue-Based Post Factum Attacks on Keyboard Data Entry. In *Proceedings of the 2019 ACM Asia Conference on Computer and Communications Security* (Auckland, New Zealand) *(Asia CCS '19)*. Association for Computing Machinery, New York, NY, USA, 586–593. https://doi.org/10.1145/3321705.3329846

[6] Duo Li, Xiao-Ping Zhang, Menghan Hu, Guangtao Zhai, and Xiaokang Yang. 2018. Physical password breaking via thermal sequence analysis. *IEEE Transactions on Information Forensics and Security* 14, 5 (2018), 1142–1154.

[7] Tao Li and Min Soo Choi. 2021. DeepBlur: A simple and effective method for natural image obfuscation. *arXiv preprint arXiv:2104.02655* 1 (2021).

[8] Yifang Li, Nishant Vishwamitra, Bart P Knijnenburg, Hongxin Hu, and Kelly Caine. 2017. Effectiveness and users' experience of obfuscation as a privacy-enhancing technology for sharing photos. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–24.

[9] Tsung-Yi Lin, Michael Maire, Serge Belongie, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C Lawrence Zitnick. 2014. Microsoft coco: Common objects in context. In *European conference on computer vision*. Springer, 740–755.

[10] Wei Liu, Dragomir Anguelov, Dumitru Erhan, Christian Szegedy, Scott Reed, Cheng-Yang Fu, and Alexander C Berg. 2016. Ssd: Single shot multibox detector. In *European conference on computer vision*. Springer, 21–37.

[11] Richard McPherson, Reza Shokri, and Vitaly Shmatikov. 2016. Defeating Image Obfuscation with Deep Learning. arXiv:1609.00408 [cs.CR]

[12] Keaton Mowery, Sarah Meiklejohn, and Stefan Savage. 2011. Heat of the moment: Characterizing the efficacy of thermal camera-based attacks. In *Proceedings of the 5th USENIX conference on Offensive technologies*. 6–6.

[13] Thermal Camera 2021. *Affordable thermal camera on amazon.* Retrieved Dec 12, 2021 from https://www:amazon:co:uk/dp/B07CMDCZGV

[14] Wojciech Wodo and Lucjan Hanzlik. 2016. Thermal Imaging Attacks on Keypad Security Systems.. In *SECRYPT*. 458–464.