

Digital Twin-Driven Semi-Supervised Knowledge Distillation for HMC Monitoring on Edge Device

Md Mahinur Alam, Kanita Jerin Tanha, and Taesoo Jun

Pervasive and Intelligent Computing Lab, Department of IT Convergence Engineering,

Kumoh National Institute of Technology, Gumi 39177, South Korea

(mahinuralam213, kanitajerin17, and taesoo.jun)@kumoh.ac.kr

Abstract—Industry 5.0 envisions harmonious collaboration between humans and machines to elevate industrial productivity, safety, and efficiency. However, achieving real-time monitoring and analysis of human-machine collaborations poses significant challenges, particularly in resource-constrained environments. Key obstacles include ensuring accurate safety monitoring, maintaining scalability, and generating robust training data for dynamic, ever-evolving industrial scenarios. To address these challenges, this paper introduces an innovative Digital Twin-Driven Semi-Supervised Knowledge Distillation (SSKD) framework. By leveraging a Digital Twin of the physical industrial environment using Unreal Engine 4, with ROS2-based communication facilitating collaboration between the virtual and physical systems, the framework creates synthetic data to support deep learning and semi-supervised pseudo-labeling using teacher model and a lightweight student model, optimized through knowledge distillation and quantization, has been deployed on custom-built edge device, achieving a 6.62% increase in accuracy while improving inference speed by 46.04%, ensuring efficient real-time performance. Extensive evaluation across state-of-the-art architectures highlights the framework's superior accuracy, scalability, and operational reliability, offering a transformative solution for monitoring and enhancing Human-Machine Collaboration (HMC) in Industries.

Index Terms—Digital Twin, edge device, human-machine collaboration, semi-supervised learning, knowledge distillation, quantization.

I. INTRODUCTION

In the era of Industry 5.0, collaborative machines are revolutionizing smart manufacturing by enhancing productivity, and efficiency while reducing the physical burden on humans [1]. Human-machine collaboration is at the heart of this transformation, where humans and machines operate together in shared environments. However, ensuring human safety remains a significant challenge in such a dynamic industrial environment [2]. Recent research has shifted toward more adaptive and flexible safety solutions, such as collision avoidance systems that enable machines to alter their actions in real-time to prevent collisions with humans [3]. While promising, these methods often struggle to reliably distinguish between human activities, introducing potential risks. Moreover, they rely heavily on aligning digital models of robots with physical observations, which is often complicated by the need for precise hand-eye calibration [4].

Digital Twin technology has emerged as a groundbreaking innovation, providing real-time virtual replicas of physical systems to simulate, analyze, and optimize industrial processes

[5], [6]. By enabling precise modeling of Human-Machine Collaborations (HMCs), Digital Twins eliminate the need for complex calibration and facilitate adaptable, cage-free safety strategies. Leveraging semi-supervised learning [7], [8], digital twins generate synthetic data that cannot be replicated in the real world, enhance scene understanding, and improve the decision-making capabilities of deep learning models in collaborative systems [9]. The integration of knowledge distillation [10] further strengthens these systems by enabling the deployment of lightweight models on edge devices. This approach ensures efficient real-time performance, maintains high accuracy, and enhances the system's ability to meet modern manufacturing environments' dynamic and evolving demands. This paper addresses the industrial HMC monitoring issue utilizing a digital twins-based SSKD framework. The main contributions of this work are as follows:

- We proposed a novel Digital Twin-Driven SSKD framework tailored for monitoring HMC in Industry 5.0 specifically designed for deployment on resource-constrained edge devices, ensuring efficiency and scalability in real-world industrial settings.
- Developed a Digital Twin of an industrial HMC environment using Unreal Engine 4 to generate synthetic data to train deep learning models and facilitate semi-supervised pseudo-label generation using a teacher model. This approach enhances the monitoring and analysis of HMC.
- A lightweight student model has been developed by integrating knowledge distillation and quantization techniques. This model has been successfully deployed on a custom-built edge device, enabling real-time inference and rigorous model evaluation to ensure operational feasibility.
- The proposed framework is thoroughly tested and evaluated using multiple state-of-the-art model architectures. Comparative analysis with existing methodologies further demonstrated the superior performance and capability of the framework across various configurations.

