Lite-DIO Is Actually What You Need for Efficient Inertial Localization

Extended Abstract

Yan Li
The School of Systems Science and Engineering, Sun Yat-Sen University
Guangzhou, China
liyan377@mail2.sysu.edu.cn

Meng Liu
College of Intelligent Systems Science and Engineering, Harbin Engineering University
Harbin, China
liumeng_heu@163.com

Zhongchen Shi
Defense Innovation Institute, Academy of Military Sciences (AMS)
Beijing, China
shizhongchen@buaa.edu.cn

Yanqing Hou
The School of Systems Science and Engineering, Sun Yat-Sen University
Guangzhou, China
houyq9@mail.sysu.edu.cn

Liang Xie
Defense Innovation Institute, Academy of Military Sciences (AMS)
Beijing, China
xielnudt@gmail.com

Hongbo Chen
The School of Systems Science and Engineering, Sun Yat-Sen University
Guangzhou, China
chenhongbo@mail.sysu.edu.cn

Erwei Yin
Defense Innovation Institute, Academy of Military Sciences (AMS)
Beijing, China
yinerwei1985@gmail.com

ABSTRACT

In this work, we propose a simple and effective framework (*i.e.,* Lite-DIO), marking the first attempt to accelerate deep inertial odometry with knowledge distillation. In Lite-DIO, we first independently construct the Transformer-based teacher model and a lightweight student network. Then, adaptive transferring knowledge is enabled between the teacher model and the student network in a dual-level contrastive distillation manner. With such design, the distilled knowledge comes from not only the teacher model's predictions but also the latent high-order collaborative semantics preserved in embeddings. Extensive experiments conducted on three real-world datasets demonstrate that the proposed Lite-DIO significantly reduces model size and inference time compared to existing popular alternatives, while the compressed model still maintains competitive localization accuracy.

KEYWORDS

Inertial measurement unit; Pedestrian localization; Deep learning; Knowledge distillation; Inertial odometry

ACM Reference Format:
Yan Li, Meng Liu, Zhongchen Shi, Yanqing Hou, Liang Xie, Hongbo Chen, and Erwei Yin. 2025. Lite-DIO Is Actually What You Need for Efficient Inertial Localization: Extended Abstract. In *Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Detroit, Michigan, USA, May 19 – 23, 2025*, IFAAMAS, 3 pages.

Corresponding author : Zhongchen Shi, Erwei Yin.
The work is supported by the National Natural Science Foundation of China (Grant Nos.: 62332019, 62076250).

Proc. of the 24th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2025), Y. Vorobeychik, S. Das, A. Nowé (eds.), May 19 – 23, 2025, Detroit, Michigan, USA. © 2025 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org).

1 INTRODUCTION

Accurate and efficient odometry with low-cost Inertial Measurement Units (IMUs) is an ideal localization solution of a moving subject (*e.g.,* robot and pedestrian), playing a crucial role in robot navigation, augmented reality (AR) [11], pedestrian localization [9] and autonomous driving.

Recent years have witnessed the great success of deep neural networks (DNN) in learning latent representations for temporal data. Inspired by such development, many efforts have introduced DNN into inertial odometry to avoid IMU integration and mitigate positioning drifts and shown its power in modeling high-order nonlinear relationships [1, 4, 6, 7]. Though promising, we notice that several key problems of deep inertial odometry (DIO) are overlooked by previous solutions: 1) Increasing the depth and size of deep neural networks has become a key strategy for enhancing localization accuracy. 2) The vast number of parameters inevitably raises computational overhead and inference time, which in turn can lead to localization delays and tracking loss.

To this end, we propose a dual-level contrastive distillation model compression framework to improve both the effectiveness and efficiency of DIO, named Lite-DIO. In detail, a Transformer-based DIO initially serves as the teacher model, transferring knowledge to a lightweight student model. Then an embedding-invariant representation learning based on a bidirectional contrastive loss is employed to capture higher-order collaborative semantics in the teacher model's embeddings. After that, we harness prediction-invariant representation learning based on teacher-bounded loss to align the predictive outputs between the teacher and student models. With this design, the distilled knowledge is derived not only from the teacher model's predictions but also from the latent higher-order collaborative semantics embedded within the representations. Extensive experiments conducted on two real-world public datasets demonstrate that our method not only matches the performance of existing state-of-the-art solutions but also achieves remarkable improvements in inference speed and model efficiency.

