DT-PosFormer : A Digital Twin Enabled Transformer Network for Precise Pose Estimation and Trajectory Prediction of MiC Modules.

Yujie Han, Zhiheng Zhao, George Q. Huang

5/31 2024

The development of MiC and digital twin

Offsite Modular Prefabrication

Industry 4.0

Industrialized Construction

Hong Kong MiC

Singapore ICPHs
bca.gov.sg hdb.gov.sg

Australia

MiC

The Chief Executive's 2022 Policy Address

To foster the development of MiC

The Chief Executive delivered the 2022 Policy Address on 19 October. To enhance Quantity, Speed, Efficiency and Quality, the Government will increase public housing and land supply by different measures, including the building of Light Public Housing by MiC to shorten the waiting time of public housing. The Housing Authority and Housing Society will adopt the MiC in more public housing projects. The DEVB will establish a cross-departmental steering committee, to streamline approval processes, and also making available land in the Northern Metropolis for manufacturing and storage of modules.

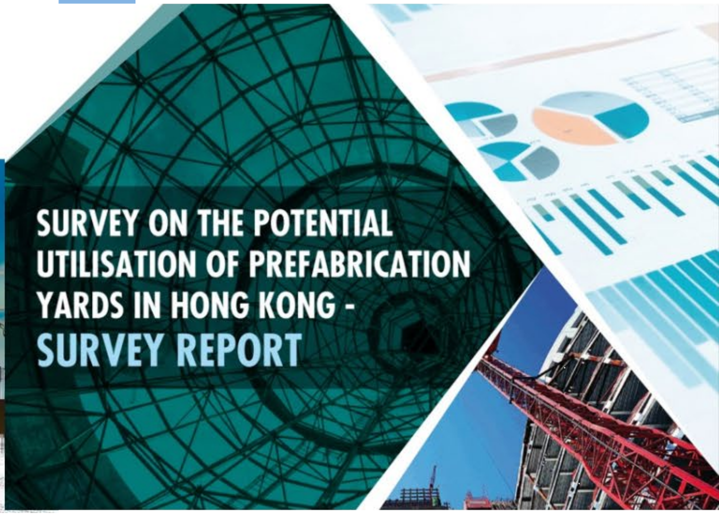
The Chief Executive's 2022 Policy Address

Uplift the Productivity of the Construction Industry

The Chief Executive's 2022 Policy Address

Public Housing Supply (1)

Every year in past
5 and 10 years



The special condition in Hong Kong

Modular Integrated Construction (MiC) is progressively replacing traditional construction.

Traditional construction site.
Complicated construction sites,
Labor-intensive industries.

Goes against the Smart Site Safety System (4S)
Construction Innovation and Technology Fund (CITF)

MiC construction site.
Improved construction quality,
simplified management,
reduced construction duration

Hong Kong

- **High density**
- **High rise**
- **High labor costs**
- **High safety requirement**
- **Limited site**

Substantial Efforts: ITFs + RIF + SPPR + Consulting

Smart Manufacturing

Smart Logistics

ITP/045/13LPI
ITF 2014-16

ITT/003/18LP
ITF 2018-19

R7027-18
RIF 2019-23

Construction Blockchain Consulting
Industry 2019

CIC MiC
CIC 2019

S2019.A8.013.19S
SPPR 2020

PRP/068/20LI
ITF 2020-22

- Substantial Funding
- Close Industrial Collaboration
- Extensive Publications
- Academic insights

Prefab Production

Cross-border Logistics

Construction Site



Research Gaps and Further Requirements

"The faster the stacking process, the less the project duration and costs"

For the construction process

- Rely on the worker observation and the expertise of crane drivers
- Limited observation perspective and delayed stacking control

Safety accidents   Installation errors   Module collision

For the stakeholders

- Slow manual stacking process
- Limited management and information sharing

Module accumulation   Delays in project delivery
Affecting customer satisfaction and the company's reputation

Further requirements

Comprehensive stacking perception

Dynamic and rapid response to emergent incidents

Predictive decision-making

Efficient information transmission and sharing among stakeholders

Predictive information and spatial and temporal traceability are the most important.

