

# DT-PoseFormer: A Digital Twin-enabled Transformer Network for Precise Pose Estimation and Trajectory Prediction of MiC Modules

Yujie Han and Zhiheng Zhao

IPIIC 2024

10th International  
Physical Internet Conference

May 29-31, 2024  
Savannah, GA USA

THE HONG KONG  
POLYTECHNIC UNIVERSITY  
香港理工大学

## ABSTRACT

Modular integrated construction (MiC) is a game-changing disruptively-innovative approach, which improves quality, simplifies management and reduces construction time due to its off-site production and on-site installation characteristics. However, as one of the most sophisticated processes of MiC, on-site assembly still faces: 1).insufficient interoperability; 2). low-efficient module alignment; 3). highly susceptible safety incidents. To address these challenges, our contributions are:

- A **digital twin (DT)**-based real-time module tracking framework for real-time monitoring and control of MiC assembly to fulfill sufficient interoperability.
- A **transformer-enabled** (DT-PosFormer) network for real-time spatial-temporal module data analysis to automatic pose estimation.
- A **trajectory prediction** of MiC assembly alignment to prevent potential structural collision and misaligned stacking.

## INTRODUCTION

The workflow of MiC mainly consists of three stages: production, logistics, and on-site installation, as shown in Fig. 1. In the production stage, modules are manufactured in the factory based on lean manufacturing principles, advanced manufacturing technologies and automated equipment to ensure quality, precision, and standardization of the modules. Effective collaboration and communication among different departments within the factory, as well as with suppliers and clients, are conducted to provide timely feedback on production status and coordinate logistics and delivery plans. In module logistics stage. dynamic transport planning are conducted based on real-time traffic condition, on-site assembly progress, and module arrival times to reduce storage space requirements and facilitate assembly sequence planning. Finally, modules in the installation queue are stacked by the crane to the target position. The complexity and uncertainty of the hoisting process seriously affects the smoothness and accuracy of the stacking.

