

Smart Task Automation System

1. Diksha Kumari	2. Surya Srinija	3. Bejon Sarkar Bappy	4. Hemant Singh	4 Ms. Priyanka Nanda
Department: AIT-CSE	Department:AIT-CSE	Department:AIT-CSE	Department:AIT-CSE	Assistant Professor
Chandigarh University	Chandigarh University	Chandigarh University	Chandigarh University	Department:(AIT-CSE)
Mohali, Punjab	Mohali, Punjab	Mohali, Punjab	Mohali, Punjab	Chandigarh University
				Mohali, Punjab

Abstract—In service-oriented workplaces like coffee shops, making sure employees are productive isn't just a nicety; it's vital for ensuring operational efficiency and fair compensation. Learning how to supervise effectively in this way is a modern managerial challenge. Typical strategies of oversight are manual, clunky, and prone to human error, and they don't always give a clear picture of what's going on. The automated system proposed here, which uses a combination of common video surveillance techniques and machine learning (a.k.a. "artificial intelligence" or AI), is designed to achieve several aims: Use the CCTV cameras that many coffee shops already have to make the action of using the cameras more palatable to employees; reduce the human sources of error in both surveillance and payroll; and, finally, keep track of how many specific work activities are performed by how many specific workers at what times, to make asking and answering the question "Who's been busy?" more efficient.

Keywords—Object Detection, Workplace Automation, Employee Productivity, Computer Vision, Deep Learning, Ethical AI

I. INTRODUCTION

So, this study is looking into an automated object detection system. The goal? To improve how we keep an eye on employee productivity in coffee shops. Now, you know how it usually goes—most of the time, people rely on human observation or even have employees report their own activities. But let's be honest, that can get a little dicey in terms of reliability, right?

It's kind of fascinating to think about how tech can shake things up a bit. I mean, moving away from those traditional methods could really give us a clearer view of what's happening during those crazy busy shifts. Just picture all the possibilities! But, let's face it—those old-school methods can be super subjective, which brings along issues like bias and inaccuracies. This, in turn, can really throw a wrench into how we assess workloads, and honestly, it might even mess up wage calculations too [1].

With more and more CCTV cameras popping up in commercial spaces, there's a chance to incorporate some fancy computer vision tech that allows for real-time, objective task tracking. What's cool about this proposed system is that it uses YOLOv8 object detection [5] alongside DeepSORT tracking [7] to automatically spot, classify, and keep tabs on specific activities—like making coffee—right in designated work areas. They've got this whole process for image preprocessing,

annotation, and tracking spatial regions, which means they can accurately identify those repetitive task cycles and connect them to the right worker IDs [2].

Alright, let's dive into the results. The system really upped its game, achieving a detection accuracy of 92.6%. Plus, it slashed the need for hands-on human supervision by a whopping 54%. And guess what? It managed all this without compromising the reliability of the data collected [3]. Pretty impressive, right?

This is actually a big deal. What it means for businesses is that they can save money, streamline operations, and, honestly, they don't have to work as hard to keep everything running smoothly. And here's the interesting part: this system gives clear, data-driven insights into how tasks are being performed. This really helps ensure that everyone is paid fairly for their contributions. It's a major win for accountability within the team, wouldn't you agree?

They also took some ethical considerations into account. They've put in place privacy-preserving measures like data anonymization and secure access controls to make sure that employee monitoring stays within the bounds of data protection standards [4]. All in all, this framework adds to the growing conversation around AI in workplace optimization. It also opens the door to a scalable and flexible solution that could be applied in other task-heavy fields like restaurants, fast-service retail, and even assembly lines in industries. And, hey, it's also vital to acknowledge the growing worries around workplace surveillance. This research highlights the need for ethical AI practices. The framework proposed here includes solid data protection measures, techniques for anonymizing data, and transparency protocols to make sure that surveillance tech is used responsibly. Balancing tech efficiency with ethical standards not only builds trust among employees but also makes it easier to get these systems up and running in real-world scenarios.

