Proceedings of the 8th European Conference on Industrial Engineering and Operations Management
Paris, France, July 2-4, 2025

Proceedings of the International Conference on Industrial Engineering and Operations Management

Publisher: IEOM Society International, USA DOI: 10.46254/EU08.20250469
Published: July 2, 2025

Smart Robotic Manufacturing Operations with Adaptive Task Planning: GPT-Augmented Multi-Channel Ensemble Learning with Situational Awareness

Yiyun Fei, Mulang Song and Roger J. Jiao
School of Mechanical Engineering
Georgia Institute of Technology, Atlanta, GA, USA
yfei38@gatech.edu

Shu Wang and Xinping Deng
TE Connectivity
Middletown, PA, USA

Abstract

This paper presents a smart robotic operations framework that integrates multi-camera perception, GPT-augmented multi-channel ensemble learning, and domain-driven task planning for dynamic manufacturing environments. By leveraging synchronized top, side, and scene cameras, the system achieves high-precision object localization and real-time situation awareness across complex work zones. Stacking-style ensemble learning fuse multi-view pose estimates into reliable robot control commands that adapt to variable spatial conditions. GPT- augmented reasoning further enables automatic generation of Planning Domain Definition Language (PDDL) models, which, together with ROSPlan integration, translate high-level task plans into executable robot actions. A preliminary use case in flexible electronics component assembly demonstrates the system's capability to handle multi-task scheduling, material changeover coordination, and task prioritization across parallel workstations. This framework bridges perception, reasoning, and manipulation within a closed-loop architecture, offering a scalable approach toward adaptive, situation-aware robotic operations in manufacturing. Future work will focus on real-world deployment, enhanced domain modeling, and broader integration with production management systems.

Keywords
Smart Robot Operations, Multi-camera Vision Fusion, GPT-augmented Ensemble Learning, Adaptive Task Planning, Situation Awareness.

1. Introduction
Robotic systems have long been integral to manufacturing automation, performing tasks such as assembly, welding, painting, packaging, and material handling with high speed and repeatability. Traditional industrial robots are typically deployed in structured environments, operating on fixed workpieces along predefined trajectories with little variability. These systems excel in high-volume, low-variability production lines where tasks are repetitive, environments are tightly controlled, and object positions are precisely specified. However, as manufacturing shifts toward high-mix, low-volume production, rapid product changeovers, and increasing demand for customization, conventional robotic solutions face significant limitations. Reprogramming these systems to accommodate new tasks can be costly and time-consuming. Moreover, the inherently unpredictable and dynamic nature of manufacturing poses significant challenges to conventional automation. Among these challenges, two critical aspects have emerged: ensuring robust perception in increasingly unstructured environments and enabling flexible task execution that can adapt to real-time production variability.

(1) Vision-Guided Robotics in Manufacturing: To improve flexibility, vision-guided robotic systems have been widely adopted across various manufacturing operations. For example, many modern pick-and-place systems utilize vibrating feeder plates combined with camera-guided robots to dynamically locate and grasp randomly distributed parts. Furthermore, camera technologies have evolved from traditional 2D systems to include 3D vision and depth sensors, expanding their capability to handle more complex geometries and spatial uncertainties. Despite these advancements, vision-guided robotic systems still encounter significant limitations when applied to highly variable or unstructured environments. Current industrial vision system for robots often rely heavily on conventional rule-based image processing and classical machine vision algorithms, which are typically optimized for well-constrained tasks. When faced with extensive variations in part geometry, placement, lighting, or unexpected occlusions, these traditional approaches can struggle to maintain reliable performance, limiting their effectiveness in increasingly diverse and dynamic production scenarios.

(2) Task Planning with Situation Awareness: While task planning technologies have advanced significantly at the production scheduling level—optimizing order sequencing, resource allocation, and overall throughput—many low-level, discrete manufacturing tasks still rely on human operators for real-time situational judgment. Tasks such as sample collection, box replacement, material replenishment, or ad-hoc tool changeovers often require on-the-spot decisions based on current shop floor conditions. Even in systems where robots leverage Manufacturing Execution System (MES) data for automated planning, the accuracy and completeness of such production data may not fully reflect the actual state of the physical environment. Variations caused by human interventions, equipment disturbances, or unexpected delays are not always captured in MES databases promptly or accurately. This mismatch between digital plans and physical reality underscores the importance of real-time situation awareness.

To address these challenges, this paper proposes a unified smart robotic operations framework that integrates multi-camera perception, GPT-augmented ensemble learning, and adaptive task planning into a closed-loop system for dynamic manufacturing environments. The system combines real-time multi-view vision with adaptive reasoning to enable robots to perceive complex scenes, interpret dynamic shop floor conditions, and autonomously generate executable task plans. A pilot application in flexible electronics component assembly demonstrates how the framework manages multi-task coordination, material changeovers, and task prioritization across parallel workstations. The proposed approach aims to enhance manufacturing flexibility, improve decision-making robustness, and reduce the reliance on human intervention for routine but situation-sensitive tasks in future smart factory environments.

2. Emerging Technologies for Smart Robotics in Manufacturing

2.1 Vision-Guided Robotics in Manufacturing

Machine vision has become a cornerstone of modern manufacturing robotics. It enables robots to recognize parts, inspect quality, and guide manipulation tasks with high precision. Recent advances in deep learning-based vision have greatly improved robotic perception on factory floors. Zhao et al. (2024) develop a deep learning framework for vision-guided robots that use semi-supervised knowledge distillation to generalize detection models across varying production scenarios. Vision-guided robots are also being applied to quality control – Chaabani et al. (2025) integrate a 2.5D camera with a Doosan robot for real-time defect detection, using multiple pretrained CNN models to identify defect parts on assembly lines. Terras et al. (2025) demonstrates that combining collaborative robots with advanced vision can significantly improve manufacturing efficiency and accuracy in tasks like part sorting, surface inspection, and precision assembly.

