Creating an AI Powered Child Monitoring Tool

Mrinall Umasudhan

Abstract Traditional child monitoring systems, commonly referred to as baby monitors, have conventionally relied on providing live video feeds to parents, requiring constant vigilance to ensure a child's well-being. However, this approach poses inherent limitations, often leading to inefficiencies and, more importantly, the potential for harmful incidents. This paper introduces an improved child monitoring system that integrates computer vision and machine learning techniques to ensure the safety of young children in household environments. The YOLO (You Only Look Once) object detection model, initialized with pre-trained weights from the COCO (Common Objects in Context) dataset, plays a central role in the system. This approach allows for accurate and real-time identification of various objects relevant to child safety, ensuring a comprehensive monitoring solution. To assess the system's functionality, we conducted tests in a real-world scenario, featuring a 10-year-old child navigating a kitchen within a four-family household. Results demonstrate the system's efficacy in alerting caregivers to potential risks, surpassing the limitations of conventional monitoring methods. This innovative child monitoring system represents a significant stride towards proactive and adaptive child safety solutions in household settings.

1 Introduction

Ensuring the safety of young children in various environments, particularly when left unattended, is a major issue for parents that has been largely unexplored. Traditional child monitoring systems, commonly referred to as baby monitors, have predominantly relied on providing live video feeds to parents. However, this conventional approach is inherently inefficient and has the potential for harmful incidents. In response to this challenge, this paper addresses the need for an improved child monitoring system that leverages cutting-edge machine learning and computer vision techniques.

The existing methods of live feed monitoring necessitate parents' consistent visual checks, making it susceptible to lapses in attention and delayed response times. The consequences of such inefficiencies are particularly alarming when considering the vulnerability of young children who are left unattended. To address these shortcomings, the proposed child monitoring system integrates a multitude of machine learning and computer vision techniques to provide a comprehensive and continuous assessment of a child's status. By leveraging modern object detection models, such as YOLO (You Only Look Once), the monitoring system ensures constant and precise tracking of the child's movements in real-time. This technological foundation empowers parents with an unprecedented level of awareness and proactive control over their child's safety.

Through an intuitive interface, parents can actively participate in shaping the monitoring process. The system allows parents to select specific regions within the scene, designating areas where the child should not approach, boundaries they should not cross, and the areas where they should not leave. This approach tailors the monitoring system to the unique dynamics of each living space, providing a personalized monitor that adapts to the child's environment.

The incorporation of object detection extends beyond tracking the child's activities, enhancing the system's vigilance to potential environmental threats. In the event that objects with the potential to endanger the child's safety, such as fire, intruders, or animals, appear in the monitoring scene, immediate alerts are sent to the parents. This real-time notification system ensures that parents can respond swiftly and effectively to any unforeseen dangers, further mitigating the risks associated with leaving a child unattended.

This paper outlines the significance of transitioning from passive live feed monitoring to an intelligent system capable of autonomously analyzing and interpreting the child's behavior and surroundings. The approach presented here aims to surpass the limitations of conventional systems by not only detecting potential dangers in real-time but also by significantly reducing the need for manual checks by parents.

Following extensive testing, the results reveal a substantial improvement over the conventional periodic checks performed by parents alone. The proposed system demonstrated an impressive 95 percent accuracy in detecting when a child enters a designated danger zone, showcasing its efficacy in proactive hazard prevention. Moreover, the system consistently identified instances when the child exited the camera's field of view, eliminating blind spots that may exist in traditional monitoring setups.

2 Related Work

While the realm of AI-based child monitoring systems is relatively unexplored, a noteworthy work that bears some similarities to this project is "An Intelligent Baby Monitor with Automatic Sleeping Posture Detection and Notification". This system utilizes machine learning techniques to monitor a baby, primarily focusing on facial expression analysis to alert parents when the baby begins crying. However, this approach is limited in scope, primarily tailored for infants in cribs and concentrating on emotional cues (Khan, 2021).

In contrast, this work takes a more comprehensive approach, extending beyond facial expression analysis to encompass the continuous tracking of a potentially older child as they move and interact with their environment. The system developed in this paper aims to address the dynamic nature of child behavior, ensuring that the child remains within a predetermined safe zone and does not approach potentially hazardous objects. Unlike the aforementioned system, which relies on a close-up view of the baby in its crib, the proposed child monitoring system integrates computer vision techniques to maintain a broader perspective, allowing for continuous monitoring of the child's activities within a larger space.

