Int j simul model 24 (2025) 4, 718-729
Original scientific paper

SIMULATION OF AIGC-ENHANCED CONGESTION MANAGEMENT IN FACTORY AUTONOMOUS DRIVING

Zhang, X. R.*,# & Zhou, J. N.**
* Zhijiang College of Zhejiang University of Technology, Shaoxing 312030, China
** Zhejiang Business College, Institute of Electronic Commerce, Hangzhou 310053, China
E-Mail: zhangxianrong@zzjc.edu.cn (# Corresponding author)

Abstract

With the advancement of intelligent manufacturing, Automated Guided Vehicles (AGVs) have become essential to material transport, and their efficiency directly influences production output. However, congestion from intersection conflicts and corridor bottlenecks remains a key bottleneck in multi-AGV scheduling. Existing control methods often lack robustness under extreme conditions, adaptability to dynamic networks, and balance among multiple objectives. This study proposes an integrated framework combining Artificial Intelligence Generated Content (AIGC)–based scenario generation with an improved Model Predictive Control (MPC) algorithm. An enhanced diffusion model generates diverse operating scenarios and uses least-squares calibration to reduce deviation from real conditions. The improved MPC employs a weighted objective and distributed optimization to enhance congestion mitigation, task efficiency, and stability. Experiments on a SUMO-ROS co-simulation platform with real factory data verify the framework's adaptability and effectiveness, providing a practical solution for autonomous traffic management in smart factories.

(Received in August 2025, accepted in October 2025. This paper was with the authors 1 month for 1 revision.)

Key Words: Factory Autonomous Driving, AGVs, Traffic Congestion Prediction and Intervention, AIGC, MPC, Simulation Modelling

1. INTRODUCTION

With the global manufacturing industry undergoing a profound transformation toward intelligent manufacturing [1-4], AGVs have become the core carriers for material handling in flexible production systems. Their traffic operational efficiency directly determines the stability of production line rhythm and the effectiveness of capacity output [5, 6]. However, as the scale of AGV deployment and the complexity of task coordination increase within factories, traffic congestion problems such as intersection conflicts [7, 8] and corridor bottlenecks [9, 10] in multi-AGV collaborative scheduling have become increasingly prominent. Industrial field data indicate that congestion-induced delays in material transport tasks can reach as high as 15–25 %, and in extreme scenarios, such congestion may even lead to production line interruptions [11], forming a critical bottleneck restricting efficiency improvement in smart factories.

In terms of factory AGV congestion prediction, existing studies can generally be classified into two categories. The first category is based on time-series models [12, 13], which learn temporal correlations in AGV operational states to achieve congestion prediction. However, these models overlook the spatial topological characteristics of factory road networks, resulting in limited prediction accuracy at critical nodes such as intersections. The second category is based on Graph Neural Networks (GNNs) and their variants [14, 15], which represent the road network as a graph structure to capture spatial dependencies. Nevertheless, the training of such models requires a large volume of labelled data, and their prediction robustness is often compromised under extreme operating conditions. Regarding intervention strategies, static path planning approaches are simple to implement but incapable of addressing dynamic congestion scenarios [16, 17]. Multi-Agent Reinforcement Learning (MARL) has been employed to enhance dynamic adaptability through agent-based collaborative optimization; however, its training process is complex, and the convergence speed is slow, making it difficult to satisfy real-time control requirements in industrial environments [18-20]. Overall, existing methods have yet to achieve an integrated framework that simultaneously ensures accurate prediction, real-time intervention, and multi-objective balance.

To address the core challenges in factory AGV congestion management – namely data scarcity, insufficient dynamic adaptability, and difficulty in achieving multi-objective coordination – this study aims to (a) design an AIGC-based scenario generation method adapted to enclosed factory road networks and dynamic operating conditions, enabling the efficient generation of both regular and extreme congestion scenarios; (b) develop an improved MPC algorithm incorporating multi-objective optimization, achieving an integrated framework for precise congestion prediction and real-time intervention; and (c) validate the proposed integrated framework through simulation experiments in terms of congestion control accuracy, system operational efficiency, and multi-objective balance capability, thereby providing an engineering-feasible solution for factory AGV congestion management.

The subsequent sections present the theoretical foundations, the design of the integrated framework, and the simulation experiments along with result analysis, followed by a summary of key findings and future research directions.

2. AIGC-ENHANCED MPC FOR CONGESTION CONTROL

2.1 Simulation framework architecture

To systematically address the core challenges in factory AGV congestion control – namely the scarcity of scenario data, insufficient adaptability of dynamic intervention, and the absence of end-to-end verification – a three-layer closed-loop integrated simulation architecture, termed the "AIGC scenario generation layer–improved MPC control layer–simulation verification layer", was proposed in this study. The core logic and interaction mechanisms of this architecture are below. The AIGC scenario generation layer takes as input the static parameters of the factory road network and dynamic operating commands. Through a customized generative model, it produces a comprehensive dataset of traffic scenarios encompassing normal, moderate, and extreme congestion conditions. The improved MPC control layer, serving as the central decision-making module, operates on two levels. On one hand, it receives the AIGC-generated scenario data to conduct offline algorithm training and parameter initialization. On the other hand, it accesses real-time operational states of AGVs and dynamic information of the road network, synchronously outputting predictions of congestion nodes and durations as well as intervention commands for AGV speed and path adjustment. The simulation verification layer, established within the SUMO-ROS co-simulation environment, replicates both AIGC-generated and real factory operational processes. It executes the MPC-based intervention commands and outputs performance indicators such as congestion mitigation rate and task completion rate. Meanwhile, the deviation between scenario data and actual control requirements is used as a feedback signal to adaptively adjust key parameters of the AIGC generative model, thereby achieving dynamic adaptation between scenario generation and algorithmic control.