The remaining part of this paper is organized as follows: **Section II** presents an overview of the related works on Digital Twin for HRC and SSKD for object detection. **Section III** describes the SSKD framework in detail. **Section IV** describes the experimental evaluation and performance analysis. **Section V** summarizes the results and discusses future work.

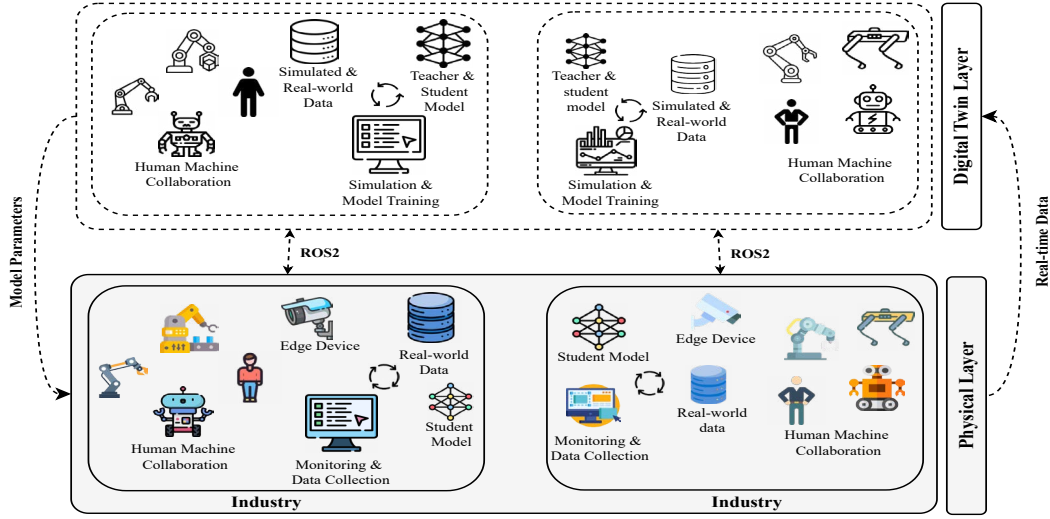


Fig. 1: The Overall Process of the Proposed Digital Twin-Driven SSKD Framework.

II. RELATED WORK

A. Digital Twin for HMC

Digital Twin enhances HMC by enabling real-time modeling and simulation of interactions in industrial settings. Digital Twins enhance decision-making, safety, and adaptability in dynamic environments by integrating synthetic data with real-world observations. Recent advancements in AI and sensing technologies have made Digital Twins an essential tool for optimizing HMC [11], ensuring efficiency and safety in Industry 5.0 applications. Malik and Brem [2] propose a framework for industrial assembly systems that uses labeled and unlabeled data to enhance safety in HRC environments by focusing on adaptability in dynamic human-machine interactions. Droder et al. [4] utilize a Digital Twin integrated with machine learning to validate deep learning models for path planning, addressing safety in collaborative scenarios. Additionally, Ghasemi et al. [12] incorporate deep learning into Augmented Reality to map virtual and physical objects during collaborations, while Park et al. [13] develop a hands-free collaboration system using a Digital Twin in mixed-reality environments.

Unlike previous studies, our Digital Twin generates synthetic datasets for training deep learning models and enables testing in robust scenarios, significantly reducing time and labor costs while improving training efficiency.

B. SSKD for Object Detection

Region-based convolutional neural networks, such as R-CNN, Fast R-CNN, and Faster R-CNN, are widely used two-stage object detection methods that rely on fully supervised learning, requiring large amounts of labeled data for training [14]. In contrast, semi-supervised approaches combine labeled and unlabeled data to reduce this dependency. Pseudo-labeling methods, like the teacher-student model, train a teacher model to generate pseudo-labels [15], which are then used with

unlabeled data to train the student model [16]. Techniques like FixMatch [17] use weak augmentations for pseudo-label creation and validate them with strong augmentations, ensuring consistency and efficiency. Similarly, soft teacher models [18], updated through exponential moving averages, refine pseudo-labeling and improve learning outcomes.