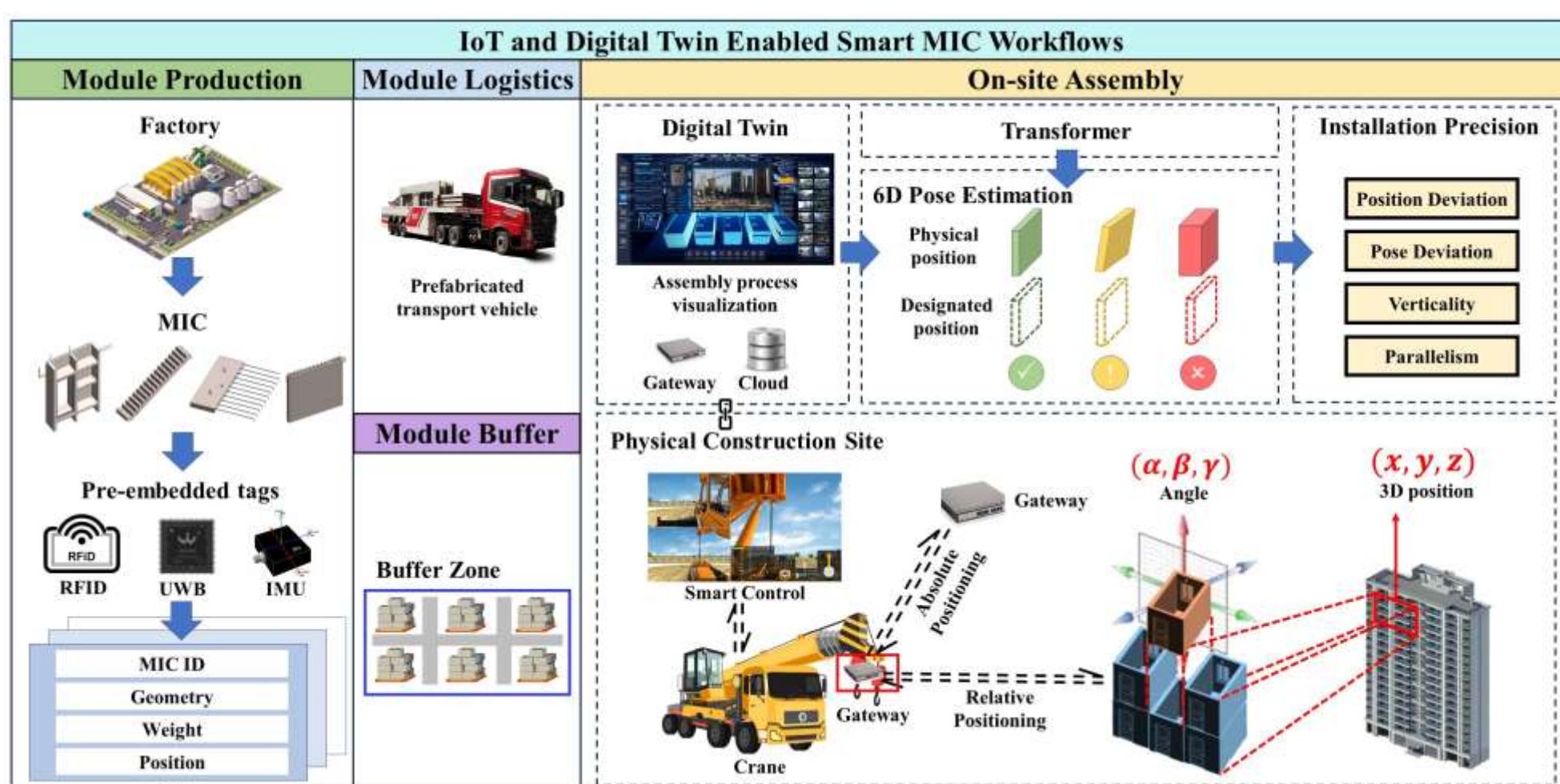


Fig. 1: The workflow of the MiC.

## ACKNOWLEDGMENT

Natural Science Foundation of China (No. 52305557), China Postdoctoral Science Foundation (Grant No. 2022M712394), Hong Kong RGC TRS Project(T32-707/22-N), Research Impact Fund (R7036-22).

## METHODOLOGY

The framework illustrates the overall structure of the DT-based pose estimation and trajectory prediction system, which includes physical layer, digital layer and service layer, as shown in Fig. 2.

### Physical Layer

Dynamic construction assets including on-site workers, crane and modules.IoT devices (UWB, IMU) for data collection, transmission.

### Virtual Layer

DT for visualization and data analysis.Static modeling includes terrain modeling, fixed facility modeling, and asset intrinsic feature modeling.

Dynamic modeling includes dynamic environment modeling, asset dynamic sensing, interaction and emergency response.

### Service Layer

Pose estimation uses UWB for 3D position and IMU for 3D attitude of modules. Trajectory prediction is based on transformer for real-time data processing and feature extraction.

