

Vision-Based Worker Pose Detection and Productivity Assessment in Lathe Machine Operations Using Deep Learning

Prashant Steele*, Alka Bani Agrawal

Department of Mechanical Engineering, University Institute of Technology,
Rajiv Gandhi Proudgyiki Vishwavidyalaya, Bhopal, India

Abstract - Effectively assessing worker performance in small and medium-scale manufacturing environments has always been a problem. This is particularly true in conventional machine shop operations where the supervisor relies on manual observation. Regular inspections or subjective assessments often do not give consistent unbiased continuous productivity measures. The efficiency of machine tools depends on the body posture of the worker, the actual time spent on operating the lathe machine and the percentage of idle or non-active moves taking place in between their operation. Lack of automatically and non-intrusively monitoring prevents objectivity performance assessment. This research proposes a framework that works through a standard RGB camera along with a technology of deep learning-based Human Pose Estimation (HPE). The suggested system tracks the skeletal keypoints of the operator over the time period. Worker activities are classified as productive machine operation, tool set up, material movement and idle inactivity/disengaging in other jobs/not present at workstation, based on the temporal pose dynamics and position features. The proposed approach provides the alternative to traditional supervisory control to enhance productivity using computer vision and deep learning in manufacturing setups.

Keywords: Human Pose Estimation (HPE), Computer Vision, Lathe Machining Operations, Worker Activity, Deep Learning, Smart Manufacturing.

I. INTRODUCTION

In small and medium scale manufacturing enterprises commonly known as MSMEs, particularly those are still in the course of using conventional lathe machines, human performance characteristics plays a important role in production output[1]. As an alternative to traditional setups, CNC systems, which automatically record detailed operational metrics such as spindle time, feed rates, and cycle duration, unlike manual lathes that offer little to no embedded data logging capability to track workers workmanship. While production outputs and shift timings are documented, the worker activities during the shift are often remain unrecorded. Historically, efforts to capture this “in-between” activity have relied on classical time-and-motion studies [2]. In such approaches, trained observers or

supervisors manually record task sequences using stopwatches and standardized data sheets in which entries are recorded which may be biased. Although these techniques can yield workers performance statistics but it is quite labour-intensive. The presence of an observer may alter worker behavior, a phenomenon widely recognized as the Hawthorne Effect, thereby compromising ecological validity. Moreover, continuous manual observation is practical not viable and impractical for round-the-clock monitoring. These limitations highlight the need for scalable, objective, and non-intrusive methods to assess worker productivity in conventional manufacturing settings [3], [4].

However, to track the workers performance, most of the currently used solutions require expensive sensors, wearable tags, or complex IoT integration that older machines simply don't support. This brings us to Computer Vision, a low cost alternative to track and assess the performance of worker automatically. Cameras are cheap, ubiquitous, and non-intrusive [5], [6]. [7], [8]. If a computer could “watch” the video feed and understand human movement just like a supervisor does, we could digitize productivity monitoring without modifying the machine at all. This paper describes a system that does just that. We employ the latest Deep Learning models for detecting the human skeleton. We also define a virtual “Machine Zone” (Region of Interest) surrounding the lathe's controls, and programmatically determine if the worker is working or idling based on the location of their hands [9], [10]. This is a simple, robust, privacy-friendly, and modern way to measure manual labor.

II. METHODOLOGY

Fundamentally, our methodology is based on “Postural Analysis”. In a lathe shop, productivity is physically expressed through the posture. A worker handling a carriage handwheel in one posture, and another who is leaning against a wall and looking at his/her smartphone in a

different posture. Our approach converts these visual signals into numbers.

A. System Architecture

- The system has three main stages:
- Video Acquisition: The workstation is captured by a monocular RGB camera, no depth sensors or special infrared equipment are necessary
- Pose Estimation: We feed the Deep Learning model (YOLOv8-Pose) with single frames to retrieve skeletal key points
- Activity Classification: We use logic engine to compare key points with relation to machine position to recognise the activity state.

B. Human Pose Estimation (HPE)

We use the YOLOv8-Pose model. While previous models had either required a lot of computational resources (OpenPose) or were limited to single person use (keypoints or other), YOLOv8 is a great trade-off between speed and accuracy. It casts pose estimation as a problem of object detection (i.e. detecting persons) and then regressing coordinates of 17 key human joints (e.g. shoulders, elbows, wrists, hips, etc.). In our case, the most important joints are the Wrists (Left: Keypoint 9, Right: Keypoint 10). The location of the hands indicates engagement with the machine.

