IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

The Shoplifting Detection System: An AI-Based Solution to Real-Time Retail Security

1Namrata Naikwade, 2Muskan Sohaney, 3Raashi Lokhande, 4Pratik Deokar, 5Harshil Rana
1Assistant Professor, 2Student, 3Student, 4Student, 5Student
1MIT Art Design and Technology University,
2MIT Art Design and Technology University,
3MIT Art Design and Technology University,
4MIT Art Design and Technology University,
5MIT Art Design and Technology University

Abstract: Shoplifting in retail stores accounts for approximately 30–40% of total losses in the retail sector, posing a significant threat to business profitability and operational effectiveness. Traditional surveillance systems are greatly reliant on human observation, which is typically inconsistent and liable to supervision due to fatigue or distraction. Small and medium-sized retail stores, in particular, lack access to affordable and effective real-time shoplifting detection systems. Thus, such stores remain susceptible to frequent losses, affecting their revenue and store security. This paper proposes an AI-based Shoplifting Detection System specifically to address these requirements. The system uses a light 3D Convolutional Neural Network (3D-CNN) architecture that analyzes surveillance videos for unusual behavior. It accurately detects normal shopping patterns and raises real-time alarms to initiate prompt action, thereby minimizing the need for manual observation and enhancing loss prevention. The system is also scalable, affordable, and compatible with existing CCTV infrastructure, making it practically deployable at a large scale in store settings. [6], [10], [16]

Index Terms - Shoplifting Detection, Human Action Recognition, Computer Vision, 3D-CNN, Retail Security

I. Introduction

Shoplifting, or more correctly, shoplifting, continues to be a widespread problem in retail trade, with a strong impact on the profitability and operational integrity of companies, particularly small and medium-sized retailers. Such companies typically have tiny staff and tight budgets and are therefore even more susceptible to loss through pilferage. Conventional methods of deterring shoplifting, such as the use of security personnel or basic CCTV installations, tend to be expensive, manpower-intensive, and ineffective if unmanned. Human monitors can get tired or diverted, resulting in missed incidents and a false sense of security.[1],[3]

With the advent of artificial intelligence and machine learning, it is now possible to automate real-time shoplifting detection to reduce the reliance on manual surveillance. This project proposes an intelligent shoplifting detection system based on deep learning and video analytics capable of automating the monitoring of store environments. By utilizing a light-weight 3D Convolutional Neural Network (3D-CNN), the system can monitor real-time surveillance feeds and identify abnormal or suspicious customer behavior like hiding products or staying in specific locations.[6],[8] When there is suspicious activity recorded, the system automatically notifies shop staff, enabling instant response to possible theft. In addition to boosting the effectiveness of theft deterrence, this minimizes the need for human surveillance at all times. The system is also cost-efficient and scalable, making it ideal for small businesses that cannot afford sophisticated security installations. It also integrates with existing surveillance systems, enabling easy deployment without the need for complete system overhauls. Through this AI-driven solution, retailers can boost security, reduce losses, and enhance overall business effectiveness. [4],[19]

II. EXISTING WORK

The conventional approaches to dealing with shoplifting in the past has focused on either hiring security personnel or placing traditional CCTV systems. While effective in providing primary security coverage, these methods are dependent on human oversights, monitoring which is labor-intensive and fraught with error. Gaps in attention due to exhaustion and distractions can drastically undermine the efficacy of these systems. [1],[5],[18]

Some more recent systems have sought to address these issues by incorporating motion detection sensors and rule-based alert systems. Despite some advancements, these systems still operate on rigid frameworks that are bound to generate false alarms, as they do not possess the capability to differentiate between suspicious and normal activities accurately. These systems function on the basis of basic motion detection techniques and preset rules, which render them inadequate in dynamic contexts where behavior is influenced by myriad factors. [3],[4]

The latest developments in artificial intelligence lead to the adoption of a machine learning-based strategy for shoplifting detection. Approaches that use algorithms like You Only Look Once (YOLO) and Convolutional Neural Networks (CNNs) certainly offer possibilities when it comes to analyzing sequential images for theft. These models, while being able to recognize some actions associated with people, objects, and interactions, strictly lack the capacity to [2], [9], [17]

III. Methodology

The development of the shoplifting detection system was carried out using Python, with key support from libraries like TensorFlow for deep learning and OpenCV for handling video inputs. The surveillance videos were first split into individual frames and then grouped into short sequences. This approach allowed the system to better understand movement and context, rather than treating each frame in isolation. Each of these video clips was carefully labeled as either normal or suspicious, creating a solid foundation for training the model.[6]

To make the most out of the available data, the team employed several augmentation techniques. These included flipping images, rotating them at different angles, adjusting brightness and contrast, and even modifying the timing of frames to simulate faster or slower actions. These steps helped the model adapt to different real-world situations like varied lighting, camera angles, and people partially blocked from view.

