



Demo Abstract: Real-Time Tracking of Smartwatch Orientation and Location by Multitask Learning

Miaomiao Liu

mliu71@ucmerced.edu

University of California, Merced
Merced, CA, USA

Wyssanie Chomsin

wchomsin@ucmerced.edu

University of California, Merced
Merced, CA, USA

Sikai Yang

syang126@ucmerced.edu

University of California, Merced
Merced, CA, USA

Wan Du

wdu3@ucmerced.edu

University of California, Merced
Merced, CA, USA

ABSTRACT

In this demo, we present *RTAT*, a real-time arm tracking system that tracks both orientation and location of a smartwatch simultaneously by a multitask learning neural network. We incorporate an attention layer and design a dedicated loss for the multitask neural network to learn the dynamic relationships among Inertial Measurement Unit (IMU) sensors. *RTAT* supports real-time tracking by performing deep learning inference on a smartphone. Finally, to train *RTAT*, we develop an easy-to-use labeled data collection system that uses a low-cost virtual reality system to measure the ground truth orientation and location of the smartwatch. Extensive experiments show *RTAT* outperforms significantly the state-of-the-art solutions in inference accuracy and latency.

CCS CONCEPTS

• Human-centered computing → Ubiquitous and mobile devices;
• Computer systems organization → Real-time systems; • Computing methodologies → Artificial intelligence.

KEYWORDS

Arm tracking, mobile sensing, multitask learning, Inertial Measurement Unit, orientation, location

ACM Reference Format:

Miaomiao Liu, Sikai Yang, Wyssanie Chomsin, and Wan Du. 2022. Demo Abstract: Real-Time Tracking of Smartwatch Orientation and Location by Multitask Learning. In *The 20th ACM Conference on Embedded Networked Sensor Systems (SenSys '22)*, November 6–9, 2022, Boston, MA, USA. ACM, New York, NY, USA, 2 pages. <https://doi.org/10.1145/3560905.3568096>

1 INTRODUCTION

Arm tracking is an essential sensing task for many mobile applications, such as gesture recognition, gym exercise assessment, and motion-based control. Real-time responses of these applications can

provide users with immediate feedback. Once we know the orientation and location of a user's wrist, we can estimate the user's elbow location via the user's forearm length, and in turn track the arm movements. As smartwatches become pervasively adopted, they provide the most convenient way to track the orientation and location of one's wrist, compared with the infrastructure-based arm tracking systems, like wireless sensing, visible light, and customized wearable sensors. The IMU sensors of smartwatches, including accelerometer, gyroscope, and magnetometer, can be used to track arm motions.

In this demo, we develop a Real-Time Arm Tracking (RTAT) system by tracking the orientation and location of a smartwatch using the IMU readings from the smartwatch. *RTAT* uses a multitask neural network for simultaneous prediction of both orientation and location. *RTAT* leverages Bidirectional Long Short-Term Memory (BiLSTM) as its backbone, considering its effectiveness for time-series data processing. Our multitask learning scheme overcomes the limitations of conventional arm tracking systems. First, *RTAT* estimates orientation and location simultaneously. Conventional methods first estimate the orientation and then use the estimated orientation to estimate the location. Small orientation errors may cause a large deviation of the location result. Our multitask neural network learns some common features for both orientation and location from the raw IMU sensor readings. Second, as a data-driven supervised learning method, *RTAT* learns the best fusion scheme of three IMU sensor data streams from the accurately labeled data, which is more immune to the noise of IMU sensor data [1–4]. Compared to the fixed data fusion methods (complementary filter), our arm tracking neural network adapts to the temporal variation of IMU sensor data. Third, *RTAT* is much faster than the conventional location estimation methods (i.e., 0.1633 ms vs. 2337.50 ms for processing 50 data samples on a desktop).

2 SYSTEM OVERVIEW

Figure 1 shows the two major parts of our system, i.e., a real-time arm tracking system, and a labeled data collection system. During the offline training phase, we feed the IMU readings of the smartwatch and the labels to *RTAT* to train a neural network model. During the online inference phase, a user can use *RTAT* by just wearing a smartwatch and running the well-trained model on a smartphone. The smartwatch transmits IMU sensor data stream to the smartphone via Bluetooth for real-time arm tracking. The labeled data collection system is no longer needed during the inference phase, *RTAT* is able

Permission to make digital or hard copies of part or all of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. Copyrights for third-party components of this work must be honored. For all other uses, contact the owner/author(s).

SenSys '22, November 6–9, 2022, Boston, MA, USA

© 2022 Copyright held by the owner/author(s).

ACM ISBN 978-1-4503-9886-2/22/11.

<https://doi.org/10.1145/3560905.3568096>

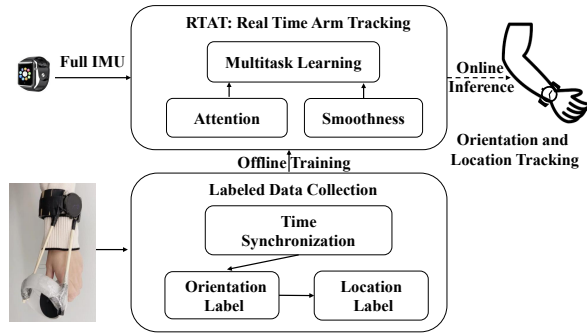


Figure 1: System Overview of RTAT

to track the orientation and location of the wrist on smartphones by the IMU readings of smartwatches.