2.2 AIGC scenario generation

Scenario generation requirement modelling serves as a fundamental prerequisite to ensure that the AIGC-generated outputs are aligned with the operational control needs of factory AGV systems. Accurate anchoring of static physical constraints and dynamic operational characteristics is required to construct a comprehensive parameter system. For static parameters, real factory Computer-Aided Design (CAD) drawings were adopted as data sources. Through a topological analysis tool, key road network information – such as intersection coordinates, corridor width (w), and turning radius – was extracted. Combined with AGV physical attributes, including rated load (m), maximum speed (v_{max}), and acceleration threshold, a static road network feature vector $\boldsymbol{X_S}$ with a dimensionality of $D_S = 12$ was constructed. For dynamic parameters, a multi-model coupling approach was employed. The AGV task model defines origin–destination pairs that follow a discrete probability distribution based on workshop material demand, assigning priority weights (ω_p) to the transport tasks of key production materials. The fault model encompasses four representative fault types – mechanical malfunction, communication interruption, and others – quantifying fault location and duration (t_f). Congestion-triggering conditions were established with a corridor occupancy threshold of $\rho_{th} = 0.75$ and a node conflict threshold of $n_{th} = 8$ within a five-minute window, thereby forming a comprehensive dynamic condition description system.

The improved diffusion model constitutes the core component for achieving high-fidelity scenario generation, enhancing scene adaptability through input-layer reconstruction and generation-process optimization. The input layer adopts a dual-modality fusion design combining feature vector and instruction encoding. The static road network feature vector ($\boldsymbol{X_S}$) is concatenated with the dynamic condition instruction vector ($\boldsymbol{X_D}$) to form a multidimensional input matrix of $D = 20$. Dynamic instructions are converted into model-interpretable numerical features using one-hot encoding. During the generation process, a dual-optimization mechanism was introduced. A spatial attention module was embedded within the U-Net backbone network, where feature weights are allocated based on the degree centrality of road network nodes, enabling the model to focus on detailed generation in high-congestion regions. In addition, an adaptive noise attenuation strategy was implemented, in which the noise scheduling factor (β_t) is dynamically adjusted according to the complexity of the operating conditions, effectively reducing redundant information unrelated to congestion features in the generated scenarios.

Scenario calibration and validity verification constitute a critical interface linking the AIGC-generated scenarios with the subsequent MPC algorithm, ensuring that the generated data maintain realism and diversity through quantitative evaluation. Calibration was conducted using three months of real factory AGV operational data as the benchmark. A three-dimensional feature indicator system was constructed, consisting of congestion duration (T_c), node occupancy ratio (ρ), and task delay rate (η). The optimal parameter set for the generated scenarios was obtained using a least-squares solution. The objective function is defined as:

$$min \sum_{i=1}^{N} \left(\alpha_1 \left| T_{c,i}^{gen} - T_{c,i}^{real} \right| + \alpha_2 \left| \rho_{c,i}^{gen} - \rho_{c,i}^{real} \right| + \alpha_1 \left| \eta_{c,i}^{gen} - \eta_{c,i}^{real} \right| \right) \tag{1}$$

where, the weight coefficients were set as $\alpha_1 = \alpha_2 = 0.4$ and $\alpha_3 = 0.2$, ensuring that the calibrated scenarios minimize deviations from real-world operational characteristics as much as possible. To assess scenario diversity, an entropy-based method was employed to quantify scenario coverage. The entropy value H of scenario distribution was computed across three dimensions: AGV density, task priority, and fault type. When $H \geq 0.85$, the scenario set is deemed to provide comprehensive coverage across three operating modes: (a) normal traffic conditions (AGV density $\leq$ 5 units/100 m^2), (b) moderate congestion (AGV density of 5–10 units/100 m^2), and (c) extreme congestion (AGV density $\geq$ 10 units/100 m^2 compounded with fault events), thereby providing full-spectrum training data support for the MPC algorithm.

2.3 MPC-based prediction and control

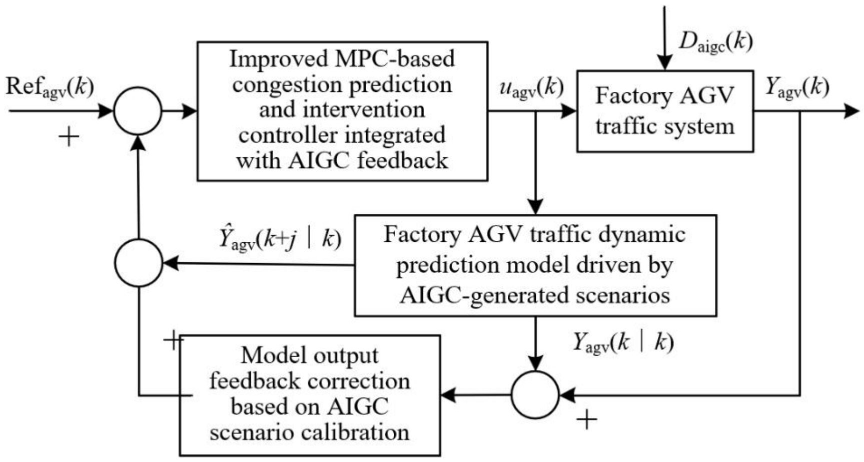


Figure 1: Framework of AIGC-integrated MPC for congestion control.