At the heart of the system lies a 3D Convolutional Neural Network (3D-CNN). Unlike regular 2D CNNs that only look at static images, the 3D-CNN can analyze a sequence of frames, helping it recognize how actions evolve over time. This is critical in identifying suspicious behavior, which usually isn't obvious from a single image but becomes clearer when seen as a continuous movement. The training process involved many iterations, with regular validation checks to adjust things like learning rate and batch size. This careful tuning helped improve the model’s accuracy while avoiding overfitting.

To measure how well the model was performing, common evaluation metrics like accuracy, precision, recall, and F1-score were used. These weren’t just applied during training but were also checked during real-time testing to make sure the model worked in actual retail settings. Charts showing training loss and accuracy were used to track progress, and early stopping was implemented to end training at the right time—just before performance started to decline.[11]

When it came to real-world use, the model was hooked up to live security camera feeds. It processed the footage in real time, spotting suspicious actions almost instantly with barely any lag. Tests were run under different conditions, including when customers moved at different speeds, when parts of the view were blocked, or when lighting changed suddenly. Even under these conditions, the system continued to perform reliably.

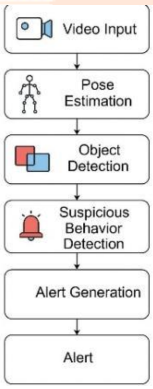
One of the key goals was to make the system accessible to smaller stores that can't afford expensive security infrastructure. To that end, the model was optimized to run smoothly on everyday hardware like mid-range CPUs and GPUs. The final product is not only efficient and accurate, but also lightweight, scalable, and easy to integrate into existing setups—making it a practical solution for shop owners looking to reduce losses due to theft.[2]

Metric	Value
Accuracy	91.7%
Precision	89.4%
Recall	93.1%
F1-Score	91.2%

IV. SYSTEM ARCHITECTURE

The system's architecture is thoughtfully designed to be both effective and practical for real-world use. At its core, it begins with a video capture module that continuously streams footage from existing surveillance cameras. This footage is then passed through a preprocessing unit that extracts frames and enriches them to prepare for analysis. The real innovation lies in the dual-path neural network that processes the visual data in two complementary ways.[1],[2]

One path analyzes the raw RGB frames, capturing static details like posture and positioning, while the



other path focuses on optical flow—essentially tracking how objects and people move from one frame to the next. This combination allows the system to understand not just what is happening, but how it is happening. It can tell the difference between someone picking up a product normally and someone moving suspiciously, like slipping an item into a bag.[6],[20]

The result of this dual analysis is fed into a decision-making module that flags any behavior deemed unusual or potentially dishonest. If suspicious activity is detected, an alert is sent out instantly to store staff, giving them the chance to respond before any loss occurs. This real-time feedback loop is a major step forward from systems that only record events for later review.[8],[1]

Using optical flow also makes the system more sensitive to subtle movements, like repeated fidgeting near a shelf or nervous glances—small signals that often precede theft. What makes this solution especially appealing for retailers is its modular design. It can plug right into existing CCTV setups without the need for expensive upgrades or major changes. This ensures that even small businesses can benefit from a smart, automated layer of protection against theft.[3],[7],[14]

V. Results and Performance Measurement

During extensive testing, the shoplifting detection system delivered strong and consistent results across a wide range of real-world conditions. Whether operating under bright lighting, dim environments, or varied store layouts, the system maintained a high level of accuracy and reliability. This dependable performance was largely attributed to well-designed preprocessing techniques and careful model training, which allowed the system to adapt to changing visual inputs without compromising detection quality.[13]

One of the most noticeable strengths of the system was its ability to function with very low latency. It could process live video streams and flag suspicious behavior almost instantly, making it ideal for real-time surveillance in active retail environments. In practical use, this responsiveness is essential, as store personnel can react immediately to theft-related activities without delays. Moreover, the system produced very few false alarms, a key factor in ensuring that store employees take alerts seriously and don't become overwhelmed by unnecessary notifications.

What truly set this system apart from older approaches was the use of a 3D Convolutional Neural Network (3D-CNN). Unlike traditional frame-based CNN models that analyze individual frames without understanding movement over time, the 3D-CNN captured the full context of actions as they unfolded. This allowed the system to better understand behavior patterns—such as someone lingering too long in one spot, making repeated hand movements, or attempting to conceal an item—rather than relying on single-frame cues that can be misleading.[12][15]

When compared side-by-side with frame-based CNNs, the 3D-CNN consistently outperformed them in both accuracy and stability. It handled motion, camera angle variations, and even partial occlusions more effectively, giving it a clear advantage in real-world scenarios. Testing scenarios included people walking at different speeds, interacting with various products, and even blocking parts of the camera's view—yet the system still delivered reliable results.