Research has shown that fusing multiple camera views can dramatically improve a robot's understanding of its workspace by overcoming occlusions, improving 3D scene reconstruction, and enhancing perception accuracy (Lin et al., 2021). For example, multi-camera tracking systems have been applied to in-plant logistics and human–robot interaction, enabling real-time 3D trajectory prediction for flying objects and more robust gesture recognition by observing actions from multiple angles (Qadeer et al., 2024; Bandi and Thomas, 2025). Recent advances also explore multi-view self-supervised learning, where fused visual representations allow policies to remain robust under camera variations during manipulation tasks (Seo et al., 2023). Moreover, ensemble learning methods, which fuse predictions from diverse classifiers or representations, have been shown to improve object recognition and enable continual learning of new objects in open-ended industrial settings (Kasaei and Xiong, 2024).

2.2 Task Planning in Dynamic Manufacturing Systems

In dynamic manufacturing settings, robotic systems must be guided by flexible task planning algorithms. Classical AI planning, specifically PDDL-based symbolic planning, has re-emerged as a powerful tool for specifying and orchestrating robot skills on the shop floor. The ROSPlan framework was an early effort to embed a PDDL planner within ROS, enabling robots to autonomously plan action sequences to achieve goals, and recent work has been built on this by introducing an Action Interface Manager to improve real-time plan execution and monitoring, allowing the system to adapt dynamically to failures or environmental changes (Cashmore et al., 2015; Bezruczko et al., 2021). Moreover, recent works address the challenges of applying planning in real factories. Heuss et al. (2024) propose a skills-based planning approach that automatically adapts abstract planning domains to specific manufacturing cases by auto-generating PDDL action models from a high-level skill library, enabling operators to specify only the goal while the system generates viable action sequences for reconfiguring industrial robots to new tasks.

2.3 Large Language Models and Generative AI for Robotics

Another emerging frontier is the integration of Large Language Models (LLMs) and generative AI into robotic frameworks. LLMs like GPT-3/4 have demonstrated remarkable abilities in reasoning and knowledge representation, and researchers are now leveraging these capabilities to enhance robot intelligence. Researchers have shown that LLMs can serve as planners by interpreting complex, ambiguous instructions and grounding them in robotic actions using knowledge of affordances and constraints (Ahn et al., 2022). Another promising direction is to have LLMs generate robot control code or programs in real time, effectively using natural language prompts to produce executable policies for manipulation tasks (Liang et al., 2023). Mon-Williams et al. (2025) proposed the ELLMER framework, which integrates large language models with robotic sensorimotor capabilities and retrieval-augmented generation to enable robots to perform complex, long-horizon tasks in unpredictable environments by dynamically adapting plans based on force and visual feedback. Recent work also demonstrates that combining GPT-4o vision models with classical computer vision techniques enables robots to achieve highly accurate object recognition and manipulation with reduced reliance on traditional kinematics (Author et al., 2025).

3. System Analysis and Design

The system architecture shown in Figure 1 illustrates a generalized smart robotic operations system capable of handling a wide range of manufacturing tasks – whether it involves assembly, inspection, material handling, packaging, or tool changeover – across points in a production line. The system begins with image acquisition, collecting visual data using top, side, and scene cameras from multiple views. The resulting images serve as inputs to both the object positioning and situation recognition modules.

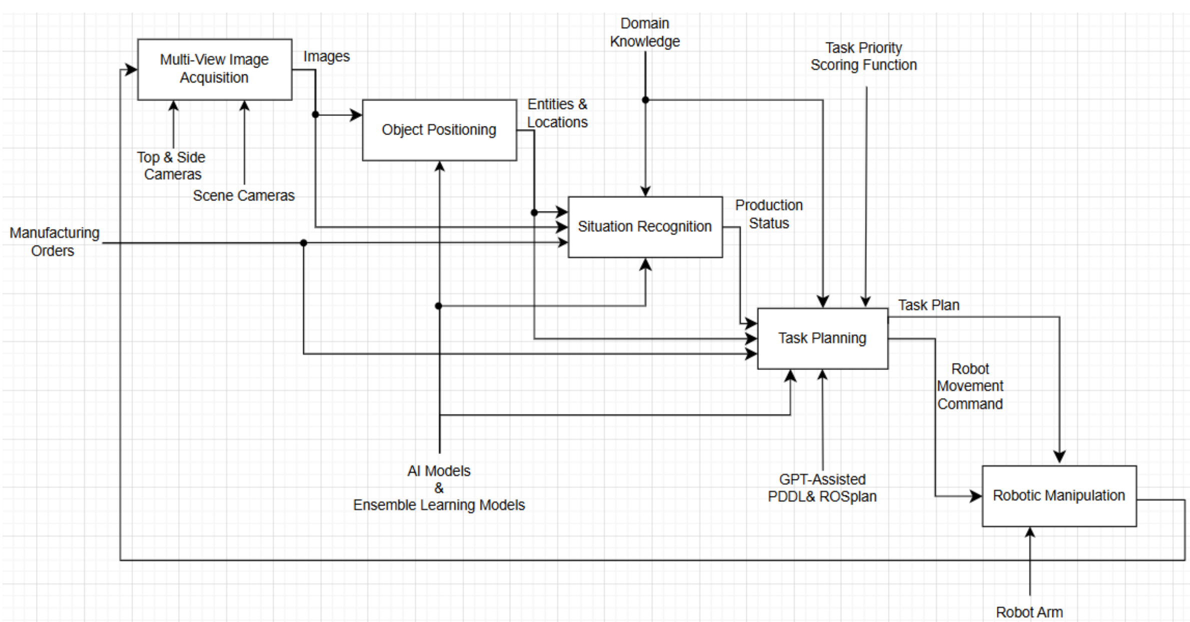

Figure 1. System Architecture of Smart Robotic Operations

In the object position module, the images are processed using ensemble learning models and AI models to identify key entities involved in the operation, for instance, machines, materials, tools, and products, along with their precise locations. These identified entities and their locations are then passed to the situation recognition module. This module interprets the images and labels in the context of current and incoming manufacturing orders and utilizes AI models and ensemble learning models as mechanisms to generate real-time production status.