Fig. 1 illustrates the principle framework of the improved MPC-based congestion prediction and intervention control system integrated with AIGC-generated scenarios. This framework demonstrates an enhanced MPC loop that integrates AIGC-generated dynamic scenarios as perturbations, coupled with a dynamic prediction model and feedback correction mechanism, aiming to achieve closed-loop optimization of congestion prediction and intervention commands for factory AGV traffic states. The construction of the state equation serves as the fundamental basis for coupling congestion prediction with intervention, requiring simultaneous characterization of the dynamic correlation between individual AGV behaviour and the overall road network state. The system state vector at time step k is defined as $\boldsymbol{x}_k = [\boldsymbol{x}_{agv,k}{}^T, \boldsymbol{x}_{net,k}{}^T]^T$, where $\boldsymbol{x}_{agv,k} = [x_{1,k}, y_{1,k}, v_{1,k}, q_{1,k}, \ldots, x_{n,k}, y_{n,k}, v_{n,k}, q_{n,k}]^T$, with n representing the number of AGVs, includes each AGV's two-dimensional coordinates $(x_{i,k}, y_{i,k})$, velocity $v_{i,k}$, and remaining task quantity $q_{i,k}$, $\boldsymbol{x}_{net,k} = [\rho_{1,k}, r_{1,k}, \ldots, \rho_{m,k}, r_{m,k}]^T$, with m representing the number of road network nodes, encompassing the occupancy ratio $\rho_{j,k}$ and congestion risk value $r_{j,k}$. The congestion risk value $r_{j,k}$ is computed as a weighted function of the node occupancy ratio and conflict frequency. Let θ_i denote the heading angle of the AGV. Based on the AGV kinematic model $\dot{x}_i = v_i cos\theta_i$ and $\dot{y}_i = v_i\, sin\theta_i$, together with road network traversal rules, a linearized state transition equation was established as follows: $\boldsymbol{x}_{k+1} = \boldsymbol{A}\boldsymbol{x}_k + \boldsymbol{B}\boldsymbol{u}_k + \boldsymbol{w}_k$, where $\boldsymbol{u}_k = [v_{1,k}{}^*, \theta_{1,k}{}^*, \ldots, v_{n,k}{}^*, \theta_{n,k}{}^*]^T$ represents the control vector, including the desired velocities and heading angles of all AGVs. The matrix $\boldsymbol{A}$ denotes the state transition matrix, quantifying the influence coefficients of AGV motion on the overall road network state, while $\boldsymbol{B}$ denotes the control matrix. The term $\boldsymbol{w}_k$ is a Gaussian disturbance component introduced to account for uncertainties such as AGV speed fluctuations, thereby ensuring accurate system state prediction.

The multi-objective optimization function serves as the decision-making core of the improved MPC, designed to quantitatively balance three key objectives: congestion mitigation, task efficiency, and operational stability. A weighted objective function is formulated as $J = \omega_1 J_1 + \omega_2 J_2 + \omega_3 J_3$, where $J_1 = \sum_{j=1}{}^m \lambda_j r_{j,k}$ represents the congestion mitigation term, in which λ_j denotes the node importance weight, assigned as 1.2 for intersections and 1.0 for ordinary corridors. The term $J_2 = \sum_{i\in\Omega} |t_{i,k}{}^{act} - t_{i,k}{}^{exp}|$ corresponds to task efficiency, where Ω denotes the set of high-priority tasks, and $t_{i,k}{}^{act}$ and $t_{i,k}{}^{exp}$ refer to the actual and expected task completion times, respectively. The term $J_3 = \sum_{i=1n} |v_{i,k} - v_{i,k-1}|$ represents operational stability, capturing the magnitude of velocity fluctuations. A dynamic weight adjustment mechanism was designed based on real-time congestion level C_k and high-priority task ratio P_k. The dynamic weights are defined as $\omega_1 = 0.4C_k/(C_k + P_k)$, $\omega_2 = 0.4P_k/(C_k + P_k)$, and $\omega_3 = 0.2$, with $\omega_1 + \omega_2 + \omega_3 = 1$. This configuration ensures that ω_1 increases under severe congestion conditions, whereas ω_2 is increased under urgent task conditions. Fig. 2 presents the fuzzy decision-making subsystem of the factory AGV congestion management framework integrating AIGC and improved MPC. The structure illustrates the processes of feature fuzzification driven by AIGC scenarios such as congestion risk and velocity deviation, multi-objective inference, and control instruction generation, providing fuzzy decision-making support for the improved MPC.

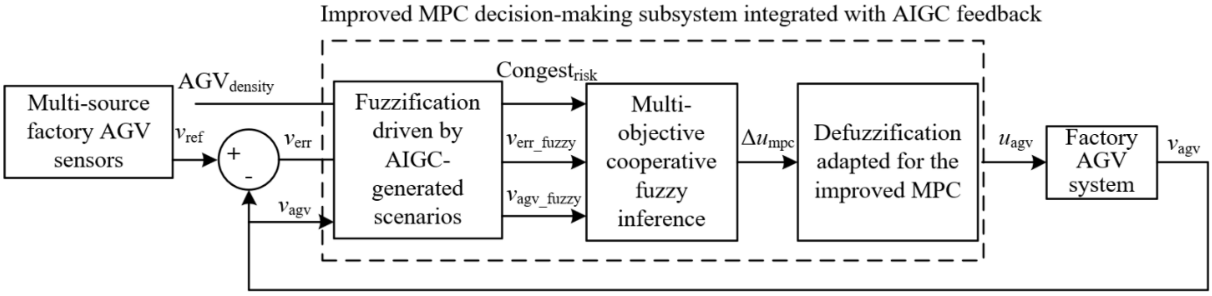


Figure 2: Core modules of the AIGC-MPC congestion management framework.