Moreover, the current work goes beyond individual child behavior tracking by implementing scene monitoring capabilities. This includes the detection and alerting of potential environmental threats, such as sudden fires or other safety hazards. This proactive approach to environmental monitoring distinguishes the proposed system from existing works that primarily focus on the child's internal states.

3 Data

In developing the monitor two crucial elements were required to train and test the AI: video data capturing child activities and the pre-trained weights for the YOLO (You Only Look Once) object detection model.

Complementing the video data, the YOLO object detection model necessitates pre-trained weights for accurate object identification. For this paper, the default weights associated with the COCO (Common Objects in Context) dataset are used. Serving as a comprehensive training set for YOLO, COCO includes a diverse array of objects commonly found in household environments. The YOLO model is initialized with weights trained on the entire COCO dataset, enabling proficient recognition of various objects relevant to child safety monitoring (Lin et al., 2014).

To evaluate the functionality of the child monitoring system, we conducted tests using a specific real-world scenario. The chosen setting was a kitchen within a four-family household, featuring a 10-year-old child moving about. This scenario serves as a practical testbed, allowing us to assess the system's responsiveness, particularly concerning its alert system, in response to the child's movements.

Crucially, no special preprocessing or filtering was required for the YOLO model's pre-trained weights, underlining their adaptability to a broad spectrum of common household objects. The integration of diverse datasets and real-world scenarios underscores the robustness and adaptability of the child monitoring system, positioning it as a comprehensive solution for ensuring the safety of children in various household settings.

Figure 1: Scene used for testing

4 Methods

To address the challenges outlined in the introduction, this approach employs the YOLO (You Only Look Once) object detection model for real-time tracking of the child within the monitoring scene. The decision to employ the YOLO (You Only Look Once) object detection model stems from a careful consideration of the balance between speed and accuracy, critical elements in an efficient child monitoring system. YOLO operates by dividing an image into a grid and predicting bounding boxes and class probabilities for each grid cell. Unlike traditional methods that might scan an image multiple times, YOLO processes the entire image in one pass. This unique approach allows YOLO to achieve remarkable speed, making it well-suited for real-time applications like child monitoring. The YOLO architecture is built upon a convolutional, deep neural network backbone composed of convolutional layers, batch normalization, and shortcut connections. It utilizes a series of convolutional blocks, each containing multiple convolutional layers, to extract hierarchical features from the input image. These features are then used to predict bounding boxes, class probabilities, and objectness scores for objects within the image. The deep neural network and convolutional layers, create an architecture that offers a high degree of accuracy in addition to fast processing times (Redmon et al., 2015). While other models, such as Faster RCNN, offer higher accuracy due to its two-stage detection process as opposed to YOLO's single-stage detection, they fall short in processing the numerous frames required for continuous child tracking. YOLO strikes a pragmatic balance, ensuring that parents receive swift notifications in case of potential incidents while maintaining reliable accuracy in detecting various objects relevant to child safety. This speed and accuracy combination positions YOLO as the ideal choice for the child monitoring system. Parents are empowered to draw specific re-

Real-Time Detectors	Train	mAP	FPS
100Hz DPM [31]	2007	16.0	100
30Hz DPM [31]	2007	26.1	30
Fast YOLO	2007+2012	52.7	**155**
YOLO	2007+2012	**63.4**	45
Less Than Real-Time			
Fastest DPM [38]	2007	30.4	15
R-CNN Minus R [20]	2007	53.5	6
Fast R-CNN [14]	2007+2012	70.0	0.5
Faster R-CNN VGG-16[28]	2007+2012	73.2	7
Faster R-CNN ZF [28]	2007+2012	62.1	18
YOLO VGG-16	2007+2012	66.4	21

Figure 2: In terms of real-time performance, YOLO trades a slight decrease in accuracy for a significant speed advantage compared to Faster R-CNN (Liu, 2023).

gions within the scene where the child should not approach and where they should not exit. Despite considering automating the generation of danger zones, it became evident that manual input was crucial. This approach respects parental preferences, allowing them to highlight areas that may not be inherently dangerous but hold potential risks. By leveraging parental insights into their home environment, we overcome limitations that automated tracking may face, such as partial occlusions hindering the YOLO algorithm's detection capabilities. Thus, a balanced synergy is achieved between AI-driven tracking and parental control.