The previous approaches often rely on extensive data pre-processing and assume labeled and unlabeled data come from the same domain, leading to reduced accuracy in unseen environments. Our approach addresses these limitations by combining data from both physical and simulated environments for efficient teacher model training and pseudo-label generation. Through knowledge distillation and quantization, we develop a lightweight student model for edge deployment, ensuring real-time performance and adaptability while reducing dependency on labeled data.

III. PROPOSED FRAMEWORK OVERVIEW

Fig. 1 illustrates the semi-supervised knowledge distillation digital twin framework, which brings together a *Digital Twin layer* and a *Physical layer* to facilitate robust human-machine collaboration detection. By combining synthetic data generated in a virtual environment with real-world data gathered from actual industrial operations.

A. Digital Twin Layer

The Digital Twin layer accurately simulates robotic operations and human involvement in a virtual environment. A teacher-student model pair is initially trained within this layer by leveraging synthetic datasets, which can be further augmented with any available labeled real-world samples. Such simulations allow for cost-effective, rapid data generation, including edge-case scenarios or new industrial layouts that may not be readily available in real life. Consequently, the need for extensive manual labeling is greatly diminished, and iterative experimentation becomes more feasible.

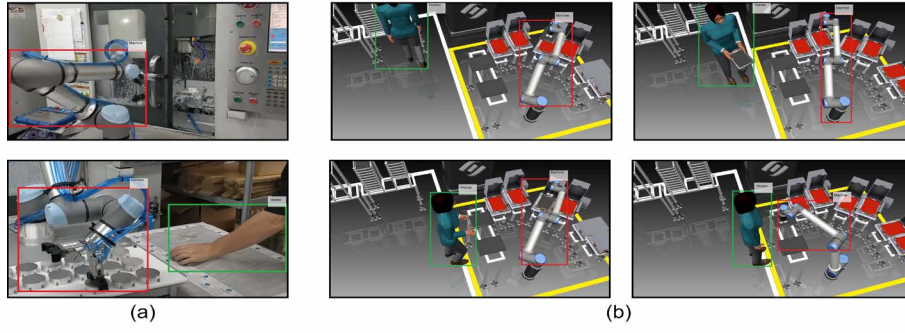


Fig. 2: Demonstration of HMC Environment & Dataset: (a) Cobot and Human Collaboration (b) Our Digital Twin Simulation of the Real-world Industrial Setting.

Within this layer, robotics processes and human-machine collaborations are closely replicated, enabling the teacher model to learn high-level patterns from a broad spectrum of simulated tasks. This teacher model then transfers its learned knowledge to a student model, accelerating the learning process while maintaining performance. ROS2 communication ensures smooth bidirectional data flows between the Digital Twin and the Physical layer, as virtual environments can request ground-truth updates or partially labeled samples from the real world for enhanced training fidelity.

B. Physical Layer

Operating in parallel, the Physical layer features real-world autonomous robots, edge devices, and on-site monitoring systems that capture live sensor data (e.g., video streams). A lightweight student model, informed by the teacher-student training loop in the Digital Twin, is deployed on these devices to detect and respond to human-machine collaborations in real time. This setup allows the model to accommodate practical challenges, including variations in lighting, object occlusions, and other unstructured conditions that are difficult to replicate even in advanced simulations.

As the Physical layer continuously streams new data, it can relay partially labeled or unlabeled observations back to the Digital Twin via ROS2, where the teacher-student model benefits from real-world updates. In turn, simulated outputs and refined model parameters from the Digital Twin are periodically pushed back to edge devices, ensuring that the student model stays current with evolving industrial workflows. This cyclical exchange of data and model updates bolsters safety, reduces labeling overhead through semi-supervised learning, and fine-tunes system performance under resource constraints. By iterating this loop where synthetic data augments real-world data and real-world feedback guides simulation improvements the framework evolves to address both expected and unforeseen collaboration scenarios, ultimately enhancing the efficiency and reliability of human-machine collaborations.