The production status, together with the identified entities and manufacturing orders, informs the task planning module. This module, governed by domain knowledge and supported by GPT-augmented reasoning, utilizes both PDDL (Planning Domain Definition Language) and ROSPlan to generate executable task plans and robot action sequences. GPT is responsible for analyzing perception and situation data to automatically construct planning models and generate appropriate PDDL and ROSPlan configurations. The resulting plans are then translated into robot commands that guide the robotic manipulation module, enabling the robot arm to carry out physical tasks in coordination with the broader manufacturing workflow. In this architecture, domain knowledge serves as a control for both situation recognition and task planning, while the task plan also acts as a control directive for robotic manipulation. This functionally structured adaptive system enables adaptive and situation-aware robotic behavior across a variety of manufacturing scenarios, not limited to specific product type.

4. Multi-Camera for Situation Awareness and Vision Guidance

In the smart robotic operations system, multi-camera input is essential for enabling perception beyond a single fixed viewpoint. The multi-camera subsystem supports comprehensive scene understanding and vision-guided robotic behavior. To achieve accurate and robust perception of the real world, the system employs ensemble learning, which fuses the output of multiple AI models trained on image data captured by the top, side, and scene cameras. Each camera provides a distinct spatial perspective, and together they enable full-scene observation with enhanced depth, pose, and occlusion handling. This multi-view image acquisition allows the system to estimate both the location and orientation of entities such as tools, parts, machines, and human operators.

The integration of multiple viewpoints also facilitates situation awareness, in which the system learns to interpret work cell conditions (e.g., workstation occupancy, tools and material availability, task progress, safety hazards) and make informed decisions during task planning stage. For example, scene cameras may detect whether a workstation is currently occupied and identify signs that material replenishment is needed, triggering the system to reprioritize tasks, initiate replenishment actions, or notify human operators. This situation awareness enables the robot to act proactively and maintain workflow continuity without manual intervention.

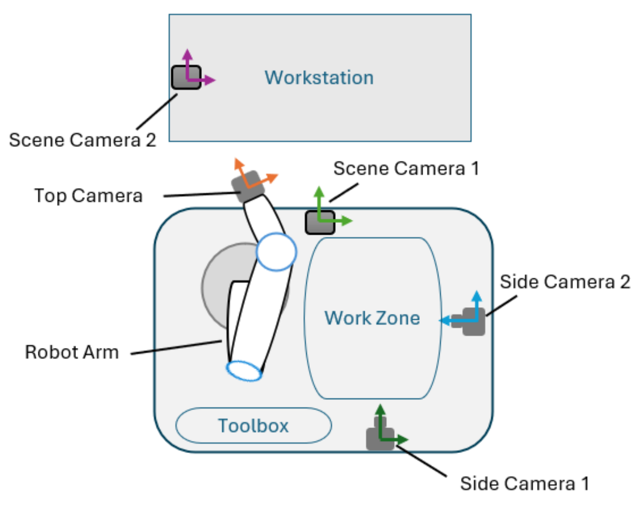

Figure 2. Multi-Camera Deployment Schematic at the Work Cell

Crucially, the multi-camera setup supports vision-guide robotic control, where the system perceives and interprets the spatial positions and orientations of target objects in real-time. These observations are transformed into executable commands within the robot's control frame, guiding its movements for precise interaction with the target objects.

Learning-based techniques are integrated to accommodate variations caused by non-linear offsets and dynamic workspace conditions. Detailed is explained in the next chapter. The system further incorporates real-time visual feedback, ensuring the robot continuously refines its actions based on live visual input. This includes correcting component misalignment, detecting dropped or shifted parts, and recognizing environmental changes or human presence. Through this feedback loop, the robot adapts its behavior dynamically, improving operational robustness and accuracy in high-mix, variable manufacturing settings.

Figure 2 presents a top-view schematic of a representative deployment of a smart robotic operations cell within a manufacturing environment. The layout includes a fixed-position workstation, which represents the manufacturing process itself. A dedicated scene camera is mounted above the workstation to monitor production status from an overhead perspective. The smart mobile robotic cell consists of a robotic arm equipped with a top-mounted camera at its end-effector, along with two side cameras positioned to capture depth information in the work zone. An additional scene camera is oriented to assist in situation awareness, particularly in observing the interaction dynamics between the robot and the workstation. The specific number and placement of cameras may vary depending on the requirements of a given application scenario.

5. GPT-Augmented Multi-Channel Ensemble Learning for Precise Robot Control

This chapter introduces a unified framework that leverages the complementary strengths of multiple vision models, large-language-model reasoning, and stacking-style ensemble strategies. Ensemble learning synthesizes the outputs of multiple, diverse predictors to achieve greater accuracy and robustness than any single model alone. In this framework, each camera view generates an independent 6 Degree of Freedom pose estimate for every object, and a meta-learner then learns to weight and combine these estimates based on situational cues—such as workspace status and objects around—into a reliable robotic control command. GPT-o3 model is utilized to automate the configuration of the meta-learner's architecture, training procedures, and runtime parameters so that the resulting fusion model can dynamically adapt to changing conditions and minimize estimation errors in real time. While individual base learners may leverage voting, bagging, or boosting to refine their predictions, the defining characteristic of our approach is the stacking-style multi-channel ensemble: a final, learned combiner that ingests all base outputs (and scene context) to produce the ultimate robot pose, thereby unifying the strengths of each strategy under a single meta-learner.