Analysis of feasible methods for realizing spatio-temporal tracking

Real-time spatio-temporal information collection, analysis, and visualization for module pose estimation and trajectory prediction may be the only viable and revolutionary solution

Mainstream method can be divided into visual-based and nonvisual based method.

Computer Vison

- **Environment**

Lighting conditions, cluttered background, and surface reflections of modules

Vibrations and shacking caused by construction activities

The camera perspectives are constrained by the deployment positions and angles

Image blurring and noise interference

- **Cost**

Full coverage camera on building sites is costly

Camera is easily vandalized and costly to maintain.

The cost of calculating full-scene data is huge.

Non-visual sensor-based method

- **Environment**

Ease of deployment

Long-term operational stability

Adaptability to variable environments

- **Cost**

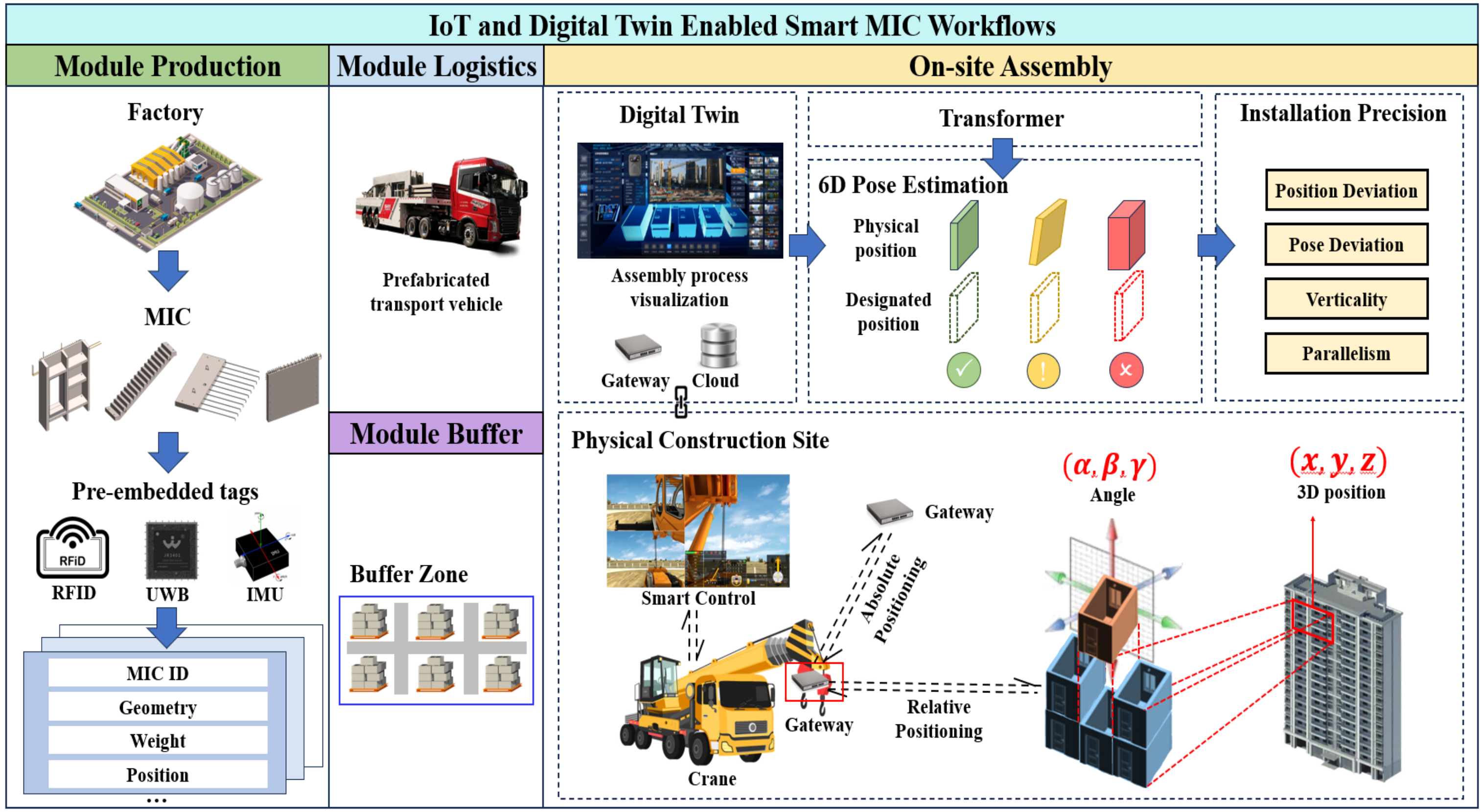
Lower cost

Real-time 3D position information –UWB (Ultra-Wide Band)