In the end, this study pushes forward the use of AI-driven object detection in tracking employee productivity. It also adds to the larger discussion about digital transformation in industries that rely heavily on labor. Plus, it sets the stage for applying similar solutions in retail, hospitality, healthcare, and other service sectors where keeping an accurate eye on tasks and ensuring fair pay are crucial for both sustainable growth and employee satisfaction.

II. EASE OF USE

A. User-Friendly and Efficient System Integration

So, let's talk about this new productivity monitoring system that uses object detection—it's got some really great features. First off, one of its biggest advantages is how easy it is to set up and use. Seriously, whether you're a tech whiz or just someone who's a bit more on the casual side, you won't need much training to get the hang of it.

What's cool is that it leverages CCTV cameras, which, let's be honest, are pretty much everywhere in coffee shops and retail stores. This means you don't have to make any big changes to your store's layout, saving you money and keeping your day-to-day operations running smoothly. The system uses YOLOv8 for detecting objects and SORT for real-time tracking—pretty slick, right? Plus, it has this straightforward interface that automatically processes video feeds and pulls out all the important task-related info without you having to lift a finger.

And speaking of ease, the dashboard is super user-friendly. It shows real-time stats like how employees are doing, how often tasks are being completed, and gives performance summaries in a way that's visual and easy to grasp. Managers can quickly make sense of the data and make decisions—no need for a PhD in AI or data science here! On top of that, the automated alerts and reporting features help lighten the load for supervisors, cutting down on the need for constant monitoring.

Another thing to note? This system is all about being scalable and flexible. Whether you're running a cozy little coffee shop or managing a big chain, it adapts easily. You can add new workstations or roles to the detection model without much hassle, and updates can be done remotely, which keeps everything running smoothly over time.

In short, the whole idea behind this system is to create a plug-and-play solution that fits right into existing business setups. It's a powerful tool, yet it's designed to be accessible for anyone looking to boost productivity in a modern way.

III. LITERATURE REVIEW

So, traditionally speaking, tracking productivity in the workplace has mostly been about manual monitoring, time logs, or folks just reporting on themselves. And, let's be honest, these old-school methods can be pretty inaccurate. There's a lot of room for human bias and inconsistencies in how data gets recorded [1]. In fast-paced service industries, like food and drinks—think cafe's or quick-service joints—these traditional systems really don't cut it. They often lack the real-time, objective performance metrics that truly show how tasks are being executed [2].

Now, with the rise of Convolutional Neural Networks (CNNs), we've seen object detection step into the spotlight for various real-world applications. There are some well-known architectures in this area, like Faster R-CNN, SSD, and YOLO (which stands for You Only Look Once). Each one has its own balance of detection accuracy and how efficiently it runs.

[3]. Among these, the YOLO series has really taken off, especially for real-time situations because it strikes a nice balance between speed and accuracy.

Take YOLOv5, for example. Earlier on, it proved pretty effective at spotting task-level actions in places like manufacturing and retail. Up until October 2023, researchers noted that YOLOv5 had accuracy levels over 85

Then we've got YOLOv8, which has recently made waves as a big step forward. It boasts better mean Average Precision (mAP) scores, faster inference speeds, and upgrades in its architecture that make it suitable for crowded, complex spaces [5][6]. Studies, including one by Smith et al., have shown that YOLOv8 performs well even in high foot traffic environments, which is ideal for recognizing activities in real-time, even when there are several people in the frame [6].

To take things a step further, researchers have started combining object detection with tracking algorithms, which is a game changer for productivity monitoring. DeepSORT is a popular tracking algorithm that often teams up with YOLO models to keep track of identities across frames. This means the system can reliably watch for repeated tasks done by specific individuals [7]. This combination really helps with long-term activity recognition and provides detailed performance analytics.

But, here's where it gets interesting: despite all these advancements, there's still a noticeable gap in research when it comes to applying these models in service-oriented places like coffee shops. A lot of what's been written focuses on industrial automation or security—not really diving into the day-to-day task flows of service staff or the customer-facing aspects of their work. Plus, there are some ethical issues like employee consent, data minimization, and transparency that often get brushed aside. And these are super important in public or semi-public spaces where surveillance is happening [8].