C. Digital Twin Simulation & Dataset

Unreal Engine 4 (UE4) provides a highly versatile platform for creating immersive, photorealistic virtual environments. We leverage UE4 to simulate real-world industrial settings

with minimal reality gaps, generating synthetic data that includes precise annotation information. This data is central to training the SSKD model and validating its performance in identifying human and machine work areas, ensuring adherence to safety standards.

Fig. 2 contrasts the real-world and simulated HMCs. The physical environment including workspace geometry, robot kinematics, and visual data were first captured and synthesized into a 3D model. This model was then imported into UE4, where the cobot and essential workspace elements were positioned to mirror the actual layout. ROS2 handled real-time synchronization, continuously transmitting joint angles, end-effector coordinates, and camera feeds from the physical robot to its virtual counterpart. A virtual human operator was also introduced to replicate the human participant’s actions. We have leveraged this setup to develop a dataset that accurately captures real-world scenarios.

D. Semi-Supervised Knowledge Distillation

Fig. 3 illustrates the proposed semi-supervised knowledge distillation method, which combines synthetic data produced by a Digital Twin with real-world observations. The pipeline begins by training a *teacher model* T on real-world and synthetic data ($\mathcal{D}_{rs}$), thereby leveraging automatically generated annotations that capture diverse robotic scenarios. Formally, the teacher training can be expressed as

$$T = \text{train}(\mathcal{D}_{rs}), \quad (2)$$

and once trained, T predicts *pseudo-labels* on unlabeled synthetic samples

$$\hat{y} = T(x), \quad x \in \{\mathcal{D}'_{\text{sim}} \cup \mathcal{D}_u\}, \quad (3)$$

where $\mathcal{D}'_{\text{sim}}$ is unlabeled simulated data and $\mathcal{D}_u$ is real-world unlabeled data. Pseudo-labels below a chosen confidence threshold can be discarded to maintain label quality. A *student model* S , designed for efficient edge deployment, is then initialized and trained on a mix of labeled and pseudo-labeled data:

$$S = \text{train}(x, \hat{y}) \quad (4)$$

This training incorporates *response-based knowledge distillation*, wherein the teacher’s output probabilities (“soft labels”)

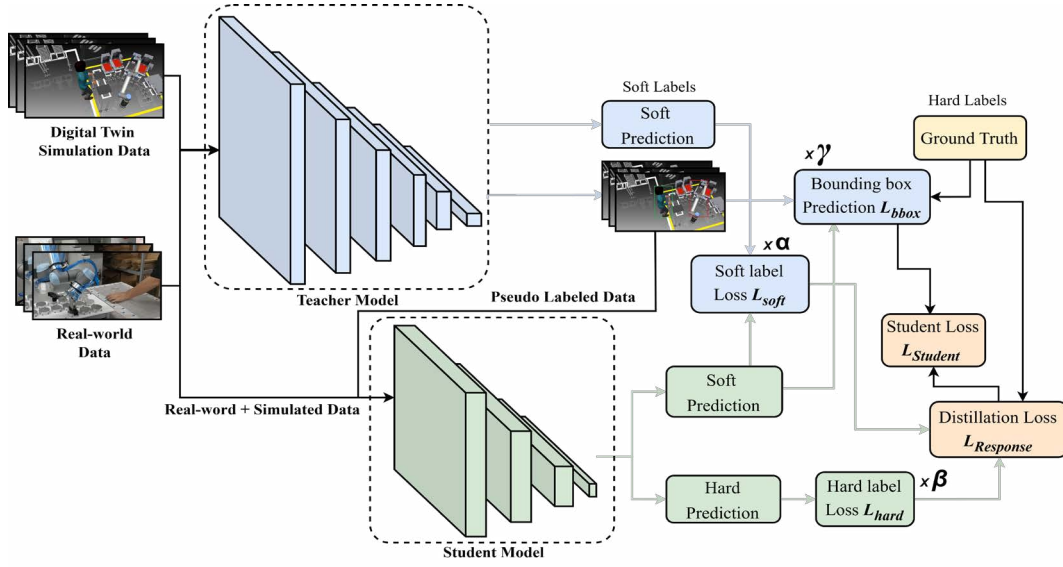


Fig. 3: Teacher and Student Model Training Scheme on the Proposed Semi-Supervised Knowledge Distillation Method.