Real-time 3D Attitude information –IMU Inertial Measurement Unit

The workflows of smart MiC

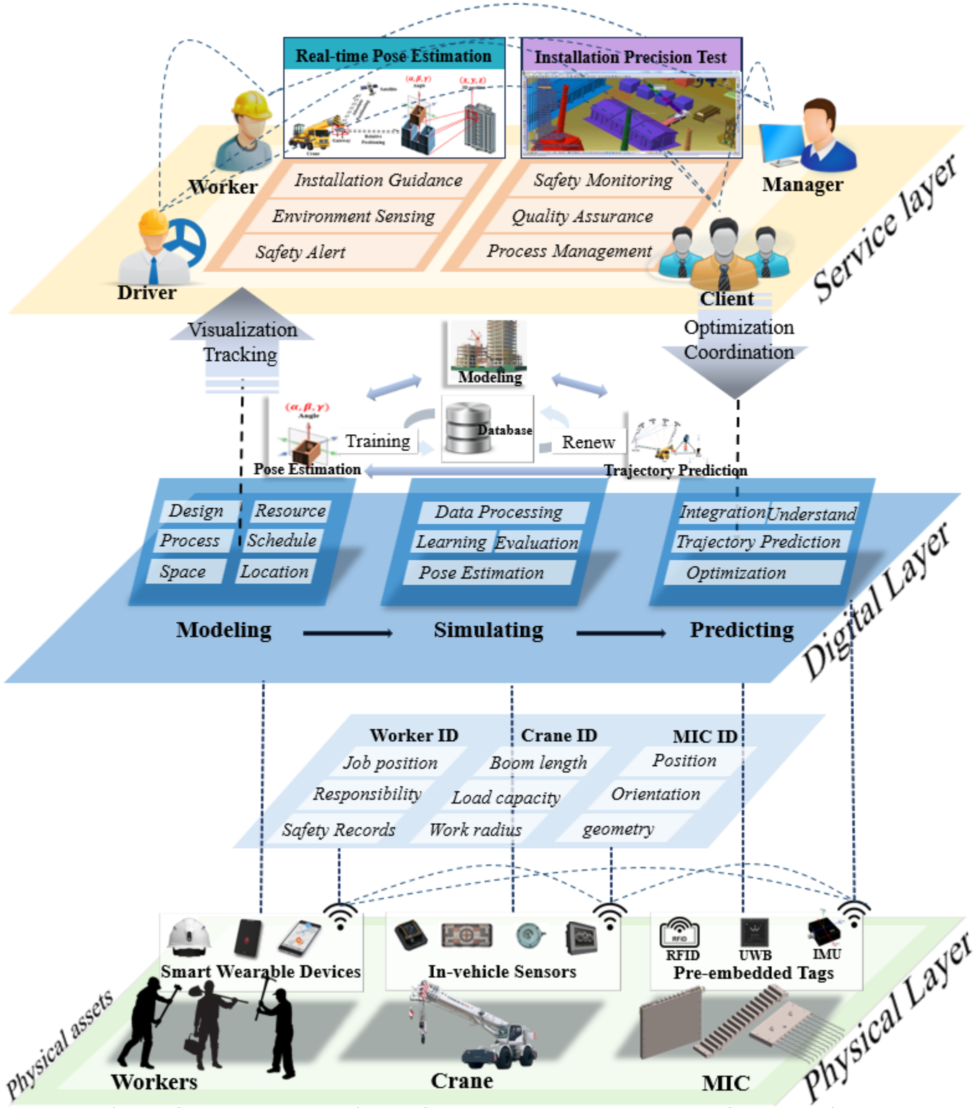
A DT-PosFormer-enabled pose estimation and trajectory prediction system (DTPTS) for MiC on-site assembly.