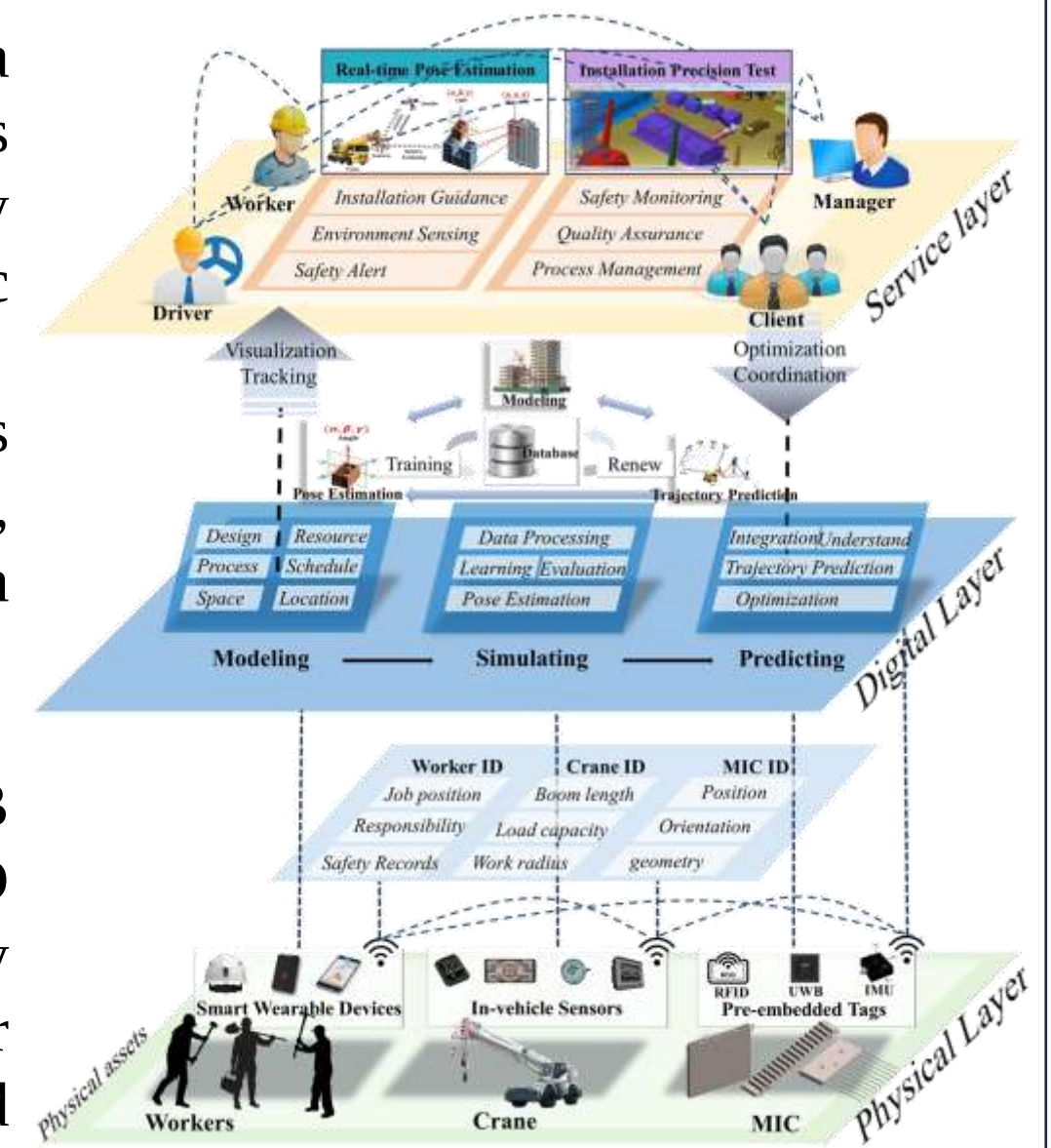


Fig. 2: The framework of DT-Poseformer.

## EXPERIMENTS

### a. Quantitative experiment

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

| TABLE I. COMPARISONS 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                                                | <b>0.004935</b> | 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)                                    | <b>0.00661</b>  | <b>23.58859</b> | <b>136.27763</b> |

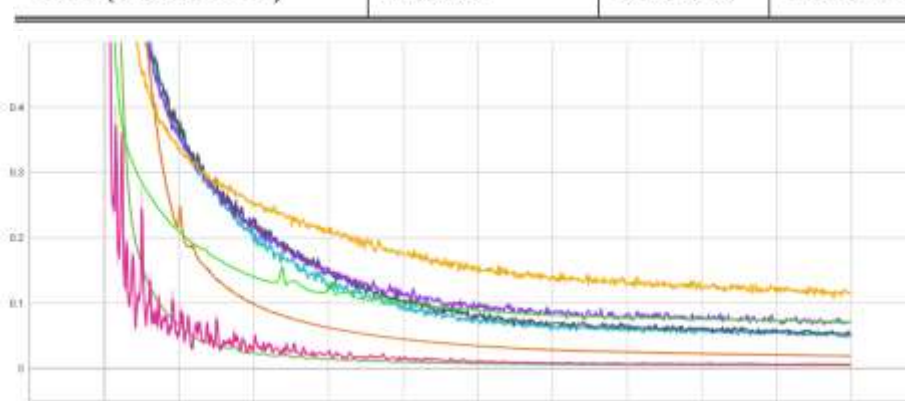


Fig. 3: The loss of different models.