2 METHOD

Teacher and Student Model. Inspired by CTIN [4], we design a Transformer-based DIO initially serves as the teacher model. In addition, we build two simple student models, *i.e.,* a single-layer unidirectional LSTM (SL-LSTM) and a pared-down ResNet (PDResNet).

Prediction-invariant Representation Learning. Conventional knowledge distillation is predominantly designed for classification tasks [5], where the predictions of the teacher network serve as a guide for the student model's learning process. Since the probability outputs of the teacher model can be intuitively interpreted and leveraged, the student model can learn more refined decision boundaries by mimicking these probability distributions.

Unlike distillation for discrete categories, the teacher model's outputs in regression tasks are continuous values, lacking a clear interpretative structure. In order to distill knowledge from the teacher model to the student model in velocity regression task, Lite-DIO first follows the paradigm of MSE-based prediction-invariant clue extraction to align the predictive outputs between the teacher and student models. However, a brute-force alignment of the outputs from the two models does not necessarily result in effective knowledge transfer. The teacher's regression outputs can provide very wrong guidance toward the student model, since the real valued regression outputs are unbounded. Thus, we extend the teacher-bounded regression loss [2] by treating the teacher's predictions as an upper bound, refraining from adding additional loss when the student model outperforms the teacher, thereby encouraging the student to learn more from the guidance of ground truth.

Embedding-invariant Representation Learning. Inspired by recent contrastive learning algorithms [3], we design a bidirectional contrastive loss COS-InfoNCE for the transfer of the intermediate inertial representation layer. It can better mimic the generalization capability of teacher.

In a training batch, IMU measurement windows $\mathcal{T} = \{\mathbf{X}_i\}$, when encoded by both the teacher and student, yield B pairs of inertial representations. For k-th teacher's representation $\mathbf{H}_k^{(t)}$, k-th student's representation is its positive $\mathbf{H}_k^{(s)}$, while other student's representations will be the negatives in this batch. Taking distilled SL-LSTM as an example, we first reshape $\mathbf{H}_i^{(t)} \in \mathbb{R}^{w \times d}$ and $\mathbf{H}_i^{(s)} \in \mathbb{R}^{w \times d_1}$ to $\mathbf{h}_i^{(t)} \in \mathbb{R}^{d_2}, \mathbf{h}_i^{(s)} \in \mathbb{R}^{d_2}$ through an adaptive layer, which is an average pooling operation followed by a linear layer. Then, we use cosine similarity to compute similarity of "teacher" to "student" and "student" to "teacher", respectively. Next, we maximize cross-model contrastive alignment between $\mathbf{h}_i^{(t)}$ and $\mathbf{h}_i^{(s)}$ while minimizing the alignment of non-similar representations.

3 RESULTS

Table 1: Position evaluation (meter) on RIDI dataset. The best distillation results are highlighted in bold. SL-LSTM2 and PDResNet2 are the student versions after Lite-DIO distillation.

Model	Params(M)	ATE	RTE
PDR [8]	–	22.76	24.89
RONIN_ResNet [4]	4.63	2.33	2.36
RONIN_TCN [4]	2.03	3.25	2.64
IMUNet [10]	3.66	2.20	2.48
SL-LSTM	0.05	2.81	3.02
PDResNet	0.23	2.78	2.91
SL-LSTM2	0.05	2.15	**2.53**
PDResNet2	0.23	**2.14**	2.66

Table 1 summarizes the overall performance of Lite-DIO and three baselines. From the results we have the following observations: 1) The deep learning-based inertial odometry outperforms the traditional PDR model. This improvement is likely due to the fact that inaccurate step length and heading estimations in PDR can lead to cumulative drift. 2) the teacher model's localization error showed a significant reduction compared to the baselines. 3) By using our proposed Lite-DIO, the knowledge distilled from the teacher model is adaptively transferred to the student model, significantly improving the localization accuracy of the student model. In particular, SL-LSTM2 and PDResNet2 improve an average ATE on RIDI seen test datasets by 23.49%, 23.02% over SL-LSTM and PDResNet.