The monitoring process unfolds through a series of steps. Initially, the system checks the detection list returned by the YOLO algorithm to confirm the child's presence in the scene. Subsequently, it scans for potentially dangerous objects, predefining this list by identifying objects in the COCO dataset that pose a risk to a child. The system then ensures the child is not within any predefined danger zones, verifying that the child's bounding box does not overlap or fall within any danger zone bounding boxes. Lastly, the system guarantees the child is within at least one designated safe zone by checking that the child's bounding box is completely contained within a safe zone.

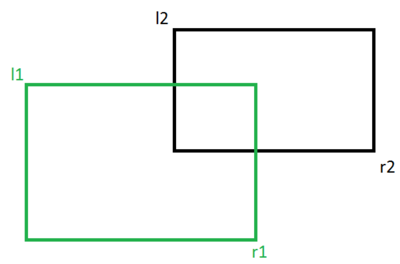

Figure 3: Danger zone alert condition

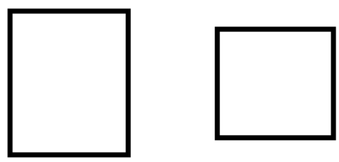

Figure 4: Safe zone alert condition

In an effort to make child monitoring effective and affordable, the hardware used to conduct this study was purposefully not cutting edge. The algorithm was run on a Lenovo ThinkPad with an i7 processor and integrated graphics, paired with a Logitech webcam boasting 720p resolution. Simpler and more budget-friendly hardware was chosen to ensure that the end product would be reasonably priced for parents from all economic backgrounds.

This comprehensive methodology, encompassing manual zone definition, object detection, and bounding box checks, significantly enhances safety compared to periodic manual checks conducted solely by parents. The integration of these methods provides a holistic child monitoring solution, leveraging the strengths of both AI-driven automation and human intuition to create a robust system that ensures the safety of children in diverse household environments.

5 Results

To comprehensively evaluate the efficacy of the child monitoring system, a series of experiments were meticulously designed and conducted within a simulated family household setting, specifically focusing on the kitchen environment. The selected test subject, a 10-year-old child, served as a representative model for real-world scenarios where children might be left unattended. The experimental design incorporated two distinct stages, introducing variability by having the child traverse the monitoring scene at both a regular walking pace and a faster running speed.

A comprehensive set of five criteria was systematically employed to gauge the system's performance across various aspects. These criteria included the accuracy percentage of child detection, the system's capability to accurately identify instances of the child entering a predefined danger zone, accuracy in detecting when the child exited a designated safe zone, the system's proficiency in autonomously flagging objects with potential danger, and accuracy in detecting when the child left the monitoring scene.

The obtained results from these experiments provided nuanced insights into the system's performance under different conditions. Remarkably, at a regular walking pace, the system exhibited an impressive 99% accuracy in detecting the child, underscoring its proficiency in routine scenarios. However, as the child's movement speed increased to a running pace, the detection accuracy experienced a slight reduction to 80%, revealing a potential challenge in accurately capturing high-velocity movements. In examining the system's

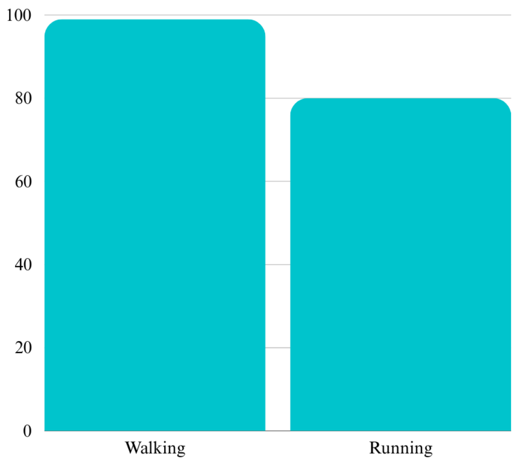

Figure 5: Child detection accuracy percentage

ability to identify the child entering a danger zone, the results showcased a robust performance with a high accuracy of 95% at a walking pace. Nevertheless, as the child's movement speed escalated to a running pace, the accuracy slightly diminished to 85%. A similar trend was observed in the accuracy of detecting the child exiting a designated safe zone, with 98% accuracy at a walking pace and 96% at a running speed.