### b. Visualization experiment

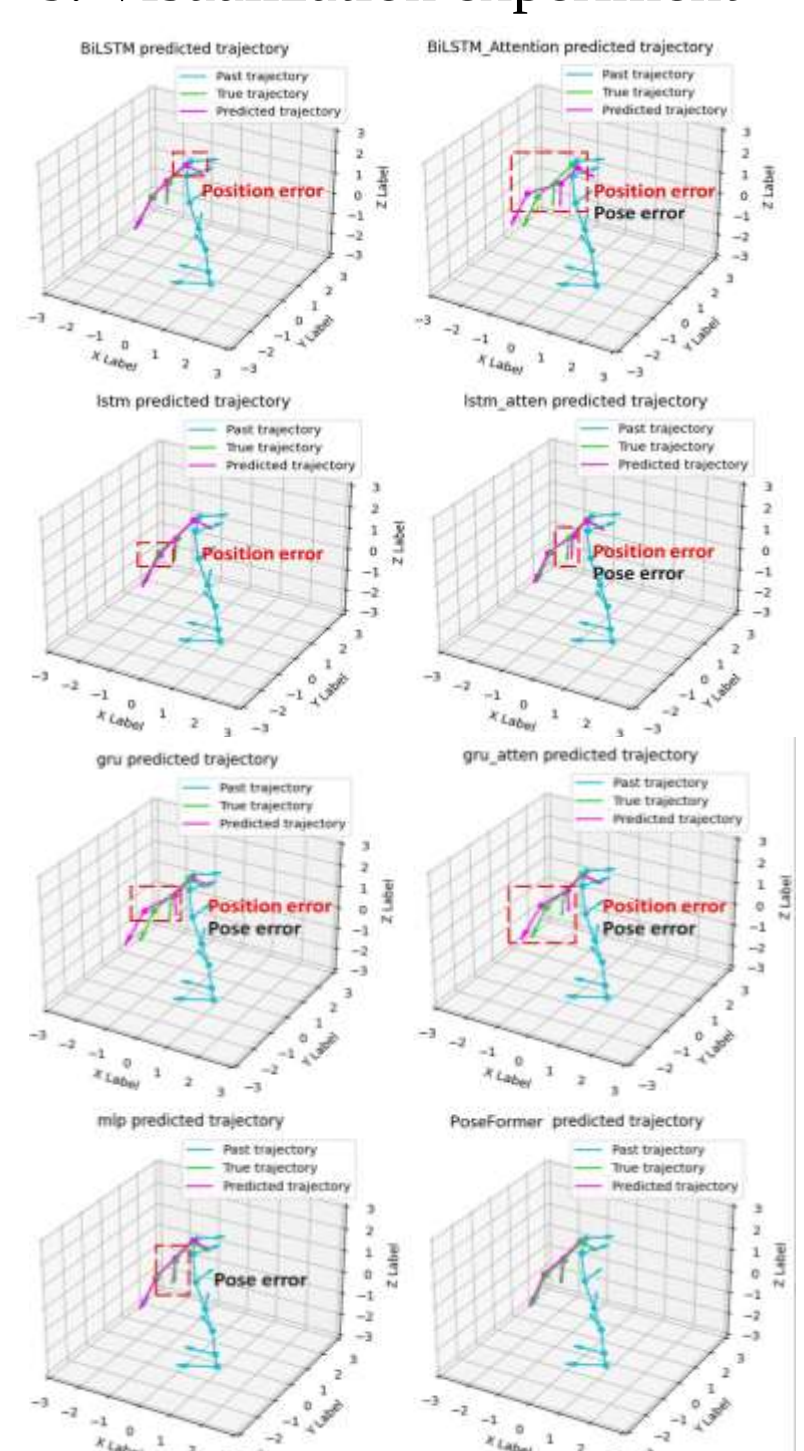


Fig. 4: The predicted trajectory.

## CONCLUSION

We propose a framework for module installation process services based on digital twin and IoT. On one hand, real-time pose estimation information of modules is obtained through UWB and IMU sensors. On the other hand, a deep neural network based on transformer, named DT-PosFormer, is employed for trajectory prediction during the module installation process.

IPIIC 2024

10th International Physical Internet Conference  
Empowering Resilient and Sustainable Supply Chain Ecosystems  
May 29-31, 2024 | The DeSoto Hotel | Savannah, GA USA