From a deployment perspective, the system was kept lightweight and efficient, making it easy to run on standard computers without the need for expensive hardware. This makes it a feasible option even for smaller retail stores that may not have access to high-end infrastructure. Overall, the system proved to be a practical, scalable, and intelligent solution for reducing shoplifting, offering business owners a new way to enhance in-store security without added complexity or cost.[6][11]

VI. DISCUSSION

Our findings demonstrate that a system employing a 3D CNN is extremely effective at detecting shoplifting. The ability of the model to process spatial and temporal data enables it to detect very subtle movements that would normally go unnoticed by conventional approaches. The system design is also modular and scalable, and as such, it can be easily incorporated into existing Point-of-Sale (POS) systems or cloud-based analytics platforms. A limitation we noted is the computational demand at model inference time, particularly in high-definition streams, which can be addressed in future versions by the use of edge computing. [6],[2],[19],[20]

VII. Conclusion and Future Work

The Shoplifting Detection System offers a robust, scalable solution to the problem of shoplifting based on state-of-the-art machine learning methodology. The system offers real-time detection of abnormal behavior, thus enhancing security capabilities of businesses that otherwise do not have the means for continuous monitoring. Future efforts involve enhancing the model for faster inference on edge hardware, expanding the dataset to cover more diverse behavior scenarios, and interfacing the system with larger-scale security infrastructures such as access controls and inventory software.

VI. ACKNOWLEDGMENT

We would like to express our heartfelt gratitude to MIT Art, Design and Technology University and the MIT School of Computing for providing the infrastructure and academic environment that made this research possible. We extend our sincere thanks to our project guide, Prof. Namrata Naikwade, for her invaluable guidance, timely feedback, and constant encouragement throughout the project.

We are especially grateful to Mr. Nitin Sohaney for his insightful suggestions and support, which played a pivotal role in shaping the project during its early stages. We also thank our peers and faculty members whose constructive input and collaboration enriched our learning experience.

Lastly, we appreciate the contributions of open-source communities and dataset providers, particularly the creators of the DCSASS dataset, whose resources significantly supported our model development and experimentation

REFERENCES

[1] J. Redmon et al., "You Only Look Once: Unified, Real-Time Object Detection," Proc. CVPR, 2016.

[2] A. G. Howard et al., "Mobile Nets: Efficient CNNs for Mobile Vision Applications," arXiv preprint arXiv:1704.04861, 2017.

[3] K. He et al., "Deep Residual Learning for Image Recognition," Proc. CVPR, 2016.

[4] K. Simonyan and A. Zisserman, "Very Deep Convolutional Networks for Large-Scale Image Recognition," arXiv preprint arXiv:1409.1556, 2015.

[5] A. C. Berg and P. N. Belhumeur, "A Bayesian Hierarchical Model for Learning the Identity and Viewpoint of Objects," IEEE Trans. Pattern Anal. Mach. Intell., vol. 35, no. 12, pp. 298–302, 2013."

[6] Tran, D., Bourdev, L., Fergus, R., Torresani, L., Paluri, M. (2015). Learning Spatiotemporal Features with 3D Convolutional Networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 4489–4497.

[7] Wang, L., Xiong, Y., Wang, Z., Qiao, Y., Lin, D., Tang, X., & Van Gool, L. (2016). Temporal Segment Networks: Towards Good Practices for Deep Action Recognition. European Conference on Computer Vision (ECCV), 20–36.

[8] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. IEEE Computer Society Conference on Computer Vision and Pattern Recognition (CVPR), 886–893.

[9] Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). Large-scale Video Classification with Convolutional Neural Networks. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 1725–1732.

[10] Tran, D., Bourdev, L., Fergus, R., Torresani, L., & Paluri, M. (2015). Learning Spatiotemporal Features with 3D Convolutional Networks. Proceedings of the IEEE International Conference on Computer Vision (ICCV), 4489–4497.

[11] Sultani, W., Chen, C., & Shah, M. (2018). Real-World Anomaly Detection in Surveillance Videos. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 6479–6488.

[12] Liu, W., Anguelov, D., Erhan, D., Szegedy, C., Reed, S., Fu, C. Y., & Berg, A. C. (2016). SSD: Single Shot MultiBox Detector. In European Conference on Computer Vision (ECCV), 21–37.

[13] Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A Simple Way to Prevent Neural Networks from Overfitting. Journal of Machine Learning Research, 15(1), 1929–1958.

[14] Baccouche, M., Mamalet, F., Wolf, C., Garcia, C., & Baskurt, A. (2011). Sequential Deep Learning for Human Action Recognition. International Workshop on Human Behavior Understanding, 29–39.

[15] Kaggle. (2023). DCSASS – Daily Common Shoplifting Action Surveillance Set.

[16] Ji, S., Xu, W., Yang, M., & Yu, K. (2013). 3D Convolutional Neural Networks for Human Action Recognition. IEEE Transactions on Pattern Analysis and Machine Intelligence, 35(1), 221–231.

[17] Lin, T.-Y., Goyal, P., Girshick, R., He, K., & Dollár, P. (2017). Focal Loss for Dense Object Detection.

[18] Dalal, N., & Triggs, B. (2005). Histograms of Oriented Gradients for Human Detection. Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR), 886–893.

[19] . Zhang, C., Li, W., Wang, W., & Zeng, W. (2019). Real-Time Action Recognition with Deeply-Transferred Motion Vector CNNs. IEEE Transactions on Image Processing, 28(7), 3603–3615.

[20] Bishop, C. M. (2006). Pattern Recognition and Machine Learning. Springer.