The workflows of smart MiC

- A digital twin platform based on IoT is proposed for real-time perception and bidirectional interoperability
- A novel network based on Transformer is embedded into the digital twin for spatio-temporal module trajectory prediction
- Comparative studies with the current state-of-the-art DL models are conducted to validate the feasibility of the system and the effectiveness of the proposed method

The framework of DTPT system for MiC



The framework of DTPT system for MiC

Service layer

The service layer creates a real-time visualization and information sharing platform for all stakeholders together. It primarily encompasses two types of services: 1) Guidance service. 2) Management service.

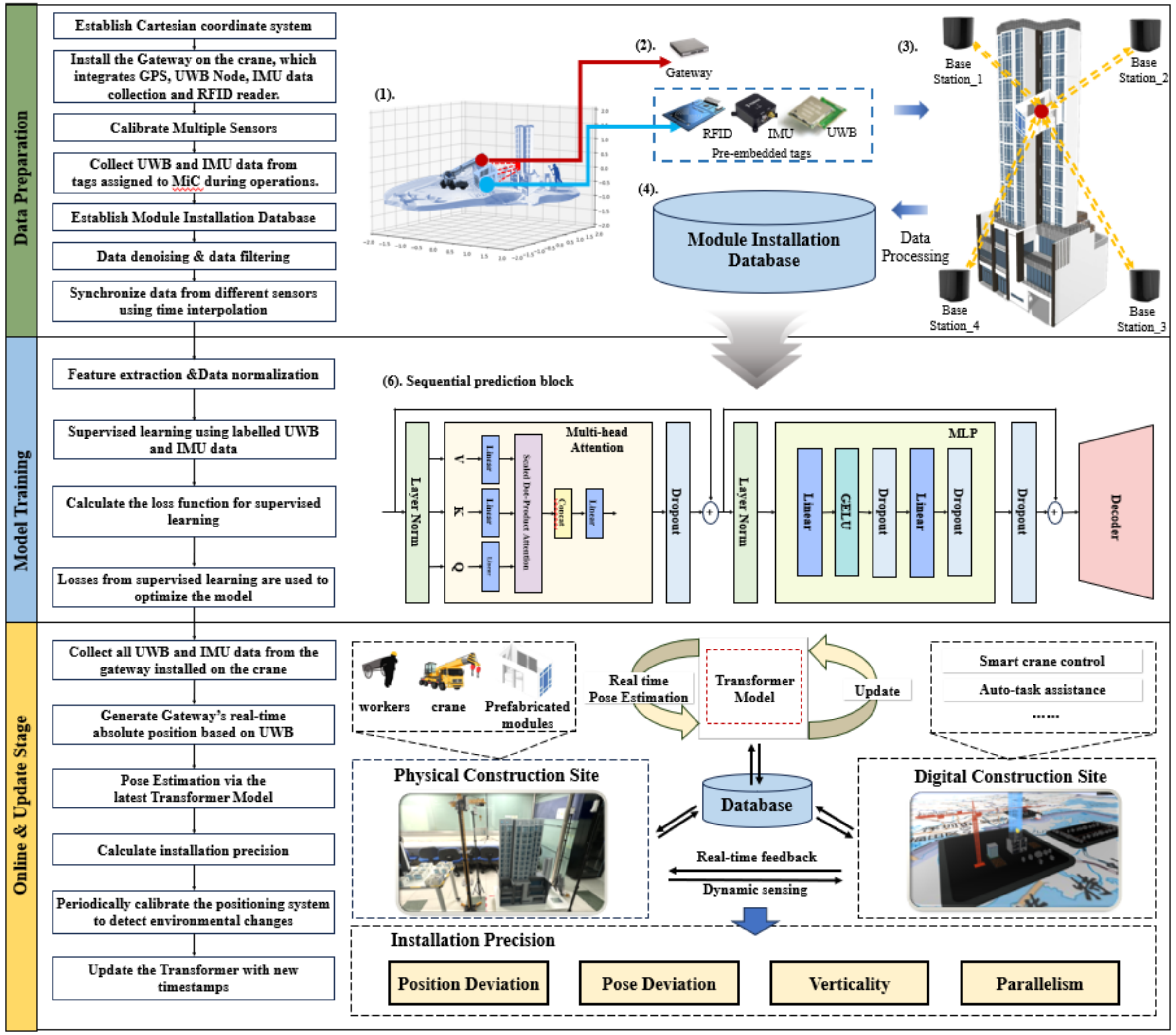
Virtual layer

The virtual layer mainly realizes the unified management of the data uploaded by the physical layer and establishes a high-fidelity digital twin of the stacking process.