TABLE I: Experimental Results of the Proposed SSKD Model with Quantization

Model	mAP (%)	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)	IoU (%)	Parameters (m)	FLOPs	Inference Time (ms)
Baseline (YOLOv5s)	79.45	88.35	84.78	76.88	80.51	77.20	7.2	16.5	129.34
Baseline + SS (Pseudo label)	85.98	90.48	85.29	78.97	81.91	78.89	7.2	16.5	130.74
Baseline + KD (from YOLOv5l)	89.76	91.10	87.38	81.75	84.30	80.88	7.2	16.5	121.54
Baseline + SSKD	91.98	95.39	88.87	84.46	85.97	87.45	7.2	16.5	125.12
Baseline + SSKD & Quantization	91.11	94.20	86.98	85.40	83.10	84.12	4.8	10.3	67.51

*SS: Semi-Supervised Learning, *KD: Knowledge Distillation, *SSKD: Semi-Supervised knowledge Distillation

provide a more informative training signal than ground-truth labels alone. In this approach, soft labels are computed by applying a temperature-scaled softmax to the teacher's logits. Here, the normalized probability i is

$$q_i = \frac{\exp(z_i/T)}{\sum_j \exp(z_j/T)} \quad (10)$$

where z_i denotes the teacher's raw logits, and T is the distillation temperature. A larger T yields a softer distribution, thereby facilitating the student's understanding of inter-class relationships. During distillation, the total loss combines *soft-label loss* and *hard-label loss*:

$$L_{\text{Response}} = \alpha L_{\text{soft}} + \beta L_{\text{hard}} \quad (11)$$

Here, L_{soft} is a measure of how well the student's temperature-scaled predictions match the teacher's temperature-scaled predictions; L_{hard} enforces consistency with ground-truth (hard) labels. Specifically, the soft-label loss takes the form of a cross-entropy between teacher and student outputs at temperature T :

$$L_{\text{soft}} = - \sum_{j=1}^N p_j^T \log(q_j^T) \quad (12)$$

where p_j^T is the teacher's probability j at temperature T , and q_j^T is the student's corresponding probability under the same scaling. The hard-label loss compares the student's predictions to ground-truth labels $c_j \in \{0, 1\}$:

$$L_{\text{hard}} = - \sum_{j=1}^N c_j \log(q_j^1) \quad (13)$$

hence q_j^1 represents the student's predicted probability at temperature 1. For classification tasks, c_j is 1 if the instance belongs to the target class and 0 otherwise. For the HMC detection, the model predicts bounding boxes $\hat{b}_i$ for each instance i . For that an additional bounding-box regression loss L_{bbox} , is implemented via Smooth L_1 or ℓ_1 distance between the predicted $\hat{b}_i$ and ground-truth b_i :

$$L_{\text{bbox}} = \sum_{i=1}^N \text{Smooth}_{L_1}(\hat{b}_i, b_i) \quad (14)$$

where each $b_i = (x_i, y_i, w_i, h_i)$ is the ground-truth box, and $\hat{b}_i$ is the student's predicted box i . The distillation loss accounts for both classification (soft + hard labels) and bounding-box regression:

$$\mathcal{L}_{\text{student}} = \alpha L_{\text{soft}} + \beta L_{\text{hard}} + \gamma L_{\text{bbox}} \quad (15)$$

Here, α , β , and γ serve as weighting factors that balance distillation signals (soft labels), supervised classification (hard labels), and bounding-box regression. By tuning these coefficients, one can prioritize the different learning signals according to application demands. When $\gamma = 0$, the student focuses solely on classification tasks; when γ is larger, bounding-box accuracy takes precedence.



Fig. 4: Custom Build Lightweight Edge Device for Edge Deployment and HMC Monitoring.