The dynamic constraint system provides feasible boundary conditions for MPC optimization, ensuring the implementability of intervention strategies by integrating three categories of constraints: physical, logical, and predictive. Physical constraints include (a) AGV velocity constraint $0 \leq v_{i,k} \leq v_{i,max}$, where $v_{i,max} \in [1.0, 2.0]$ m/s and dynamically adjusts according to load; (b) acceleration constraint $|a_{i,k}| \leq a_{max} = 0.2$ m/s^2; (c) safety distance constraint $(x_{i,k} - x_{j,k})^2 + (y_{i,k} - y_{j,k})^2 \geq d_{safe} = 0.5$ m $(i \neq j)$, preventing collisions among AGVs. Logical constraints are defined as follows: (a) right-of-way constraint for high-priority AGVs: when $\omega_{p,i} > \omega_{p,j}$, then $v_{i,k} \geq 0.8\, v_{i,max}$ and $v_{j,k} \leq 0.5\, v_{j,max}$; (b) restricted entry constraint for fault regions: when a fault occurs at node j (i.e., $\rho_{j,k} = 1$), AGV coordinates must not fall within the influence zone of the fault node. Predictive constraints are aligned with the AIGC-generated scenario outputs. Based on the predicted congested node set $\Phi_{k:T}$ over the future T time steps, the velocity of AGVs passing through these nodes is restricted by $v_{i,k+t} \leq 0.3\, v_{i,max}$ $(t = 1, \ldots, T)$, thereby enabling proactive avoidance of predicted congestion zones.

Computational complexity optimization is critical to ensuring MPC real-time performance. This is achieved through a distributed architecture and model simplification, maintaining control latency within 100 ms. A distributed MPC strategy based on “workshop partitioning and sub-controller coordination” was implemented. The factory was divided into M independent control zones, each assigned a sub-controller responsible for the regulation of n_m AGVs within its zone, satisfying $\sum_{m=1}^{M} n_m = n$. Sub-controllers exchange boundary AGV state information via Ethernet and achieve inter-zone coordination through a consensus algorithm, effectively preventing congestion propagation between regions. A model simplification strategy was further applied to reduce the state-space dimensionality for non-interacting AGVs. A distance threshold $L_0 = 50$ m is defined. For external AGVs located beyond this threshold from the current region, an “average velocity–equivalent occupancy ratio” model replaces their individual state representations. This simplification reduces the dimensionality of each sub-controller’s state variables by more than 40 %. Complexity analysis shows that the computational cost of the distributed architecture was $M \cdot O(n_m^3)$, which is approximately 65 % lower than that of a centralized MPC, denoted as $O(n^3)$. Under a scenario involving 50 AGVs, control latency stabilized around 85 ms, fully satisfying the real-time operational requirements of factory-scale AGV traffic management.

3. SIMULATION EXPERIMENT DESIGN AND IMPLEMENTATION

The design of the experimental environment and dataset was conducted based on the core principles of realism, comprehensive scenario coverage, and reproducibility, ensuring reliable support for method validation. The hardware platform consisted of an Intel Xeon 8375C CPU, an NVIDIA A100 GPU, and 128 GB DDR4 memory, which provided sufficient computational capacity for both AIGC model training and real-time computation of the improved MPC. On the software layer, a SUMO–ROS co-simulation environment was established. SUMO was employed for traffic flow modelling and congestion state visualization, while ROS, equipped with the MoveIt! package, was utilized to perform AGV motion control. The development framework adopted PyTorch 2.0 for AIGC model training and MATLAB R2023a for MPC algorithm solving.

The factory road network was reconstructed from the actual CAD drawings of an automotive parts manufacturing facility. The reconstructed layout comprised five workshops covering a total area of 4,800 m², eight main corridors with widths ranging from 1.5 to 3.0 m, and twelve intersections. The AGVs adopted in the simulation were of the latent-type model, featuring a rated load capacity of 1,000 kg and a maximum velocity of 1.5 m/s. The AGV quantity was dynamically adjustable from 5 to 50 units to simulate varying load intensities across different operational conditions. The dataset was divided into two categories. The real dataset was collected from AGV operational data recorded between January and March 2024 in the aforementioned factory. It comprised 200 regular-condition cases and 50 congestion-condition cases, with the latter encompassing scenarios such as AGV faults and sudden task surges. Key features – including congestion duration and node occupancy ratio – were extracted and used for model calibration. The AIGC-generated dataset was produced by the improved diffusion model, containing 1,000 generated scenarios that comprehensively covered all combinations of AGV quantities (5, 20, and 50 units), task density levels (low, medium, and high), and fault types (mechanical, communication, and battery-related). This dataset was utilized for the offline training and generalization performance verification of the improved MPC.