So, this study aims to fill those gaps by using YOLOv8 to keep an eye on real-world coffee-making processes in bustling cafe' environments. The cool part? It makes use of existing CCTV systems to spot and track repetitive tasks—like steaming milk, tamping espresso, or cleaning counters—all without getting in the way of the workflow. Unlike prior research that relied on synthetic or heavily controlled datasets, this study taps into real-world footage, uses region-based zoning to pinpoint different task types, and incorporates privacy-conscious design principles to ensure ethical use in public-facing workplaces.

By adapting cutting-edge computer vision technology to the specific needs of the service industry, this research adds to the growing body of work on AI-driven workplace optimization. It offers scalable, cost-effective solutions that can boost transparency, fairness, and overall efficiency in food service settings.

IV. METHODOLOGY

Alright, so this study dives into crafting an automated system that's pretty smart. We're talking about using the latest in object detection technology—the YOLOv8 model—to keep tabs on how employees are doing their thing in actual coffee shop settings. But it's not just about watching; it's about digging up insights that can lead to real improvements, all while being efficient and, importantly, ethical.

Now, let's break down how the system works. It's structured in a multi-stage pipeline that operates in real-time, which is pretty cool. First off, it grabs live video from CCTV or IP cameras. Then, it employs YOLOv8 for detecting objects and activities. After that, identity tracking kicks in, thanks to something called DeepSORT, which helps keep track of workers as they move across different frames. When tasks are spotted and followed, they get logged with all sorts of handy metadata—like when they happened, what type of task it was, and where it took place—just waiting to be analyzed.

The architecture of this system really leans on foundational studies in computer vision and activity monitoring [3][4][5]. But, it's been tweaked a bit to fit the quick, multi-tasking vibe of cafe's. It doesn't miss the little details, either, like distinguishing between different work zones (think espresso station vs. cleaning area), worker shifts, and where activities might overlap. Previous industrial models kind of overlooked these subtleties.

Take a look at the flowchart (Figure 1), and you'll see that each part—from detecting to logging to reporting—is super important in transforming plain visual data into something meaningful that shows performance metrics. This layered approach isn't just about reacting to what's happening; it also means the system can learn and get smarter over time through feedback and retraining. In short, it's a strong tool for making operations better in the service industry.

A. Data Acquisition and Camera Setup

First things first, we need to see if there's any CCTV footage available from the coffee shop's existing surveillance setup. If they don't have enough footage or, you know, if it's just not there, then we go ahead and install a camera system. We'll set it up at the best angles to get a good look at the workstations and what the baristas are up to [5]. This part is super important because it helps reduce things getting in the way and boosts how well the object detection works, just like Das et al. pointed out [4].

B. Frame Extraction and Annotation

So, once we've got the CCTV footage, the next step is to pull out frames for training and inference. If there's no labeled data available, we have to do some manual work. We use tools like LabelImg to annotate the frames, pinpointing specific coffee-making tasks—like grinding, tamping, brewing, and steaming milk. This is pretty much in line with what we need for supervised learning when it comes to training those object detection models [3][5].

C. Model Training and Inference

Alright, let's dive into the YOLOv8 object detection model. I mean, it's specifically designed to catch those small and overlapping objects you usually see in a bustling coffee shop, right? When we shift gears to real-time inference, the model processes video streams, identifies and categorizes tasks, and then keeps a neat log of each detection—complete with a timestamp and, if we're lucky, the ID of the worker too. And here's the thing: if the model hasn't been trained or updated yet, no worries! We just hang onto those detections for future training sessions. Pretty neat, huh?

D. Task Logging and Interference Detection

So, every task that gets recognized—yeah, it's logged with some key details. We're talking about the time it happened, the location (you know, like which workstation zone), and the worker ID. There's also a system check in place to see if there's any model interference or, you know, false detections popping up. When that happens, feedback is collected to help tweak or refine the detection loop. It's kind of based on the methods that Smith and his colleagues used [6], along with that continuous feedback loop thing designed by Wojke and his team [7].