RTAT’s Arm Tracking System. The inputs of the multitask neural network are the IMU readings from a smartwatch, and the outputs are the corresponding orientation and location series. Since the data from different IMU sensors plays varying importance for the orientation estimation along with time, we design an attention mechanism on top of our multitask neural network. The attention mechanism will adjust the input focuses of the network automatically according to the varying sensor noises. Furthermore, to guarantee the smoothness of the inferred arm motions, we employ smooth losses for both orientation and location tracking.

RTAT’s Labeled Data Collection. To train a model based on supervised learning, we need to build a training dataset. The dataset should consist of a large amount of time-series IMU readings from the smartwatch, and the corresponding orientations and locations of the smartwatch (labels). However, it is non-trivial to acquire accurate labels for the smartwatch. Thus we develop a labeled data collection system to derive the training labels. As depicted in Figure 1, the training data collection process requires volunteers to wear a smartwatch and hold a VR controller at the same time. The IMU sensor readings of the smartwatch will be the inputs to RTAT’s neural network, and the labels are collected from a VR system, which can provide accurate and real-time orientations and locations of the controller held by the users’ hand. However, the readings we collect from the VR system are the orientations and locations of the VR controller, but not the smartwatch. To fill in the gap between the acquired orientations and locations of the VR controller and the required ones of the smartwatch, we first synchronize these two systems and then design a set of experiments to convert the orientation and location of the VR controller to the smartwatch. Details can be found in our full conference paper [5].

3 DEMONSTRATION

We plan to demonstrate RTAT from two perspectives, i.e., latency and accuracy. To demonstrate latency, a user wears a smartwatch and draws some letters or numbers in the air. As shown in Figure 2, the smartwatch senses and transmits the IMU readings from the smartwatch to a paired smartphone via Bluetooth. The smartphone receives the IMU readings and then forwards the IMU readings to the TensorFlow Lite model on the smartphone. The model then predicts the smartwatch’s orientation and location. We will use an interface

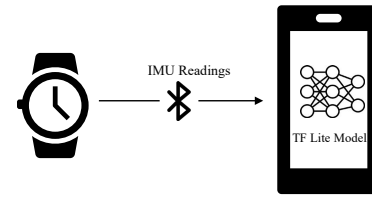


Figure 2: Data Flow for Inference on Smartphones

to show the model inference latency. For accuracy, we record a video to visualize the predicted orientation and location traces by RTAT.

Multitask Model. RTAT is implemented and trained by Keras in Python. Each BiLSTM layer has 32 units, and each fully-connected layer has 3 units. The optimizer is RMSprop, we set the learning rate as 0.0001. The training epoch is 100, the batch size is 128, and the time step we consider in BiLSTM is 32. The computer we use to collect data and develop our framework is Alienware Aurora R7, which contains one Intel Core i5 8400 CPU (6-core) and one NVIDIA GeForce GTX 1070 GPU (8GB memory).

TensorFlow Lite model. To execute RTAT on smartphones, we convert the well-trained TensorFlow model on the desktop to the TensorFlow Lite model capable of inference on mobile devices by TensorFlow Lite Converter. To do that, we froze the model as a TensorFlow concrete function. This process fixes the input size, which was specified as a tensor. It also specifies a *tf.function* which can be saved into a .pb file, the preferred input format for the TensorFlow Lite Converter. Running the TensorFlow Lite Converter on the saved model returned a FlatBuffer file with a .tflite extension which can be imported and used on mobile devices running Android or iOS, as well as some embedded devices and microcontrollers.

Experiment Platform. The smartwatch we use is Fossil Gen 5. It includes a LSM6DSO 6-axis accelerometer + gyroscope, and a AK0991X magnetometer. The sampling rate of the three sensors is set at 50 Hz. The smartphone we use to run RTAT is Google Pixel3. Meta Quest 2 is used to provide the ground truth for training and evaluating RTAT.

4 ACKNOWLEDGMENTS

We would like to thank our anonymous reviewers for their constructive comments. This research was partially supported by the NSF grant #CCF-2008837, two Seed Fund Awards (2020 and 2022 respectively) from CITRIS and the Banatao Institute at the University of California (UC), and the UC National Laboratory Fees Research Program grant #69763.

REFERENCES

- [1] Scott Sun, Dennis Melamed, and Kris Kitani. Idol: Inertial deep orientation-estimation and localization. *arXiv preprint arXiv:2102.04024*, 2021.
- [2] Martin Brossard, Silvere Bonnabel, and Axel Barrau. Denoising imu gyroscopes with deep learning for open-loop attitude estimation. *IEEE Robotics and Automation Letters*, 5(3):4796–4803, 2020.
- [3] Changhao Chen, Xiaoxuan Lu, Andrew Markham, and Niki Trigoni. Inonet: Learning to cure the curse of drift in inertial odometry. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 32, 2018.
- [4] Mahdi Abolfazli Esfahani, Han Wang, Keyu Wu, and Shenghai Yuan. Orinet: Robust 3-d orientation estimation with a single particular imu. *IEEE Robotics and Automation Letters*, 5(2):399–406, 2019.
- [5] Miaomiao Liu, Sikai Yang, Wyssanie Chomsin, and Wan Du. Real-time tracking of smartwatch orientation and location by multitask learning. In *Proceedings of the 20th ACM Conference on Embedded Networked Sensor Systems*, 2022.