The experimental evaluation and comparative design were developed to address three core objectives – method effectiveness, module contribution, and performance superiority – to ensure the scientific rigor and persuasiveness of the validation results. The evaluation index system was categorized into three groups. Prediction performance was assessed using Mean Absolute Error (*MAE*), Mean Squared Error (*MSE*), and congestion prediction accuracy. Intervention performance was evaluated based on congestion mitigation rate, high-priority task completion rate, and average AGV travel efficiency. System performance was measured in terms of control command latency and algorithmic computational complexity. Four types of comparative experiments were designed: (a) Scenario generation effectiveness comparison: Real factory scenarios were used as benchmarks to compare the generated scenario feature errors of the improved diffusion model against those of traditional Generative Adversarial Network (GAN) and baseline diffusion models. (b) Prediction algorithm comparison: Under different AGV population scales, the prediction performance of the improved MPC module was quantitatively compared with those of conventional MPC, Long Short-Term Memory (LSTM), and Graph Attention Network (GAT) models. (c) Intervention algorithm comparison: The intervention performance of the improved MPC was evaluated against Dijkstra-based static path planning and MARL methods. (d) Ablation experiments: The contributions of key components – including the AIGC scenario calibration mechanism, MPC dynamic weight adjustment, and distributed optimization strategy – were isolated and verified by sequentially removing each module from the framework. All experiments were repeated ten times, and the mean results were computed. Statistical significance was tested using *t*-

tests, with $p < 0.05$ indicating significance, thereby ensuring the reliability of the experimental conclusions.

4. RESULTS AND ANALYSIS

To validate the fidelity and extreme-condition coverage of the scenarios generated by the improved diffusion model and to ensure that the resulting datasets could effectively support subsequent MPC algorithm training, the above comparative experiments were conducted. As shown in Table I, the improved diffusion model demonstrates significant superiority over both the baseline diffusion model and the traditional GAN model across key performance indicators. The comprehensive congestion feature error of the improved diffusion model is only 4.20 %, representing reductions of 31.15 % and 51.72 % compared to the baseline diffusion and traditional GAN models, respectively, exhibiting a very high degree of consistency with real-world scenario features. In terms of scenario diversity and adaptability to extreme conditions, the improved diffusion model achieves a scenario diversity entropy of 0.89, which closely approximates that of the real scenario (0.92), and an extreme congestion scenario coverage of 97.30 %, substantially exceeding the coverage of the baseline diffusion (86.40 %) and traditional GAN (71.20 %) models. These results clearly indicate that the improved diffusion model effectively resolves the long-standing limitations of conventional AIGC-based scenario generation – specifically the issues of insufficient realism and inadequate extreme-condition coverage, providing high-quality and representative training and testing datasets for the MPC algorithm.

Table I: Comparison of AIGC scenario generation performance.

Evaluation metric	**Real scenario**	**Improved diffusion model**	**Baseline diffusion model**	**Traditional GAN model**
Comprehensive congestion feature error (%)	0.00	4.20	6.10	8.70
Scenario diversity entropy	0.92	0.89	0.76	0.61
Extreme congestion scenario coverage (%)	100.00	97.30	86.40	71.20

To validate the accuracy, temporal robustness, and stability under high-dynamic scenarios of the improved MPC prediction module, a series of multidimensional comparative experiments were conducted to evaluate its capability in overcoming the performance limitations of conventional prediction algorithms. As shown in Table II, the improved MPC exhibits significant advantages across all major prediction metrics. The overall *MAE* reaches 0.32 s, representing reductions of 28.6 %, 19.3 %, and 8.6 % compared with the traditional MPC, LSTM, and GAT networks, respectively. The congestion prediction accuracy achieves 91.50 %, marking an improvement of up to 17.01 % over the highest-performing comparison algorithm, demonstrating more precise identification of congestion nodes and duration. From the perspective of temporal scale and adaptability to high-load conditions, the improved MPC achieves an *MAE* of 0.33 s under a 10-second prediction horizon. In high-dynamic scenarios with 50 AGVs, the *MAE* remains as low as 0.38 s, consistently outperforming the comparison algorithms and highlighting its superior stability. These findings confirm that the improved MPC, through its multi-objective optimization structure and dynamic constraint design, effectively captures both the topological correlations of factory road networks and the dynamic task characteristics of AGVs. Consequently, the prediction accuracy and robustness achieved by the improved MPC surpass those of traditional algorithms, providing a reliable decision-making foundation for subsequent congestion intervention.

Table II: Comparison of congestion prediction performance.

Evaluation metric	Improved MPC	Traditional MPC	LSTM network	GAT network
MAE (s)	0.32	0.45	0.39	0.35
Congestion prediction accuracy (%)	91.50	78.20	83.60	86.40
MAE for 10 s prediction horizon (s)	0.33	0.47	0.40	0.36

a) AIGC-generated regular operating condition

b) AIGC-generated fault condition

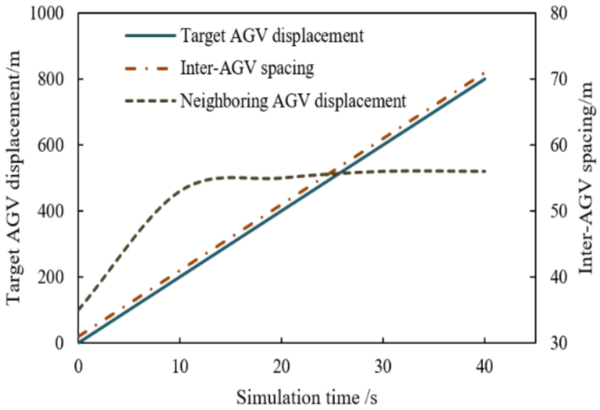
c) AIGC-generated task surge condition

Figure 3: AIGC-generated simulation scenarios of AGV congestion.