TABLE II: Performance Comparison of the Proposed Method with Existing Approaches

Teacher	ViT-Base(85.55M)		WRN40-4(8.97M)		ResNet50(23.73M)	
Student	ViT-Tiny(4.24M)		MobileNetV2(1.24M)		ResNet18(0.5M)	
Method	Accuracy (%)	mAP (%)	Accuracy (%)	mAP (%)	Accuracy (%)	mAP (%)
Chen et al. [19]	90.15	88.21	85.38	84.75	83.95	81.91
Cao et al. [10]	87.13	82.37	82.47	81.10	81.26	80.54
Cui et al. [20]	88.98	87.10	86.29	84.98	85.19	83.95
Zhao et al. [7]	91.19	88.21	88.13	85.22	86.19	83.91
SSKD	91.58	88.23	87.94	85.67	85.87	84.10

Our SSKD method allows the student model to benefit from teacher-provided soft labels capturing nuanced class relationships while still adhering to essential supervised signals in classification and localization. As a result, the student model can effectively handle sophisticated vision tasks in HMC scenarios, even under the resource constraints typical of edge devices. The Digital Twin supplies labeled synthetic data and generates unlabeled simulated samples that the teacher model can pseudo-label, thus augmenting the student’s training set without intensive human annotation. As the final step, the student model S is validated on a small portion of fully labeled real-world data to ensure robust generalization.

E. Quantization of the Student Model

post-training integer quantization is being applied to the final student model S . Specifically, all floating-point weights and activations are converted into 8-bit integers, reducing memory usage by $\approx 30\%$ without significantly impacting accuracy. During inference, integer-arithmetic kernels replace their floating-point counterparts, allowing the compressed student model $\hat{S}$ to run efficiently on our resource-constrained embedded controllers. Our testing showed that this quantized model maintains real-time performance for HMC, even under limited compute and memory budgets.

IV. EXPERIMENTAL EVALUATION

A. Experimental Environment

The study used a server with an Intel Core i9-10980XE (3 GHz), three NVIDIA GeForce RTX 3090 GPUs, 130 GB

RAM, and Ubuntu 22.04.4 LTS. The deep learning setup used NVIDIA driver 535.161.08 with CUDA 12.2 and Python (2.16.1 and 3.12.2) for TensorFlow for Digital Twin simulation and model training. And a custom AArch64 edge device (shown in Fig. 4) with a 6-core CPU (500 MHz base, 2208 MHz boost) for student model deployment and testing.

B. Evaluation of the proposed framework

A comprehensive evaluation of the proposed method is presented, employing different metrics such as mean Average Precision (mAP), Intersection over Union (IoU), Inference Time, and FLOPs. The enhancements achieved through SS, KD, and quantization in improving the baseline performance of the student model are summarized in Table I. The baseline YOLOv5s detector achieves an mAP of 79.45% with a real-time inference time of 129 ms. Enhancing this baseline with SS improves detection mAP (85.98%) while incorporating KD from a larger YOLOv5l model further improves accuracy (91.10%) and mAP (89.76%). Combining Semi-Supervised Learning and Knowledge Distillation into our proposed SSKD method leads to significant performance gains across all metrics, notably increasing the IoU from 80.88% to 87.45%. Further applying integer quantization to the SSKD model significantly reduces the model parameters, FLOPs, and inference time with a minimalist accuracy drop of 1.19%. The average inference time improvement gain through quantization is shown in Fig. 5 from 125.12 ms to 67.51 ms.

Table II presents a comparative evaluation of the proposed method against existing approaches under different model architectures. Three teacher-student pairs are analyzed: ViT-Base with ViT-Tiny, WRN40-4 with MobileNetV2, and ResNet50 with ResNet18. Several studies, including those by Chen et al. [19], Cao et al. [10], Cui et al. [20], and Zhao et al. [7], have utilized SSKD approaches similar to ours. However, the methods for KD and SS pseudo-label generation vary uniquely across these works. Notably, Zhao et al. [7] employed three KD techniques in combination, achieving slightly higher accuracy and mAP at times, with peak values of 91.19% and 88.21%, respectively. In contrast, our proposed SSKD method consistently delivers superior performance across all configurations, achieving the highest recorded accuracy and mAP of 91.58% and 88.23%, respectively, using the ViT-Base/ViT-Tiny pairing. This underscores the efficacy of the proposed method in enhancing student model performance while leveraging diverse teacher-student architectures. These results illustrate that the quantized SSKD model effectively balances inference speed, memory footprint, and detection accuracy, making it well-suited for industrial human-machine collaboration tasks.