Physical layer

The on-site assets are embedded in smart IoT devices for intelligent sensing. Multi-dimensional and multi-scale data of the internal characteristics and dynamic activities are captured by the IoT.

Pose estimation and trajectory prediction



The past trajectory of module i:

$$X_i = (x_i^t, y_i^t, z_i^t, \varphi_i^t, \theta_i^t, \psi_i^t)$$

The true trajectory of module i :

$$Y_i = \left(x_i^{t+t_{\text{pred}}}, y_i^{t+t_{\text{pred}}}, z_i^{t+t_{\text{pred}}}, \varphi_i^{t+t_{\text{pred}}}, \theta_i^{t+t_{\text{pred}}}, \psi_i^{t+t_{\text{pred}}}\right)$$

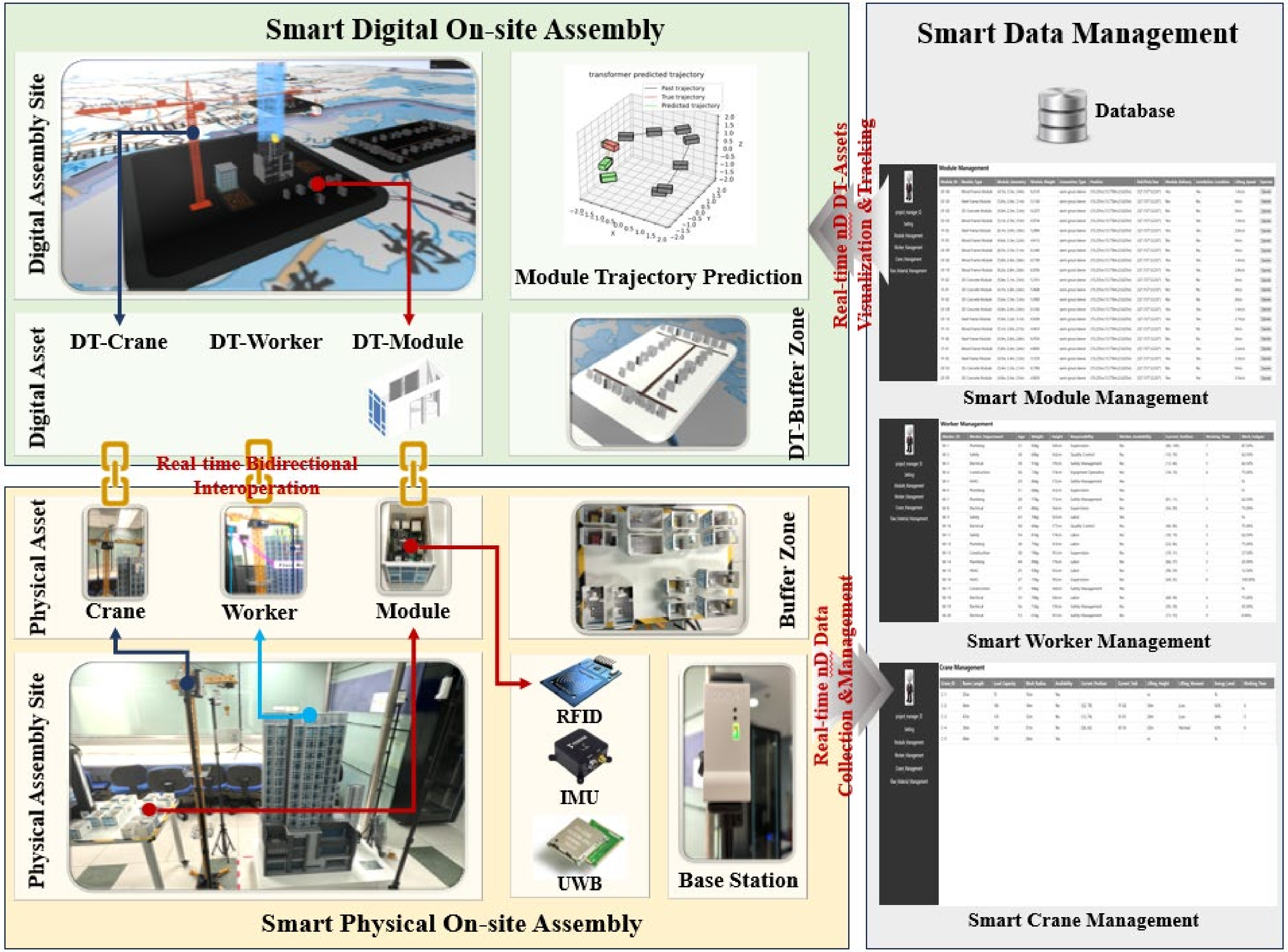
The predicted trajectory of module i :