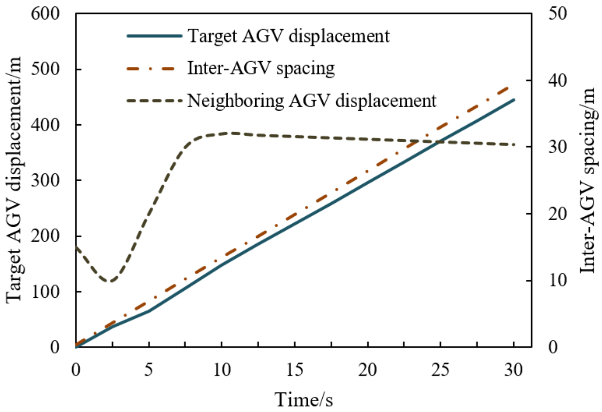
Fig. 3 illustrates the simulation scenarios of factory logistics AGV traffic congestion generated through AIGC under various operational conditions. In the regular operating condition (Fig. 3 a), logistics AGVs of different colours maintain orderly spacing along dual-lane pathways. The AIGC model accurately reproduces the spatial distribution and kinematic characteristics of AGVs in normal factory logistics operations. Under the fault condition (Fig. 3 b), the purple AGV is shown halted due to a malfunction, forming a localized congestion zone, while the green AGVs exhibit dynamic avoidance behaviour realistically captured by the AIGC model. This demonstrates the model's precision in modelling abnormal operational scenarios. In the task surge condition (Fig. 3 c), multiple AGVs of various colours are densely distributed across dual-lane routes, presenting the typical characteristics of high-load congestion. This verifies the capability of AIGC to comprehensively cover extreme operational conditions.

To verify the coordinated control performance of the improved MPC in managing AGV displacement and inter-vehicle spacing under AIGC-generated full-condition scenarios, a series of comparative experiments across multiple operating conditions was conducted. As illustrated in Fig. 4, under the AIGC-generated regular operating condition (Fig. 4 a), the displacement of the target AGV exhibits a linear and synchronized growth pattern with that of neighbouring AGVs. The inter-AGV spacing remains consistently within the safe interval of 50 ± 5 m, demonstrating precise control capability for regular traffic flow regulation. Under the fault condition (Fig. 4 b), where neighbouring AGVs experience displacement fluctuations caused by malfunctions, the improved MPC dynamically adjusts the velocity of the target

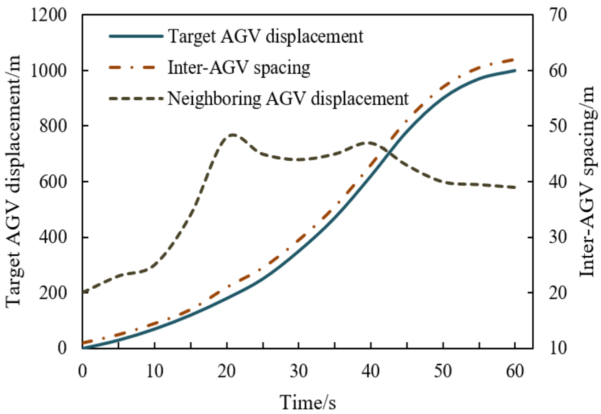
AGV. This adjustment maintains spacing fluctuations within 10 m and subsequently restores the distance to a stable equilibrium of 30 m, confirming the model's robustness in dynamic intervention under fault-induced disturbances. In the task surge condition (Fig. 4 c), although the displacements of neighbouring AGVs exhibit multi-peak fluctuations due to task concentration, the improved MPC enables the target AGV to maintain smooth displacement following, with inter-AGV spacing consistently controlled within a reasonable range of 40 ± 10 m. This demonstrates that the improved MPC successfully achieves a multi-objective balance between displacement efficiency and spacing safety under high-load operating conditions.



a) AIGC-generated regular operating condition



b) AIGC-generated fault condition



c) AIGC-generated task surge condition

Figure 4: Comparison of factory AGV displacement–spacing control performance of the improved MPC under AIGC-generated scenarios.

Table III: Comparison of congestion intervention performance.

Evaluation metric	Improved MPC	Traditional MPC	Dijkstra static planning	MARL
Congestion mitigation rate (%)	31.20	18.50	12.30	24.70
High-priority task completion rate (%)	95.30	80.10	75.60	88.40
Average task completion time (s)	286.50	352.80	389.20	315.70
Average AGV travel efficiency (km/h)	1.27	1.00	0.92	1.13
Average AGV energy consumption (kWh/10 km)	2.35	2.78	3.02	2.51
Extreme congestion intervention success rate (%)	89.60	65.30	52.10	78.20
Control command latency (ms)	85.00	150.00	40.00	220.00
Computational complexity (Giga Floating Point Operations per Second, GFLOPs)	28.60	49.30	8.20	65.70

Table IV: Comparison of ablation experiment performance.