C. Region of Interest (ROI) and Logic

Creating the system contextualization involves defining a Region of Interest (ROI) or what we call Machine Zone. A virtual bounding box is drawn over the video feed encircling the important controls of lathe such as the chuck, tool post and carriage handwheels.

- The classification logic is simple yet effective.
- Productive: If the worker is detected AND at least one of their wrists is inside the Machine Zone with high confidence. It means the worker is in charge of the machine.
- Idle: If the worker is detected in the frame, but their hands lie outside the Machine Zone for a long time, then Idle. This will record instances of standing near, talking, or resting.
- If there is no human skeleton, then Away.

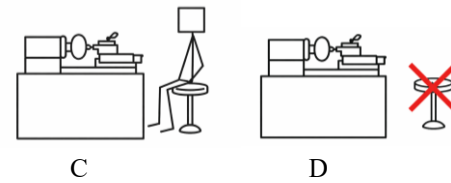
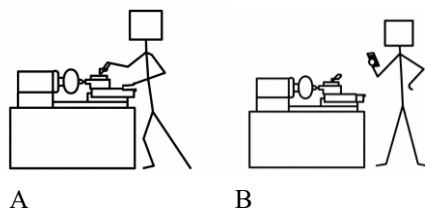


Fig 1. Different Working States of Worker in a Lathe Machine

D. Implementation

We implemented the proposed system using Python, leveraging the Ultralytics library for the AI model and Streamlit for the user interface. A standard workstation running Windows was used to deploy the application with following parameters.

- The video feed is analyzed frame by frame.
- 1080p video feed webcam or file
- The yolov8n pose.pt model (Nano version) was selected for its capability to run in real-time on consumer hardware.
- An interface was designed to allow the shop floor managers to set the ROI visually. As every lathe setup differs, hard-coding of machine controls coordinate would fail.
- Users can use our UI to easily frame the machine by drawing a box around it in the video feed.

III. RESULTS AND DISCUSSION

To validate the effectiveness of the LatheSense system, we deployed it for a period of 7 days in a controlled lathe workshop environment. The automated procedures continuously monitored the worker's posture and proximity analysis and interaction with the machine.

A. Productivity Metrics

The core metric for this study is the 'Productivity ratio', which quantifies the percentage of the total shift time the worker spent actively operating the machine. This is calculated using Eq. (1):

$$\text{Productivity ratio (\%)} = \left(\frac{\text{Time}_{\text{Productive}}}{\text{Time}_{\text{Total}}} \right) \times 100$$

Where:

- 'Time_{Productive}' is the duration where the worker's hands were detected inside the Machine Zone.
- 'Time_{Total}' is the sum of Productive, Idle, and Away durations.

B. Weekly Performance Table

Table 1 below summarizes the daily worker recorded data with machine interaction over the one-week period.

Table 1: Worker’s Daily Productivity Analysis (Week of 02-02-2026 to 07-02-2026)

Date	Total Shift Time (Hours)	Productive Time (Hours)	Productivity Ratio (%)
02-02-26	8	4.1	51.1%
03-02-26	8	6.0	74.9%
04-02-26	8	4.1	50.8%
05-02-26	8	5.2	65.1%
06-02-26	8	5.5	68.6%
07-02-26	8	5.3	66.5%

C. Discussion of Results

The data collected (Table 1) inferred an Average Productivity Ratio of worker as 62.8% for the observed week. The fluctuations in daily productivity can be attributed to varying workload complexities. For instance, days with lower ratios often correspond to lack of productive performance where the worker spends significant time gathering tools or reading blueprints outside the Machine Zone (classified as 'Idle'). Conversely, higher ratios indicate determined worker characteristics in continuous turning operations. The logic that is based on Region of Interest really worked well for telling the difference between work and just doing nothing. The Region of Interest based logic was a success, in figuring out when people were actually working and when they were just being idle as shown in Fig 1. The various state of worker posture detection is classified as under.

- Scenario A (Working) is about a Turning Operation. The worker had their hand on the carriage wheel the whole time. Their left hand was near the clutch. The system said this was Productive and it was right. The worker was being Productive, in this Scenario A. The worker leaned in to measure the workpiece with calipers. This is happening inside the Machine Zone. So it was also counted as Productive time. This is correct because setup time is part of the job when you are working with the Machine Zone and using tools like calipers for measurement.
- Scenario B (Using Phone): The worker stepped back two feet and sitting on the chair. Hands exited the ROI. The system switched to Idle.
- Scenario C (On Chair): The worker stepped back two feet and pulled out a phone. Hands exited the ROI. The system switched to Idle.
- Scenario D (Away Zone): The worker is not with the machine working area. The system switched to Idle.