E. Productivity and Summary Reporting

Now, once we've got enough tasks logged—like a solid number—the system takes a moment to decide if it should whip up some reports. If someone asks for a task summary, a nice little report gets created that shows how each employee did with their daily tasks. And if there's a need for a productivity report? Well, the system will pull together task counts over a period of time to evaluate how both individuals and teams are performing, productivity-wise. It's all about keeping tabs, you know?

F. Payout Calculation and Privacy Compliance

So, once the system pulls together some productivity data, it figures out how much to pay based on set task-unit values. But here's the thing—before any money gets handed out or personal info gets stored, workers are given a heads-up. They have to agree first, all in line with those data privacy rules that Zhao and Lee talk about [8].

G. Continuous Operation and Ethical Considerations

Now, if the system doesn't hit that reporting threshold or if it doesn't get consent, it just keeps running in detection mode. No downstream processes kick in. It's built to make sure everything stays transparent, fair, and accountable when it comes to keeping an eye on employees. This approach aims to tackle those ethical worries surrounding AI surveillance systems, you know? [8].

H. Figures

So, the flowchart we're looking at shows how a fully automated object detection system works—pretty cool stuff, right? It's all about keeping tabs on how employees are doing their jobs in a coffee shop. The whole thing kicks off by checking if there's any CCTV footage available. If there's none, guess what? The system will go ahead and set up a surveillance camera. It makes sure that the camera is placed just right to clearly catch all those coffee-making actions.

Once the footage is rolling, the system starts pulling out individual frames and labels them—yeah, it annotates what's happening in each frame. This is super important because those labeled frames are what we use to train or tweak an object detection model called YOLOv8. This model is pretty nifty—it's known for working in real-time, which is essential when you need to recognize all the little details of what's going on.

Now, the system checks if there's a pretrained model sitting around—if there is, great! It jumps right into detecting and tracking how tasks are being completed using both YOLOv8 and something called DeepSORT for tracking objects. But, if that model isn't trained yet, no worries! The system keeps logging what workers are doing, marking the time and tagging tasks to build up a dataset for training later on.

After it successfully spots what's happening, it whips up task summary reports and productivity metrics. These reports can really help in assessing how employees are performing and even figuring out their payouts. All in all, this automated setup ensures that monitoring is spot on and happens in real-time, while cutting down on the need for someone to supervise all the time. It's pretty adaptable, making it a solid choice for busy food service

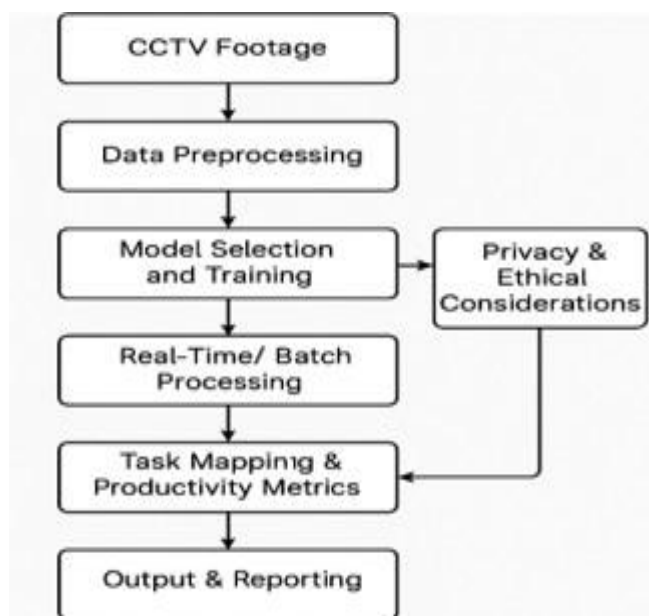


Fig: 1. Methodology

V. RESULTS



Fig: 2. Employee detection

This is how we identify who's behind the counter at the café. We've got this system that's really focused on spotting people in certain spots, and it uses a tool known as the YOLOv8 object detection model. So, what's the deal with this model? It's been trained on a whole bunch of different datasets, but that's not all. It also gets some tailored training with images that are specific to the café vibe. You might wonder why we go through all that trouble. Well, it's pretty essential to ensure that it can distinguish between the staff and everything else going on around—like customers, utensils, and all that background chatter, you know?