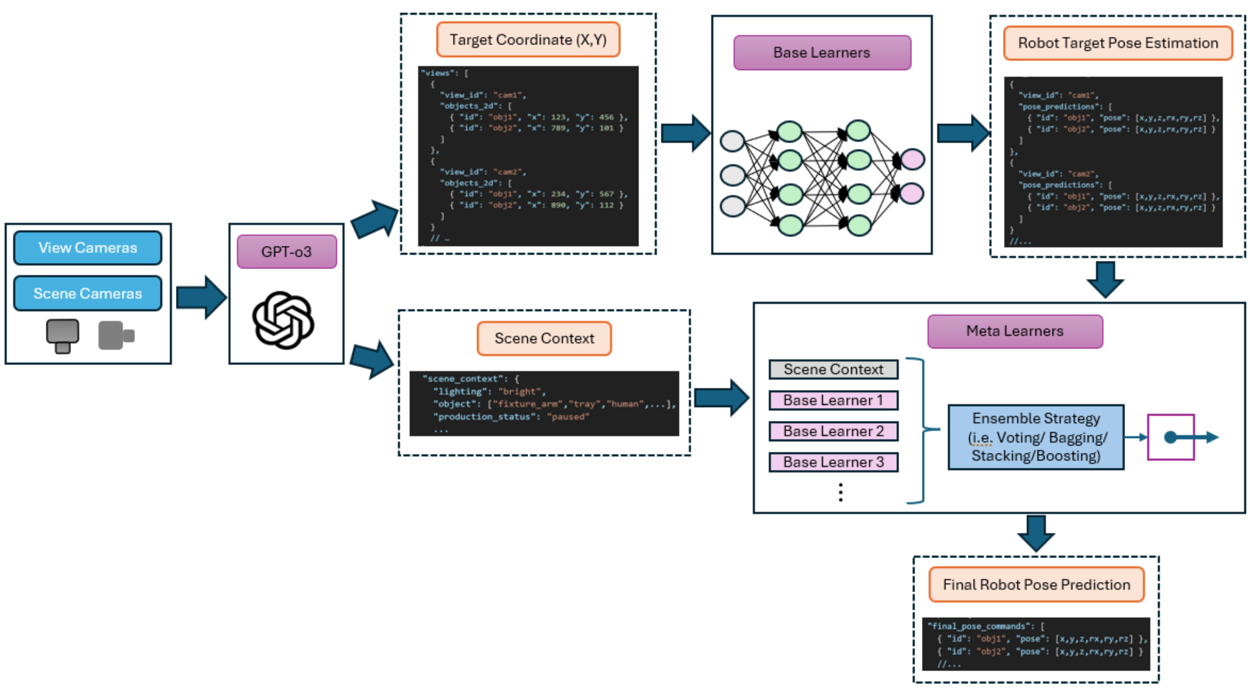

Figure 3. Vision-driven Ensemble Pipeline that Fuses Multi-view Camera Data into Precise Robot Pose Commands

Figure 3 illustrates the end-to-end vision-driven control pipeline for precise robot manipulation. First, a set of synchronized view cameras and fixed scene cameras capture both localized target perspectives and broad situational information. All images are sent together to a GPT-o3 model, which performs two tasks in sequence: it produces a high-level "situation" annotation and extracts 2D pixel coordinates for each target object in every view. Next, each view's 2D detections are passed through GPT-augmented neural networks, yielding per-object 6 Degree of Freedom pose estimates (x, y, z, rx, ry, rz) along with confidence scores. Finally, these multi-view pose predictions and the scene context are fused by a GPT-o3-based meta-learner ensemble model—to generate a single, robust set of final robot pose commands. These commands are then forwarded to ROS Plan for seamless integration with downstream task planning, closing the loop between precise object positioning and high-level robotic task execution.

6. Task Planning for Situation-Aware Robotic Operations

Effective task planning is the cornerstone of intelligent robotic operations in dynamic manufacturing environments. In this system, task planning is not an isolation function but is closed integrated with situation awareness, real-time perception, and system-wide coordination. A structured task representation forms the foundation of the the adaptive task planning framework. GPT-augmented reasoning is used to analyze the perception data and generate planning artifacts, including PDDL (Planning Domain Definition Language) models that formally encode domain knowledge, resource constraints, and action effects. These PDDL models are then integrated with ROS Plan to create executable task sequences for the robot arm. This combined planning approach systematically determines what actions to perform (e.g., pick/place, insert, screw, sort), where to perform them (e.g., identifying trays or sockets), when to perform them (e.g., based on system state or task priority), and how to execute them (e.g., selecting appropriate paths, tools, cameras, or grippers).

Accurate and real-time environmental understanding is essential for robotic decision-making. A multi-camera vision system provides continuous situation awareness. This perception data is actively integrated into the planning loop to ensure the feasibility and situational relevance of generated actions. Given the inherent uncertainties of manufacturing environments—such as missing components, occupied workstations, or mid-execution failures—the system supports reactive re-planning to dynamically adjust action sequences based on real-time feedback. Vision-based execution monitoring detects anomalies like missed grasps or misalignments, triggering automatic retry mechanisms or contingency plans. This closed-loop paradigm enhances system robustness and enables online updates to the domain model, such as marking tray slots as unavailable after repeated failures.