Evaluation metric	Improved MPC (baseline)	Without AIGC scenario calibration	Without dynamic weight adjustment	Without distributed optimization
Prediction MAE (s)	0.32	0.41	0.33	0.32
Congestion mitigation rate (%)	31.20	22.10	25.80	29.70
High-priority task completion rate (%)	95.30	90.20	82.70	94.80
Total task delay rate (%)	8.70	13.50	16.20	9.20
Average AGV energy consumption (kWh/10 km)	2.35	2.53	2.68	2.38
Control command latency (ms)	85.00	83.00	86.00	180.00
Computational complexity (GFLOPs)	28.60	27.90	28.30	49.30

To validate the comprehensive advantages of the improved MPC in congestion intervention, multi-objective balance, and system real-time performance, as well as to quantify the performance gap relative to conventional intervention algorithms, multidimensional comparative experiments were conducted. As shown in Table III, the improved MPC demonstrates significant superiority across all key intervention metrics. The congestion mitigation rate reaches 31.20 %, representing improvements of 68.65 %, 153.66 %, and 26.32 % compared with traditional MPC, Dijkstra static planning, and MARL, respectively. The high-priority task completion rate achieves 95.30 %, exceeding the comparative algorithms by 19.09 %, 26.06 %, and 7.81 %, respectively, while the average task completion time is reduced to 286.50 s, highlighting the model's strong capability in ensuring task execution efficiency. From the perspective of operational efficiency and adaptability to extreme scenarios, the improved MPC enhances average AGV travel efficiency to 1.27 km/h and reduces average energy consumption to 2.35 kWh/10 km. Furthermore, the intervention success rate under extreme congestion achieves 89.60 %, demonstrating superior robustness and adaptability compared with the benchmark algorithms. At the system level, the improved MPC maintains a control command latency of approximately 85 ms, satisfying the real-time control requirements of factory operations. Its computational complexity is only 28.60 GFLOPs, reflecting a 42.0 % reduction compared with traditional MPC, thus effectively addressing the computational burden of MARL and the limited real-time responsiveness of conventional MPC. These findings confirm that the improved MPC, through multi-objective optimization and distributed strategy design, achieves coordinated optimization across congestion mitigation, task efficiency, and system stability, delivering superior intervention performance and engineering applicability relative to conventional algorithms.

To quantitatively evaluate the independent contribution of the three core modules – AIGC scenario calibration, MPC dynamic weight adjustment, and distributed optimization – and to clarify their respective impacts on the overall framework performance, a series of ablation experiments was conducted. As presented in Table IV, all three modules play critical roles in supporting the performance of the integrated framework. When the AIGC scenario calibration mechanism was removed, the prediction *MAE* increased from 0.32 s to 0.41 s, the congestion mitigation rate declined to 22.10 %, and the total task delay rate rose to 13.50 %. These results confirm that the calibration mechanism significantly enhances the realism of generated scenarios, thereby providing high-quality data foundations for precise congestion prediction and intervention. Upon the removal of the dynamic weight adjustment mechanism, the high-priority task completion rate dropped sharply from 95.30 % to 82.70 %, while both the total task delay rate and average AGV energy consumption increased. This outcome highlights the mechanism's central function in maintaining balance among the multiple objectives of congestion mitigation, task efficiency, and operational stability. The removal of the distributed optimization strategy caused the control command latency to surge from 85 ms to 180 ms, and the computational complexity to increase from 28.60 GFLOPs to 49.30 GFLOPs, far exceeding the real-time control threshold required in factory operations. This result clearly demonstrates the irreplaceable role of distributed optimization in reducing computational load and ensuring real-time responsiveness. In addition, the removal of individual modules did not significantly affect non-associated core metrics – for instance, the prediction *MAE* remained unchanged after the distributed optimization module was removed – further validating the modular rationality of the framework design. In summary, the three core modules jointly support the high-performance operation of the improved framework, demonstrating both the necessity and scientific soundness of their design and integration.

5. CONCLUSION

To address the core challenge of traffic congestion management in factory autonomous driving systems, an integrated solution combining AIGC scenario generation and improved MPC was developed in this study. This framework achieves a full-chain innovation that encompasses scenario generation, congestion prediction, and dynamic intervention. The research findings indicate that the improved diffusion model outperforms traditional generative models in both congestion feature consistency and coverage of extreme operating conditions, providing high-fidelity and full-condition training data for the subsequent algorithm. The improved MPC, through multi-objective optimization, dynamic constraint design, and a distributed architecture, exhibits comprehensive superiority in congestion prediction accuracy, intervention effectiveness, and system real-time performance. These advancements enable coordinated optimization of congestion mitigation, task efficiency, and system stability within factory AGV traffic networks. This study provides a novel technological pathway for the intelligent upgrading of factory AGV traffic systems. It not only promotes the practical implementation of AIGC technology in complex industrial scenarios but also extends the optimization boundaries of MPC in dynamic multi-objective systems. The proposed framework possesses significant theoretical innovation value and engineering application relevance.