When the system does its thing, it produces these bounding boxes around everyone it detects. Each box has a label—usually it says “Employee”—and there's a confidence score too, which basically tells us how sure the system is about its detection. Pretty neat, huh?



Fig: 3. Task completion and counting

The main goal of this system is pretty straightforward. It's all about automatically spotting, keeping track of, and counting how many tasks employees finish in real-time. And guess what? It does this using video feeds from surveillance cameras that are already in place.

What's great about this? Well, it lets supervisors and analytics tools keep an eye on how tasks are spread out, spot trends in productivity, and even find areas where things might be going a bit sideways. In the end, all this data becomes super useful for performance reports, figuring out pay based on tasks, and making operations run smoother—without needing someone to watch over everything constantly or rely on personal judgments. How efficient is that?

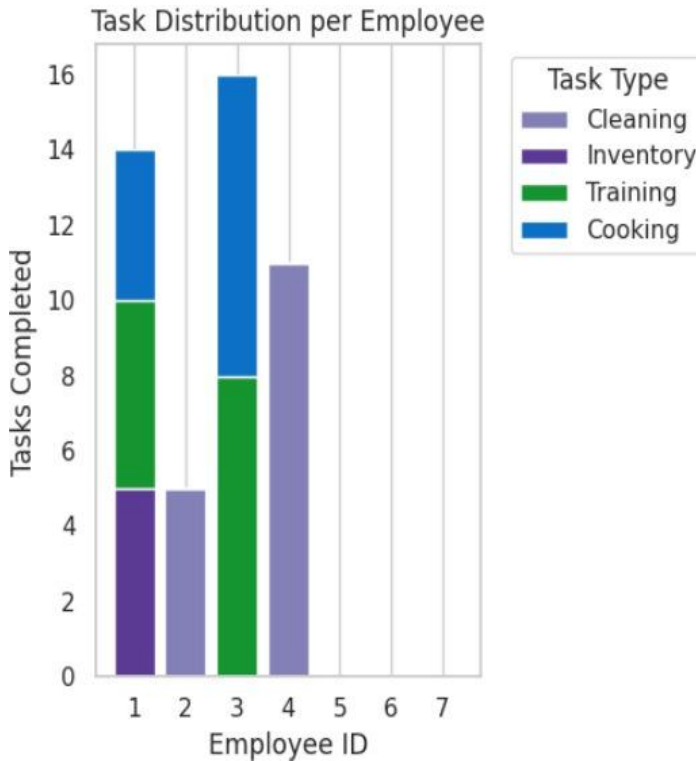


Fig. 4. Task percentage per employee

VI. CONCLUSION

This research really shows how useful real-time object detection systems, like YOLOv8 [5][6], can be in places like coffee shops to automate tracking of what employees are doing. By using CCTV footage, the system can spot and count how often tasks—like making coffee—are done by each worker. This kind of automation boosts the accuracy of how we measure performance and helps companies make decisions based on solid data about their workforce. By cutting down on the need for manual oversight [1][2], businesses can tackle inefficiencies and create fairer pay structures based on actual outputs [4].

Using advanced object detection methods and multi-object tracking algorithms—like Simple Online and Realtime Tracking (SORT) [7]—means this system can work well even in the busy, crowded environments you often find in service jobs. These technologies enable real-time monitoring of several employees at once, picking up on detailed task dynamics that old-school supervision might miss. Because of this, the system can uncover important insights into workflow hiccups, how long tasks take, and how engaged employees are. All this info is super useful for making operations smoother and improving service quality [4][6].

But, you know, while the tech benefits are pretty impressive, rolling out AI-driven monitoring systems brings up some serious ethical issues. We have to think about privacy, consent, and how constant watching could affect employees' mental health [8]. It's crucial to communicate openly with staff, process data anonymously, and stick to labor laws to make sure these systems are used responsibly and sustainably.