$$\hat{Y}_i = \left(\hat{x}_i^{t+t_{\text{pred}}}, \hat{y}_i^{t+t_{\text{pred}}}, \hat{z}_i^{t+t_{\text{pred}}}, \hat{\varphi}_i^{t+t_{\text{pred}}}, \hat{\theta}_i^{t+t_{\text{pred}}}, \hat{\psi}_i^{t+t_{\text{pred}}}\right)$$

The loss function :

$$\text{SmoothL1} = \frac{1}{n}\sum_{i=1}^{n}\begin{cases} 0.5 * \left(p_{i_p}^t - p_{i_{GT}}^t\right)^2, & \text{if } \left|p_{i_p}^t - p_{i_{GT}}^t\right| < 1 \\ \left|p_{i_p}^t - p_{i_{GT}}^t\right| - 0.5, & \text{otherxise} \end{cases}$$

Experiments



Evaluation metrics

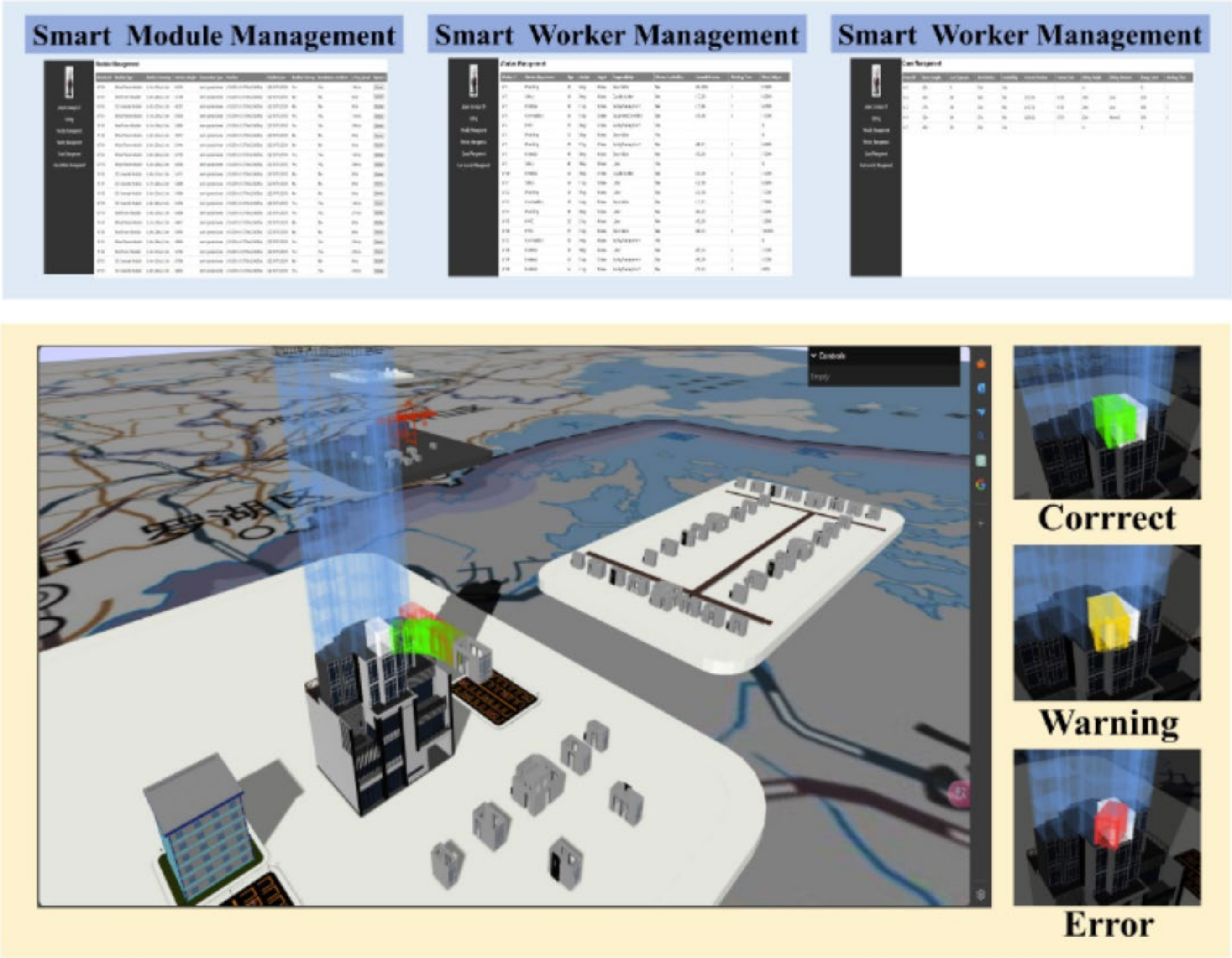
1) Final Displacement Error (FDE) :
This is the distance between the last point of the predicted trajectory and the last point of the real trajectory.