When multiple mobile robot cells operate across distributed workstations, task planning must scale to system-level coordination. To support real-time scheduling decisions, a priority scoring function is implemented to rank available tasks based on perceptual and operational situations. A simplified version of the task priority function is defined as:

$$Priority(T_i) = \alpha U_i + \beta A_i + \gamma R_i + \delta F_i - \lambda C_i \quad (1)$$

where:

- U_i: Task urgency (e.g., order due date)
- A_i: Accessibility and executability (e.g., object readiness, station availability, absence of human interference)
- R_i: Robot-task suitability (e.g., availability, proximity, tool readiness)
- F_i: Failure recovery or safety-critical boost (e.g., task is a reattempt after a mid-execution failure, or has high safety implications such as proximity to humans or critical equipment)
- C_i: Estimated execution cost (task duration × failure risk × travel distance)

The weighting factors (α, β, γ, δ, λ) can be tuned to reflect strategic priorities in different production scenarios. This formula enables efficient, situation-aware, and tradeoff-balanced task allocation for multi-robot coordination.

7. Pilot Application in Flexible Electronics Component Assembly

7.1 Problem Description

In flexible electronics component assembly, the efficient handling of material changeover poses a critical challenge not only at individual workstations but also at the system-wide level, where multiple changeover operations must be coordinated across parallel production lines. The assembly system involves multiple workstations operating simultaneously, each processing a diverse set of components—such as resistors, capacitors, and transistors—supplied through various packaging formats including tape-and-reel and box tapes. The material carriers differ in size, material composition (e.g., cardboard vs. plastic reels), and hole configurations, introducing significant variability that complicates automated handling. Adding to this complexity is the dynamic nature of production orders. The sequence, quantity, and combination of required components can shift frequently, demanding that the material feeding system remain highly flexible and responsive to change at different stations. For example, while Workstation 1 may require reel replenishment at its input feeder, Workstation 2 may simultaneously require a tray swap at its finished goods output zone, and Workstation 3 may also signal a full tray ready for replacement. These concurrent needs create a multi-task scheduling problem, where the robotic system must prioritize and allocate tasks across multiple work zones in real time.

Currently, the responsibility of managing material changeovers falls largely on human operators. The high-speed insertion process means that reels can be consumed rapidly, demanding close attention and timely intervention. Each reel may only hold enough components for approximately 30 insertions, requiring frequent replenishment. If the operator is delayed or an issue arises during the changeover, the entire production line may be paused, directly impacting productivity and increasing operational costs. A particularly delicate step during changeover is ensuring the continuous feed of components—this involves aligning and connecting the tip of a new reel to the tail of the consumed one. This task must be performed with high precision to avoid misfeeds, jamming, or misalignment that could interrupt the insertion process. The variability in reel geometry, hole alignment, and posture sensitivity further increases the technical demands on robotic gripper precision. Gripper alignment poses a significant challenge, particularly when dealing with hole-based positioning systems. Slight posture errors or hole mismatches during pick-and-place can lead to reel slippage or improper tape feed, making robust gripper control and fine alignment capabilities essential.

The robotic system must handle multiple parallel workstation-level tasks alongside sequential robotic operations. For material handling at input stations, robotic tasks include retrieving the spent reel, fetching a new reel from an input buffer rack, aligning and connecting the tape ends at a designated work zone on a mobile robot cart, and reinstalling the new reel in its original position. In addition to material loading, the mobile robotic cell is also responsible for finished goods tray changeovers. Once a tray is full, the robot transfers it to a collection cart designated for completed goods and replaces it with an empty tray to maintain continuous operation. Each task presents distinct geometric, spatial, and operational constraints, which are continuously updated via multi-camera vision systems that monitor object availability, workstation accessibility, human presence, and unexpected disturbances across all work zones.