IV. CONCLUSION & FUTURE WORK

This study shows that you do not need sensors to make manual manufacturing digital. If you use a camera with modern Deep Learning you can create a "virtual supervisor" that is fair and never gets tired. The LatheSense system is

very good, at figuring out how workers are standing and sitting and it can calculate how productive they are away. The Lathe Sense system does this by using the camera and the modern Deep Learning to watch the workers and see what they are doing. Future work will focus on Action Recognition and production item count analysis. Instead of just checking if hands are in the zone, we want to train a model to recognize specific movements—like "tightening chuck" vs "turning handwheel". This would allow us to break down the "Productive" time into specific tasks, offering even deeper insights into the manufacturing process.

CONFLICT OF INTEREST

The authors declare that they have no conflict of interest.

FUNDING

This research did not receive any specific grant

REFERENCES

- [1] M. Kessler and J. C. Arlinghaus, "A framework for human-centered production planning and control in smart manufacturing," *J. Manuf. Syst.*, vol. 65, pp. 220–232, Oct. 2022, doi: 10.1016/j.jmsy.2022.09.013.
- [2] F. Tao, Q. Qi, A. Liu, and A. Kusiak, "Data-driven smart manufacturing," *J. Manuf. Syst.*, vol. 48, pp. 157–169, Jul. 2018, doi: 10.1016/j.jmsy.2018.01.006.
- [3] M. Bhattacharya, M. Penica, E. O’Connell, M. Southern, and M. Hayes, "Human-in-Loop: A Review of Smart Manufacturing Deployments," *Systems*, vol. 11, no. 1, p. 35, Jan. 2023, doi: 10.3390/systems11010035.
- [4] W. Tao, Z.-H. Lai, M. C. Leu, and Z. Yin, "Worker Activity Recognition in Smart Manufacturing Using IMU and sEMG Signals with Convolutional Neural Networks," *Procedia Manuf.*, vol. 26, pp. 1159–1166, 2018, doi: 10.1016/j.promfg.2018.07.152.
- [5] R. Venkata Rao and B. K. Patel, "Decision making in the manufacturing environment using an improved PROMETHEE method," *Int. J. Prod. Res.*, vol. 48, no. 16, pp. 4665–4682, Aug. 2010, doi: 10.1080/00207540903049415.
- [6] C. C. Martin *et al.*, "A real-time ergonomic monitoring system using the Microsoft Kinect," in *2012 IEEE Systems and Information Engineering Design Symposium*, Charlottesville, VA: IEEE, Apr. 2012, pp. 50–55. doi: 10.1109/SIEDS.2012.6215130.
- [7] A. Toshev and C. Szegedy, "DeepPose: Human Pose Estimation via Deep Neural Networks," in *2014 IEEE Conference on Computer Vision and Pattern Recognition*, Columbus, OH, USA: IEEE, Jun. 2014, pp. 1653–1660. doi: 10.1109/CVPR.2014.214.
- [8] E. Insafutdinov, L. Pishchulin, B. Andres, M. Andriluka, and B. Schiele, "DeeperCut: A Deeper, Stronger, and Faster Multi-person Pose Estimation Model," in *Computer Vision – ECCV 2016*, vol. 9910, B. Leibe, J. Matas, N. Sebe, and M. Welling, Eds., in Lecture Notes in Computer Science, vol. 9910, Cham: Springer International Publishing, 2016, pp. 34–50. doi: 10.1007/978-3-319-46466-4_3.
- [9] J. Yan and Z. Wang, "YOLO V3 + VGG16-based automatic operations monitoring and analysis in a manufacturing workshop under Industry 4.0," *J. Manuf. Syst.*, vol. 63, pp. 134–142, Apr. 2022, doi: 10.1016/j.jmsy.2022.02.009.
- [10] S. Ren, K. He, R. Girshick, and J. Sun, "Faster R-CNN: Towards Real-Time Object Detection with Region Proposal Networks," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 39, no. 6, pp. 1137–1149, Jun. 2017, doi: 10.1109/TPAMI.2016.2577031.