REFERENCES

[1] Mou, J. B. (2024). Multi-objective optimization for resource allocation in intelligent manufacturing, *International Journal of Simulation Modelling*, Vol. 23, No. 2, 359-370, doi:10.2507/IJSIMM23-2-CO9

[2] Babaeimorad, S.; Fattahi, P.; Fazlollahtabar, H.; Shafiee, M. (2024). An integrated optimization of production and preventive maintenance scheduling in Industry 4.0, *Facta Universitatis, Series: Mechanical Engineering*, Vol. 22, No. 4, 711-720, doi:10.22190/FUME230927014B

[3] Weng, L. L. (2024). Dynamic scheduling for manufacturing workshops using Digital Twins, Competitive Particle Swarm Optimization, and Siamese Neural Networks, *Advances in Production Engineering & Management*, Vol. 19, No. 3, 301-314, doi:10.14743/apem2024.3.508

[4] Monek, G. D.; Szántó, N.; Korpai, R.; Kocsis Szürke, S.; Fischer, S. (2023). Enabling legacy lab-scale production systems: a Digital Twin approach at Széchenyi István University, *Journal of Engineering Management and Systems Engineering*, Vol. 2, No. 4, 224-230, doi:10.56578/jemse020404

[5] Isik, M.; Sahin, C.; Hamidy, S. M. (2023). Novel dispatching rules for multiple-load automated guided vehicles, *International Journal of Simulation Modelling*, Vol. 22, No. 1, 76-87, doi:10.2507/IJSIMM22-1-632

[6] Khalil, R. I.; Edan, N. M. (2025). Real-time object detection for forklift automated guided vehicles using deep learning, *International Journal of Transport Development and Integration*, Vol. 9, No. 3, 665-676, doi:10.18280/ijtdi.090319

[7] Nishi, T.; Hiranaka, Y.; Grossmann, I. E. (2011). A bilevel decomposition algorithm for simultaneous production scheduling and conflict-free routing for automated guided vehicles, *Computers & Operations Research*, Vol. 38, No. 5, 876-888, doi:10.1016/j.cor.2010.08.012

[8] Digani, V.; Sabattini, L.; Secchi, C.; Fantuzzi, C. (2015). Ensemble coordination approach in multi-AGV systems applied to industrial warehouses, *IEEE Transactions on Automation Science and Engineering*, Vol. 12, No. 3, 922-934, doi:10.1109/TASE.2015.2446614

[9] Ye, W.-B., Wang, H.-F., Chen, W.-D. (2025). Path coordination for robust, fast, and scalable multi-agent path finding under unforeseen delays, *IEEE Transactions on Automation Science and Engineering*, Vol. 22, 22396-22409, doi:10.1109/TASE.2025.3615989

[10] Vivaldini, K.; Rocha, L. F.; Martarelli, N. J.; Becker, M.; Moreira, A. P. (2016). Integrated tasks assignment and routing for the estimation of the optimal number of AGVS, *The International Journal of Advanced Manufacturing Technology*, Vol. 82, Nos. 1-4, 719-736, doi:10.1007/s00170-015-7343-4

[11] Kabir, Q. S.; Suzuki, Y. (2018). Increasing manufacturing flexibility through battery management of automated guided vehicles, *Computers & Industrial Engineering*, Vol. 117, 225-236, doi:10.1016/j.cie.2018.01.026

[12] Jiang, W.; Luo, J. (2022). Graph neural network for traffic forecasting: a survey, *Expert Systems with Applications*, Vol. 207, Paper 117921, 28 pages, doi:10.1016/j.eswa.2022.117921

[13] Bui, K.-H. N.; Cho, J., Yi, H. (2022). Spatial-temporal graph neural network for traffic forecasting: an overview and open research issues, *Applied Intelligence*, Vol. 52, No. 3, 2763-2774, doi:10.1007/s10489-021-02587-w

[14] Guo, S.; Lin, Y.; Feng, N.; Song, C.; Wan, H. (2019). Attention based spatial-temporal graph convolutional networks for traffic flow forecasting, *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 33, No. 1, 922-929, doi:10.1609/aaai.v33i01.3301922

[15] Derrow-Pinion, A.; She, J.; Wong, D.; Lange, O.; Hester, T.; Perez, L.; Nunkesser, M.; Lee, S.; Guo, X.; Wiltshire, B.; Battaglia, P. W.; Gupta, V.; Li, A.; Xu, Z.; Sanchez-Gonzalez, A.; Li, Y.; Velickovic, P. (2021). ETA prediction with graph neural networks in Google Maps, *Proceedings of the 30th ACM International Conference on Information & Knowledge Management (CIKM '21)*, 3767-3776, doi:10.1145/3459637.3481916

[16] Ma, J.; Xu, M.; Meng, Q.; Cheng, L. (2020). Ridesharing user equilibrium problem under OD-based surge pricing strategy, *Transportation Research Part B: Methodological*, Vol. 134, 1-24, doi:10.1016/j.trb.2020.02.001

[17] Murakami, K. (2020). Time-space network model and MILP formulation of the conflict-free routing problem of a capacitated AGV system, *Computers & Industrial Engineering*, Vol. 141, Paper 106270, 10 pages, doi:10.1016/j.cie.2020.106270

[18] Sartoretti, G.; Kerr, J.; Shi, Y.; Wagner, G.; Satish Kumar, T. K.; Koenig, S.; Choset, H. (2019). PRIMAL: pathfinding via reinforcement and imitation multi-agent learning, *IEEE Robotics and Automation Letters*, Vol. 4, No. 3, 2378-2385, doi:10.1109/LRA.2019.2903261

[19] Wallner, M.; Peráček, T. (2023). Predictors of average wait time at an intersection using artificial neural network, *International Journal for Quality Research*, Vol. 17, No. 3, 867-890, doi:10.24874/IJQR17.03-15

[20] Rashid, T.; Samvelyan, M.; Schroeder De Witt, C.; Farquhar, G.; Foerster, J.; Whiteson, S. (2020). Monotonic value function factorisation for deep multi-agent reinforcement learning, *Journal of Machine Learning Research*, Vol. 21, No. 1, Paper 178, 7234-7284