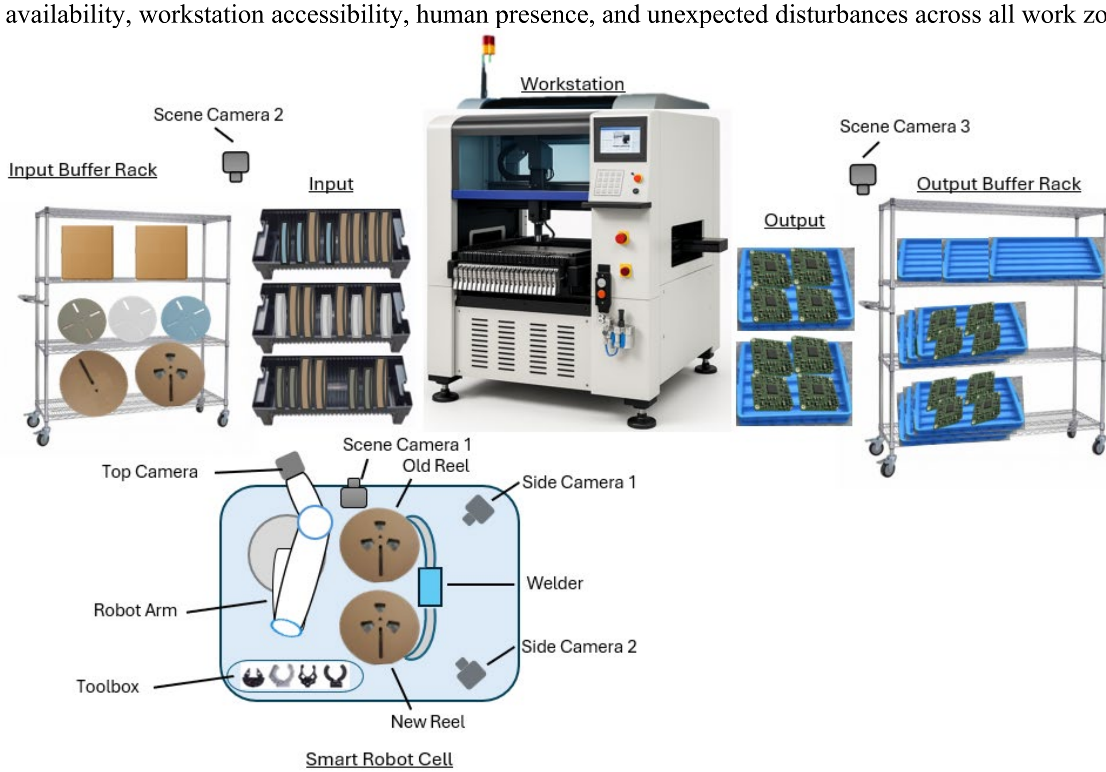

Figure 4. Example Workstation Layout for Electronics Assembly Lines

Figure 4 illustrates the layout for the smart robotic operations cell deployed in a flexible electronics component assembly environment. The setup consists of three primary zones: the input zone, the output zone, and the mobile cell zone, each monitored by dedicated scene cameras (Scene Cameras 1–3). Reels of various sizes and types are stored in the input buffer rack and input slots are visually tracked by Scene Camera 2. During operation, the mobile robotic cell—equipped with a robot arm, top camera, side cameras, and a scene camera—performs reel changeovers and tape alignment tasks within a designated workspace. A welder located on the mobile cell enables the precise connection of new and old tapes. The gripper toolbox ensures flexibility to handle different materials and tasks. After component insertion at the workstation, completed units are placed into trays. The mobile cell also handles finished goods tray changeovers, transporting full trays to the output buffer rack and replacing them with empty ones, as monitored by Scene Camera 3.

7.2 Task Planning

As shown in Table 1, considering three pending tasks in the system: T1 involves reel replenishment at Workstation 1 where the reel is 80% depleted; T2 requires tray replacement at Workstation 2 with the tray 90% full; and T3 is a tray replacement at Workstation 3 with the tray 50% full. Using perception data and system state, the normalized factor values are summarized in Table 2.

Table 1. Three Tasks Waiting for Planning and Execution

Task ID	**Workstation**	**Task Type**	**Description**
T1	WS1	Input Reel Replenishment	Reel is 80% depleted
T2	WS2	Output Tray Replacement	Tray is 90% full
T3	WS3	Output Tray Replacement	Tray is 50% full

While all tasks remain accessible according to real-time vision feedback, the reel replenishment task (T1) exhibits slightly reduced pose confidence due to its higher alignment sensitivity. Reel changeover requires precise positioning to correctly align the tape ends, which increases its failure risk and consequently affects both its failure sensitivity factor *Fi* and execution cost *Ci*. In contrast, tray replacements (T2 and T3) generally involve simpler positioning, leading to lower manipulation complexity and lower failure penalties. Robot-task suitability *Ri* further differentiates task priorities by incorporating spatial proximity: the robot is currently positioned closer to Workstation 2 (T2), followed by Workstation 1 (T1), and furthest from Workstation 3 (T3). This proximity advantage favors T2 by reducing travel time and energy consumption.

Table 2. Normalized Factor and Weight of each Task

Factor	**Weight**	**T1**	**T2**	**T3**
U_i	$\alpha = 5$	0.8	0.9	0.5
A_i	$\beta = 3$	0.9	1.0	1
R_i	$\gamma = 2$	0.85	0.95	0.7
F_i	δ=2	0.2	0	0
C_i	λ=4	0.4	0.3	0.2

Based on the priority function, the resulting scores are 7.2 for T1, 8.2 for T2, and 6.1 for T3.The result indicates that Task T2 receives the highest priority and should be executed first, followed by Task T1 and then Task T3. This example demonstrates how the priority function integrates real-time situational factors and trade-offs for efficient task scheduling. Moreover, in this scenario, even though reel replenishment is critical for maintaining continuous component feeding, the scheduler correctly elevates tray replacement (T2) to the highest priority because of its near-full capacity and minimal execution cost. This adaptive balance between upstream material availability and downstream output flow prevents potential bottlenecks across the full assembly process.