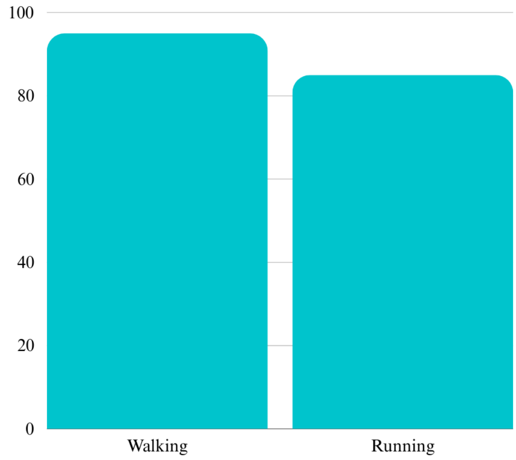

Figure 6: Danger zone detection accuracy percentage

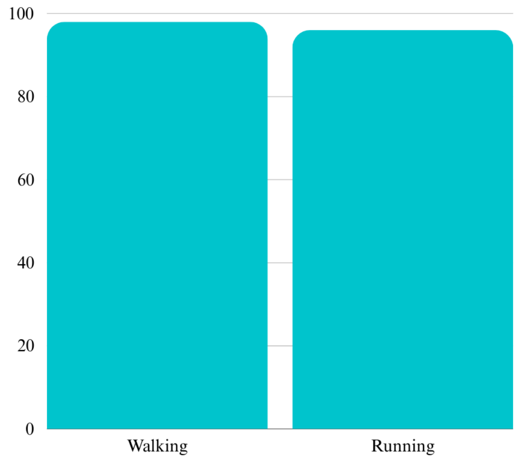

Figure 7: Safe zone detection accuracy percentage

Despite these commendable achievements, challenges surfaced in scenarios involving high-speed movements and partial occlusions. The system demonstrated a discernible decrease in accuracy when the child moved at a running pace, attributed to motion blur effects. Instances of partial occlusion, such as the child running behind a table, posed challenges for the model, indicating a potential area for improvement.

Addressing these challenges requires a nuanced approach, possibly through the incorporation of advanced computer vision techniques, such as multi-view tracking. While the system exhibited commendable performance in detecting various scenarios, these identified areas for improvement serve as valuable insights for refining the model and enhancing its robustness in real-world monitoring scenarios.

6 Conclusion

In summary, the experimentation and evaluation of the AI-powered child monitoring system have yielded compelling results, demonstrating its efficacy in notifying parents of potential dangers and surpassing traditional methods reliant on periodic live feed checks. The system proves particularly adept at detecting and alerting parents when their child enters a potentially hazardous situation. However, certain limitations have been identified, including a decrease in accuracy during high-speed movements of the child and instances of partial occlusion in the scene.

The key takeaway is that while the system excels in routine monitoring scenarios, addressing challenges related to rapid movements and partial occlusions presents opportunities for refinement and extension. To enhance the system's capabilities, one potential extension involves the incorporation of multiple cameras, offering varied perspectives of the scene. This addition ensures minimized occlusion of objects from the monitor's view and maintains continuous visibility of the child.

Furthermore, avenues for improving the robustness of the monitoring system and providing additional information to parents are worth exploring. Integrating a large language model with a vision component, such as GPT, could offer insights into the child's sentiment, enabling parents to receive alerts when the child exhibits signs of distress, such as crying. Additionally, leveraging the language model to automatically flag dangerous objects in the scene may enhance accuracy compared to utilizing pre-defined COCO labels.

While the primary focus of this paper is child monitoring, the concepts and enhancements proposed have the potential for broader applications in surveillance. Extending these ideas to general surveillance scenarios opens new possibilities for improving safety and situational awareness beyond the realm of child monitoring. As technology continues to evolve, these extensions and adaptations hold promise for creating more robust and versatile monitoring systems for diverse applications.

References

Khan, Tareq. 2021. An intelligent baby monitor with automatic sleeping posture detection and notification. *AI*, 2(2):290–306.

Lin, Tsung-Yi, Michael Maire, Serge J. Belongie, Lubomir D. Bourdev, Ross B. Girshick, James Hays, Pietro Perona, Deva Ramanan, Piotr Dollár, and C. Lawrence Zitnick. 2014. Microsoft COCO: common objects in context. *CoRR*, abs/1405.0312.

Liu, Ruiqi. 2023. Deep learning for detecting quilt status of sleeping children. *Highlights in Science, Engineering and Technology*, 39:673–680.

Redmon, Joseph, Santosh Kumar Divvala, Ross B. Girshick, and Ali Farhadi. 2015. You only look once: Unified, real-time object detection. *CoRR*, abs/1506.02640.