V. CONCLUSION

This paper introduces a cutting-edge Digital Twin-Driven Semi-Supervised Knowledge Distillation framework to address the challenges of monitoring human-machine collaborations in Industry 5.0. The framework ensures efficient real-time performance on resource-constrained edge devices by combining

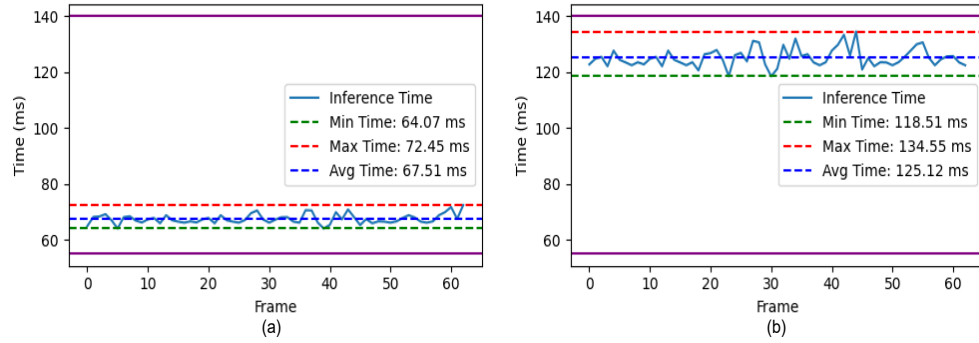


Fig. 5: Average Inference Time (a) with Quantization and (b) without Quantization on the Edge Device.

synthetic data generated via UE4-based Digital Twins and ROS2 communication with semi-supervised pseudo-labeling and knowledge distillation. The lightweight student model, enhanced through quantization, achieves a remarkable accuracy of 94.20% and a significantly low inference latency of 67.51 ms from 125.12 ms, making it highly suitable for industrial deployment. Extensive evaluations highlight the framework's superior accuracy, scalability, and operational efficiency compared to state-of-the-art methods. This study lays the foundation for integrating advanced simulation and AI-driven solutions to enhance industrial safety and collaboration. Future work will focus on extending the framework to accommodate more complex and dynamic human-machine collaboration scenarios, integrating multi-modal sensor inputs, and exploring federated learning to improve model adaptability and data privacy in distributed environments.

ACKNOWLEDGMENT

This work was supported in part by IITP (MSIT) under the Innovative Human Resource Development for Local Intellectualization (IITP-2025-RS-2020-II201612; 34%) and the ITRC program (IITP-2025-RS-2024-00438430; 33%), and by the NRF Basic Science Research Program (Ministry of Education, 2018R1A6A1A03024003; 33%).