The primary limitation of existing baselines lies in their inability to effectively capture the spatio-temporal information and inter-modal correlation dependencies within IMU sequences. Additionally, the extensive number of parameters can result in inference delays and localization latency. Therefore, the Transformer-based teacher model can effectively transfer the spatiotemporal information and modal dependencies it captures to the student network. Concurrently, our proposed Lite-DIO significantly reduces the model's parameters. As shown in Table 1, The localization accuracy of the SL-LSTM2 and PDResNet2 can reach a level comparable to the baselines. Two students not only have the fewest parameters, but their inference time by GPU during testing are also the lowest among all the models. This further demonstrates the effectiveness of our proposed Lite-DIO.

4 CONCLUSION

In this paper, we introduce Lite-DIO, a novel KD method designed to accelerate DIO. Lite-DIO achieves bi-level distillation through embedding-invariant representation learning and prediction-invariant representation learning, ensuring the maximization of adaptive knowledge transfer between the teacher and student models. Our results have shown that Lite-DIO can improve the student DIO efficiency and achieve competitive performance. In the future, we will further extend our method to other relevant fields, *e.g.,* visual-inertial odometry.

REFERENCES

[1] Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. 2018. Ionet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 32.

[2] Guobin Chen, Wongun Choi, Xiang Yu, Tony Han, and Manmohan Chandraker. 2017. Learning efficient object detection models with knowledge distillation. *Advances in neural information processing systems* 30 (2017).

[3] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. 2020. Momentum contrast for unsupervised visual representation learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. 9729–9738.

[4] Sachini Herath, Hang Yan, and Yasutaka Furukawa. 2020. Ronin: Robust neural inertial navigation in the wild: Benchmark, evaluations, & new methods. In *2020 IEEE international conference on robotics and automation (ICRA)*. IEEE, 3146–3152.

[5] Geoffrey Hinton, Oriol Vinyals, and Jeff Dean. 2015. Distilling the knowledge in a neural network. *arXiv preprint arXiv:1503.02531* (2015).

[6] Wenxin Liu, David Caruso, Eddy Ilg, Jing Dong, Anastasios I Mourikis, Kostas Daniilidis, Vijay Kumar, and Jakob Engel. 2020. Tlio: Tight learned inertial odometry. *IEEE Robotics and Automation Letters* 5, 4 (2020), 5653–5660.

[7] Bingbing Rao, Ehsan Kazemi, Yifan Ding, Devu M Shila, Frank M Tucker, and Liqiang Wang. 2022. Ctin: Robust contextual transformer network for inertial navigation. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 36. 5413–5421.

[8] Qinglin Tian, Zoran Salcic, I Kevin, Kai Wang, and Yun Pan. 2015. An enhanced pedestrian dead reckoning approach for pedestrian tracking using smartphones. In *2015 IEEE Tenth International Conference on Intelligent Sensors, Sensor Networks and Information Processing (ISSNIP)*. IEEE, 1–6.

[9] Yingying Wang, Hu Cheng, and Max Q-H Meng. 2022. Spatiotemporal co-attention hybrid neural network for pedestrian localization based on 6D IMU. *IEEE Transactions on Automation Science and Engineering* 20, 1 (2022), 636–648.

[10] Behnam Zeinali, Hadi Zanddizari, and Morris J Chang. 2024. IMUNet: Efficient Regression Architecture for Inertial IMU Navigation and Positioning. *IEEE Transactions on Instrumentation and Measurement* (2024).

[11] Baoding Zhou, Zhining Gu, Wei Ma, and Xu Liu. 2020. Integrated BLE and PDR indoor localization for geo-visualization mobile augmented reality. In *2020 16th International Conference on Control, Automation, Robotics and Vision (ICARCV)*. IEEE, 1347–1353.