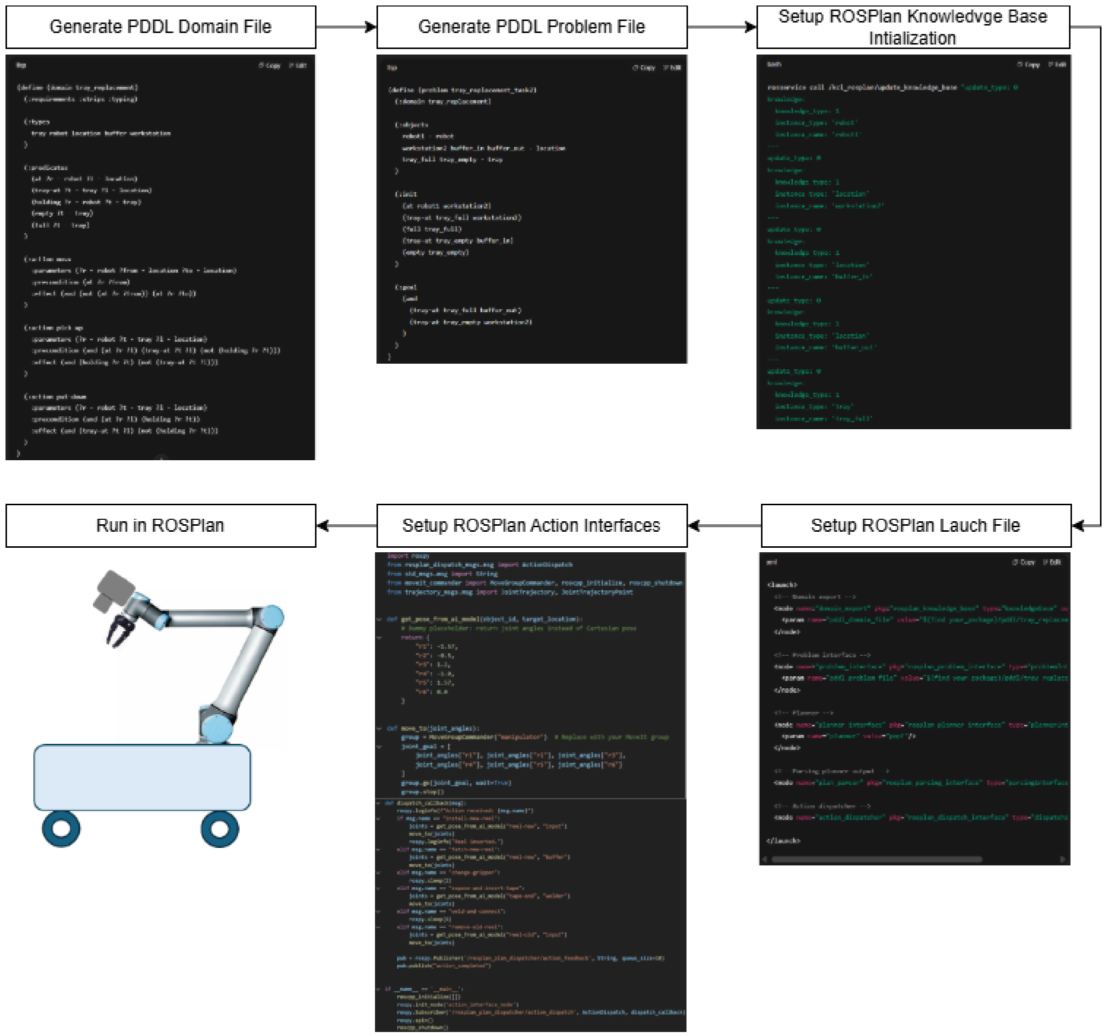

Figure 5. ROSPlan Execution Pipeline for Flexible Robot Task Planning

Figure 5 summarizes the workflow for generating and executing the robot motion task plan using the ROSPlan framework. The process begins with modeling the task scenario chosen by the priority function as PDDL domain and problem files, where robot actions, object types, and environmental states are formally defined. These files are then integrated into ROSPlan's knowledge base, initializing the system with the current scene configuration. Upon launch, the planner computes a feasible action sequence based on the provided domain knowledge and task goals. The generated plan is parsed into discrete executable actions, which are dispatched sequentially by ROSPlan's action dispatcher. Each dispatched action triggers corresponding robot control interfaces that interact with the physical robot and its perception modules, enabling real-world execution of the task plan. This integrated workflow ensures that high-level symbolic planning can be systematically translated into low-level robot operations, bridging the gap between abstract task reasoning and concrete manipulation in dynamic manufacturing environments.

8. Conclusion and Future Work

This paper presents a comprehensive smart robotic operations framework that integrates multi-camera perception, GPT-augmented ensemble learning, domain-driven task planning, and system-level coordination for dynamic manufacturing environments. By leveraging multiple synchronized camera views, the system achieves high-precision object localization and real-time situation awareness across complex work zones. The integration of multi-channel ensemble learning models enables adaptive fusion of multi-view pose estimates, yielding reliable robot control commands even under variable geometric and spatial conditions. Through the incorporation of GPT-augmented reasoning, the system automatically generates PDDL domain models and task plans, which are seamlessly executed via the ROSPlan framework. This closed-loop architecture bridges high-level symbolic reasoning with low-level robotic manipulation. The pilot application illustrates a representative use case in flexible electronics component assembly, demonstrating how the proposed system can be applied to coordinate material changeover and task prioritization across multiple workstations. While these results showcase the system's potential applicability, full-scale deployment and real-world validation remain as important next steps.

Future work will focus on conducting real-world experimental testing to verify system performance under industrial conditions, including robustness to unexpected environmental changes, human-robot interaction safety, and long-term operational stability. Additionally, efforts will be directed toward further automating domain knowledge generation, improving perception accuracy under highly cluttered or occluded scenes, extending task planning capabilities for more complex multi-step operations, and exploring integration with upstream production management systems to enable fully autonomous, closed-loop manufacturing optimization.

References

Ahn, M., et al., Do As I Can, Not As I Say: Grounding Language in Robotic Affordances, *arXiv preprint arXiv:2204.01691*, 2022.

Bandi, C., and Thomas, U., Action Recognition via Multi-View Perception Feature Tracking for Human–Robot Interaction, *Robotics*, vol. 14, no. 4, Article 53, 2025.

Bezrucav, S.-O., Canal, G., Cashmore, M., and Corves, B., An Action Interface Manager for ROSPlan. *9th ICAPS Workshop on Planning and Robotics at The 31st International Conference on Automated Planning and Scheduling (PlanRob)*, 2021.

Cashmore, M., Fox, M., Long, D., and Magazzeni, D., ROSPlan: Planning in the Robot Operating System, *Proceedings of the ICAPS 2015 PlanRob Workshop*, Jerusalem, Israel, 2015.

Chaabani, A., Cherif, R., and Yaddaden, Y., Automating quality control: real-time defect detection and automated decision-making with AI and Doosan robotics, *International Journal of Intelligent Robotics and Applications*, 2025.