REFERENCES

- [1] A. Baratta, A. Cimino, F. Longo, and L. Nicoletti, "Digital twin for human-robot collaboration enhancement in manufacturing systems: Literature review and direction for future developments," *Computers & Industrial Engineering*, vol. 187, p. 109764, 2024.
- [2] A. A. Malik and A. Brem, "Digital twins for collaborative robots: A case study in human-robot interaction," *Robotics and Computer-Integrated Manufacturing*, vol. 68, p. 102092, 2021.
- [3] J.-H. Chen and K.-T. Song, "Collision-free motion planning for human-robot collaborative safety under cartesian constraint," in *2018 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2018, pp. 4348–4354.
- [4] K. Dröder, P. Bobka, T. Germann, F. Gabriel, and F. Dietrich, "A machine learning-enhanced digital twin approach for human-robot-collaboration," *Procedia Cirp*, vol. 76, pp. 187–192, 2018.
- [5] M. Krupas, E. Kajati, C. Liu, and I. Zolotova, "Towards a human-centric digital twin for human-machine collaboration: A review on enabling technologies and methods," *Sensors*, vol. 24, no. 7, p. 2232, 2024.
- [6] M. M. Alam, M. Golam, E. A. Tuli, M. R. Subhan, D.-S. Kim, and T. Jun, "Dcfl-chain: Digital-twin-based collaborative fl-integrated energy consumption prediction for smart factory," , pp. 310–311, 2024.
- [7] Z. Zhao, J. Lyu, Y. Chu, K. Liu, D. Cao, C. Wu, L. Qin, and S. Qin, "Toward generalizable robot vision guidance in real-world operational manufacturing factories: A semi-supervised knowledge distillation approach," *Robotics and Computer-Integrated Manufacturing*, vol. 86, p. 102639, 2024.
- [8] M. M. Alam, G. Mohtasin, M. R. Subhan, D.-S. Kim, and T. Jun, "Federated semi-supervised digital twin for enhanced human-machine interaction in industry 5.0," in *2024 15th International Conference on Information and Communication Technology Convergence (ICTC)*. IEEE, 2024, pp. 1270–1275.
- [9] M. R. Subhan, M. Golam, M. M. Alam, D.-S. Kim, and T. Jun, "Energy-twin: Explainable ai-driven energy optimization using digital-twin for smart industry," , pp. 275–276, 2024.
- [10] Y. Cao, Y. Song, X. Xu, S. Li, Y. Yu, Y. Zhang, and W. Shen, "Semi-supervised knowledge distillation for tiny defect detection," in *2022 IEEE 25th International Conference on Computer Supported Cooperative Work in Design (CSCWD)*. IEEE, 2022, pp. 1010–1015.
- [11] J. Fan, P. Zheng, and C. K. Lee, "A vision-based human digital twin modeling approach for adaptive human-robot collaboration," *Journal of Manufacturing Science and Engineering*, vol. 145, no. 12, p. 121002, 2023.
- [12] Y. Ghasemi, H. Jeong, S. H. Choi, K.-B. Park, and J. Y. Lee, "Deep learning-based object detection in augmented reality: A systematic review," *Computers in Industry*, vol. 139, p. 103661, 2022.
- [13] K.-B. Park, S. H. Choi, J. Y. Lee, Y. Ghasemi, M. Mohammed, and H. Jeong, "Hands-free human-robot interaction using multimodal gestures and deep learning in wearable mixed reality," *IEEE Access*, vol. 9, pp. 55 448–55 464, 2021.
- [14] R. Girshick, J. Donahue, T. Darrell, and J. Malik, "Rich feature hierarchies for accurate object detection and semantic segmentation," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, 2014, pp. 580–587.
- [15] S. Manivannan, "Collaborative deep semi-supervised learning with knowledge distillation for surface defect classification," *Computers & Industrial Engineering*, vol. 186, p. 109766, 2023.
- [16] K. Sohn, Z. Zhang, C.-L. Li, H. Zhang, C.-Y. Lee, and T. Pfister, "A simple semi-supervised learning framework for object detection," *arXiv preprint arXiv:2005.04757*, 2020.
- [17] K. Sohn, D. Berthelot, N. Carlini, Z. Zhang, H. Zhang, C. A. Raffel, E. D. Cubuk, A. Kurakin, and C.-L. Li, "Fixmatch: Simplifying semi-supervised learning with consistency and confidence," *Advances in neural information processing systems*, vol. 33, pp. 596–608, 2020.
- [18] M. Xu, Z. Zhang, H. Hu, J. Wang, L. Wang, F. Wei, X. Bai, and Z. Liu, "End-to-end semi-supervised object detection with soft teacher," in *Proceedings of the IEEE/CVF international conference on computer vision*, 2021, pp. 3060–3069.
- [19] D. Chen, A. Ma, and Y. Zhong, "Semi-supervised knowledge distillation framework for global-scale urban man-made object remote sensing mapping," *International Journal of Applied Earth Observation and Geoinformation*, vol. 122, p. 103439, 2023.
- [20] X. Cui, C. Wang, D. Ren, Y. Chen, and P. Zhu, "Semi-supervised image deraining using knowledge distillation," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 32, no. 12, pp. 8327–8341, 2022.