$$FDE = \frac{1}{n}\sum_{i=1}^{n}\left|p_{i_p}^{tf} - p_{i_{GT}}^{tf}\right|$$

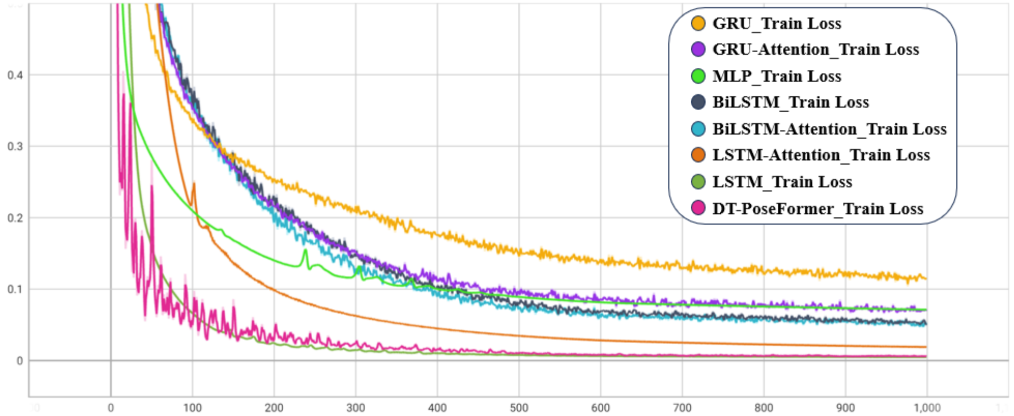
2) Average Displacement Error (ADE) :
This is the average distance between all points of the predicted trajectory and the corresponding points of the real trajectory.

$$ADE = \frac{1}{n}\sum_{i=1}^{n}\frac{1}{t_f}\sum_{t=1}^{t_f}\left|p_{i_p}^{t} - p_{i_{GT}}^{t}\right|$$

Results



The digital platform of MiC stacking process.