Heuss, L., Gebauer, D., and Reinhart, G., Concept for the automated adaption of abstract planning domains for specific application cases in skills-based industrial robotics, *Journal of Intelligent Manufacturing*, vol. 35, pp. 4233–4258, 2024.

Kasaei, H., and Xiong, S., Lifelong ensemble learning based on multiple representations for few-shot object recognition, *Robotics and Autonomous Systems*, vol. 174, p. 104615, 2024.

Liang, J., Huang, W., Xia, F., Xu, P., Hausman, K., Ichter, B., Florence, P., and Zeng, A., Code as Policies: Language Model Programs for Embodied Control, *Proceedings of the IEEE International Conference on Robotics and Automation (ICRA)*, pp. 9493–9500, 2023.

Lin, Y., Tremblay, J., Tyree, S., Vela, P. A., and Birchfield, S., Multi-view fusion for multi-level robotic scene understanding, *Proceedings of the IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pp. 6817–6824, 2021.

Mon-Williams, R., Li, G., Long, R., Du, W., and Lucas, C. G., Embodied large language models enable robots to complete complex tasks in unpredictable environments, *Nature Machine Intelligence*, vol. 7, pp. 592–601, 2025.

Qadeer, N., Shah, J. H., Sharif, M., Dahan, F., Khokhar, F. A., and Ghazal, R., Multi-camera tracking of mechanically thrown objects for automated in-plant logistics by cognitive robots in Industry 4.0, *The Visual Computer*, vol. 40, pp. 9063–9082, 2024.

Seo, Y., Kim, J., James, S., Lee, K., Shin, J., and Abbeel, P., Multi-View Masked World Models for Visual Robotic Manipulation, *Proceedings of the 40th International Conference on Machine Learning (ICML)*, PMLR vol. 202, pp. 19411–19427, 2023.

Tan, R., Yang H., Jiao S., Shan L., Tao C., and Jiao R., Smart Robot Manipulation using GPT-4o Vision, *Proceedings of the 7th European Industrial Engineering and Operations Management Conference*, Augsburg (Greater Munich), Germany, July 16, 2024,

Terras, N., Pereira, F., Ramos Silva, A., Santos, A. A., Lopes, A. M., Silva, A. F. d., Cartal, L. A., Apostolescu, T. C., Badea, F., and Machado, J., Integration of Deep Learning Vision Systems in Collaborative Robotics for Real-Time Applications. *Applied Sciences*, *15*(3), 1336, 2025.

Zhao, Z., Lyu, J., Chu, Y., Liu, K., Cao, D., Wu, C., Qin, L., and Qin, S., Toward generalizable robot vision guidance in real-world operational manufacturing factories: A Semi-Supervised Knowledge Distillation approach, *Robotics and Computer-Integrated Manufacturing*, vol. 86, Article 102639, 2024.

Biographies

Yiyun (Cindy) Fei is a Ph.D. candidate in Mechanical Engineering at Georgia Institute of Technology, Georgia, USA, where she also earned her bachelor's and master's degree in the same field. Her research focuses on the modeling, analysis, and optimization of complex operational systems to support data-driven design and decision-making. Yiyun worked as a Manufacturing and Process Development Engineer at TE Connectivity, where she gained four years of hands-on industry experience. During her time there, she led and contributed to multiple projects involving AI-based machine vision, process automation, and discrete event simulation of manufacturing systems. More information is available on her LinkedIn profile: www.linkedin.com/in/yiyun-fei

Mulang Song is a Ph.D. candidate in the School of Mechanical Engineering at the Georgia Institute of Technology (USA). He is graduating in Mechanical Engineering, with research that spans both mechanical and industrial engineering domains. He received both his B.S. and M.S. degrees in Mechanical Engineering with a focus on Computer Science, also from Georgia Tech. He has led and contributed to engineering projects through direct collaboration with industry, applying research in real-world production environments. His research interests include advanced manufacturing systems, intelligent production systems, and industrial AI applications. More info about his research: https://scholar.google.com/citations?user=jLL4nycAAAAJ

Shu Wang is a smart manufacturing engineer at TE Connectivity. His expertise spans in industrial AI and data analytics, engineering design decision-making, human-automation interaction, and operations optimization. He holds a Ph.D. in Mechanical Engineering from Georgia Tech, along with dual master's degrees in Mechanical Engineering and Electrical and Computer Engineering. He earned his bachelor's degree in Electrical engineering from Wuhan University. More info about his research: https://scholar.google.com/citations?user=_R1ldZkAAAAJ&hl=en

Xinping Deng is an engineering manager focusing on smart factory with a focus on automation, robotics, industrial AI applications, manufacturing data analysis and innovation. He holds a master degree in Mechanical and Aerospace Engineering from Cornell University and a bachelor degree in Mechanical Engineering from Worcester Polytechnic Institute. More information is available on his LinkedIn profile: www.linkedin.com/in/xinping-deng

Roger Jiao is the editor-in-chief of Journal of Engineering Design and an associate professor of mechanical and industrial engineering at Georgia Tech, USA. Prior to joining the School of Mechanical Engineering at Georgia Tech in December 2008, he was an Assistant Professor and then Associate Professor in the School of Mechanical and Aerospace Engineering at Nanyang Technological University, Singapore. Before his career in Singapore, he was a Visiting Scholar in the Department of Industrial Engineering and Engineering Management at Hong Kong University of Science and Technology from 1998 to 1999. From 1993 to 1994, he was a Lecturer of Industrial Engineering in the School of Management at Tianjin University, China, and from 1988 to 1990, he worked as an Associate Lecturer in the Department of Industrial Design at Tianjin University of Science and Technology, China. More info about his research: https://scholar.google.com/citations?user=9yikEHAAAAAJ&hl=en&oi=ao.