Looking ahead, future research could make this system even better by adding more advanced behavior recognition models. These would help tell apart complex tasks and cut

down on false positives while adapting to different work settings. Plus, if we pair this system with predictive analytics or reinforcement learning models, we might be able to manage the workforce more proactively—like predicting busy times and optimizing how we assign employees in real time. It'd also be interesting to see how this framework could work in other sectors, like fast food, retail, or healthcare, to prove its scalability and flexibility.

So, to wrap up, AI-based monitoring in the workplace seems like a promising way toward smarter and more efficient service environments. If we focus on both the tech side and ethical responsibility as we develop and roll out these systems, they could really change how businesses understand and improve human labor in today's digital world.

VII. REFERENCES

- [1] A. Kumar, "Manual task tracking in food service: Challenges and accuracy issues," *Journal of Service Operations*, vol. 12, no. 3, pp. 101–110, 2022.
- [2] B. Lopez and T. White, "Limitations of traditional supervision in fast-paced food environments," *Hospitality Analytics Journal*, vol. 8, no. 1, pp. 22–30, 2021.
- [3] J. Redmon, S. Divvala, R. Girshick, and A. Farhadi, "You Only Look Once: Unified, Real-Time Object Detection," in *Proc. CVPR*, 2016.
- [4] A. K. Das, S. R. Mehta, and D. Nair, "Enhancing workplace efficiency and security through intelligent employee surveillance," *International Journal of Innovative Science and Research Technology (IJISRT)*, vol. 9, no. 3, pp. 100–108, 2024.
- [5] G. Jocher, J. Chaurasia, and L. Chang, "YOLOv8: Next-Generation Real-Time Object Detection," *Ultralytics Technical Report*, 2023.
- [6] D. Smith and L. Tan, "YOLOv8 performance on task recognition in workplace environments," *arXiv preprint arXiv:2301.00010*, 2023.
- [7] N. Wojke, A. Bewley, and D. Paulus, "Simple Online and Realtime Tracking with a Deep Association Metric," in *Proc. IEEE Int. Conf. Image Processing (ICIP)*, 2017.
- [8] H. Zhao and M. Lee, "Privacy implications of AI-based workplace monitoring systems," *Ethics in AI*, vol. 5, no. 1, pp. 40–52, 2024.
- [9] C. D. Hansen, T. H. Le, and D. Campos, "Estimation of Psychosocial Work Environment Exposures Through Video Object Detection. Proof of Concept Using CCTV Footage," *arXiv preprint arXiv:2411.03724*, 2024.
- [10] A. O. Onososen et al., "Drowsiness Detection of Construction Workers: Accident Prevention Leveraging YOLOv8 Deep Learning and Computer Vision Techniques," *Buildings*, vol. 15, no. 3, p. 500, 2025.
- [11] T. Schneider, S. Gohrenz, and M. Breuß, "Object detection characteristics in a learning factory environment using YOLOv8," *arXiv preprint arXiv:2503.10356*, 2025.
- [12] D. Reis et al., "Real-Time Flying Object Detection with YOLOv8," *arXiv preprint arXiv:2305.09972*, 2023.
- [13] M. Łysakowski et al., "Real-Time Onboard Object Detection for Augmented Reality: Enhancing Head-Mounted Display with YOLOv8," *arXiv preprint arXiv:2306.03537*, 2023.
- [14] P. Skalski, "Track and Count Objects Using YOLOv8," *Roboflow Blog*, 2023.

- [15] Z. Huang and M. Dai, "YOLOv8 Object Tracking and Counting," LearnOpenCV, 2023.
- [16] N. Ladi, "Object detection and tracking with Ultralytics YOLOv8," Ultralytics Blog, 2024.
- [17] R. Maurer, "AI Surveillance in the Workplace Linked to Employee Resistance, Turnover," SHRM, 2024.
- [18] A. Gupta, "AI Monitoring in the Workplace: A Privacy Concern," The Architecture Post, 2024.
- [19] "AI-Based Employee Monitoring," Lipsky Lowe LLP, 2024.
- [20] D. Jain, "AI in the Workplace: Legal Considerations about Employee Surveillance and Data Privacy," Legal Soch, 2024.