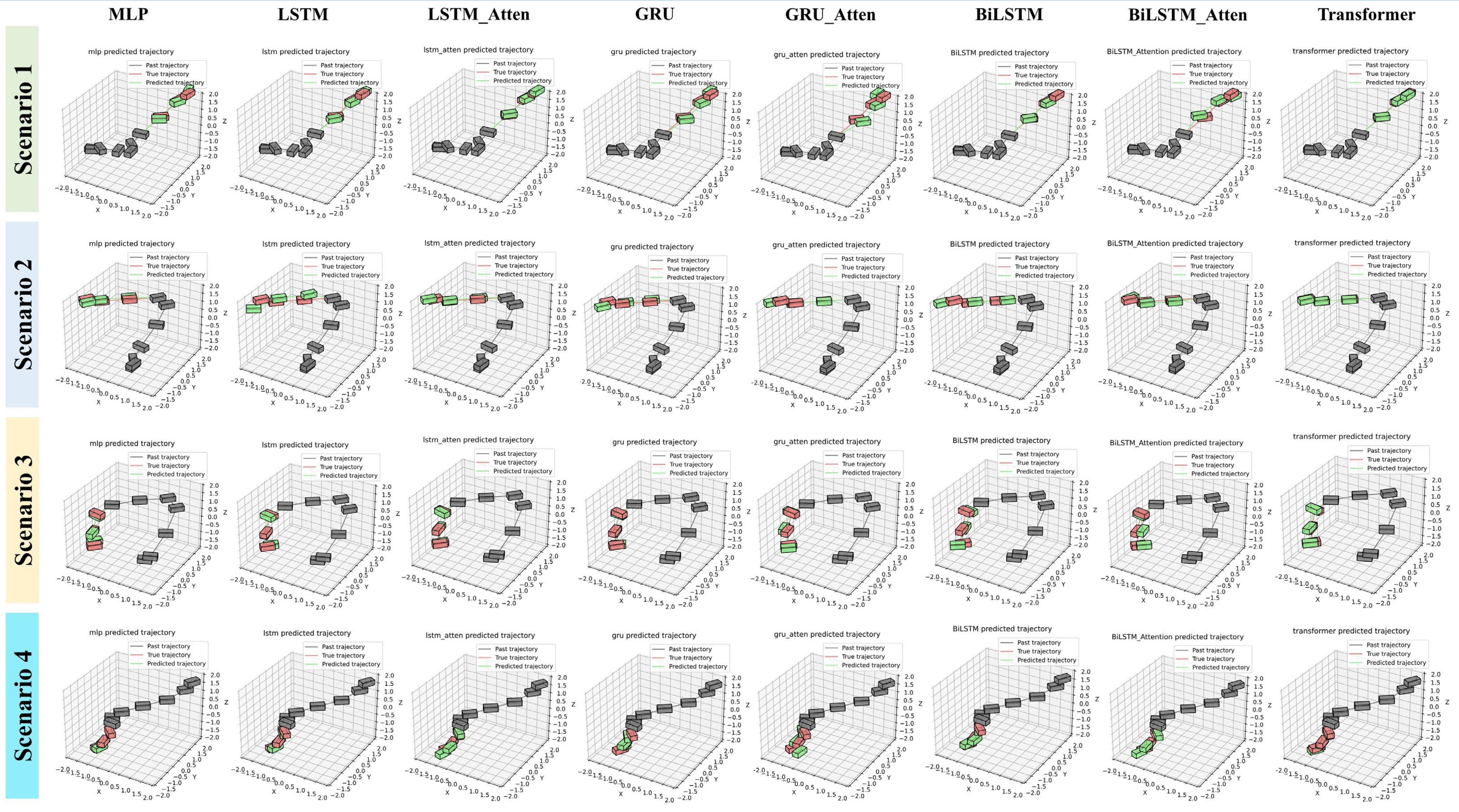
The loss of different methods.

TABLE I. COMPARIONS ON EVALUATION METRICS AND LOSS

Model	Smooth L1 loss	ADE	FDE
MLP	0.07129	57.32236	316.55729
GRU	0.1141	26.19518	137.12731
GRU-Attention	0.07162	26.17367	138.24061
LSTM	**0.004935**	26.30367	139.94350
BiLSTM	0.052	26.01963	139.14750
BiLSTM-Attention	0.05083	26.02726	139.12281
LSTM-Attention	0.01908	26.01323	138.46538
Ours(PoseFormer)	**0.00661**	**23.58859**	**136.27763**

The evaluation matrices of different methods.

Results



The visualization of pose estimation and trajectory prediction in different scenarios.

Yujie Han, Zhiheng Zhao, George Q. Huang

5/31 2024