Excerpt of Auritus: An Open-Source Optimization Toolkit for Training and Development of Human Movement Models and Filters Using Earables

Swapnil Sayan Saha, Sandeep Singh Sandha, Siyou Pei, Vivek Jain, Ziqi Wang, Yuchen Li, Ankur Sarker, and Mani Srivastava
swapnilsayan@g.ucla.edu
University of California, Los Angeles
Los Angeles, CA, USA

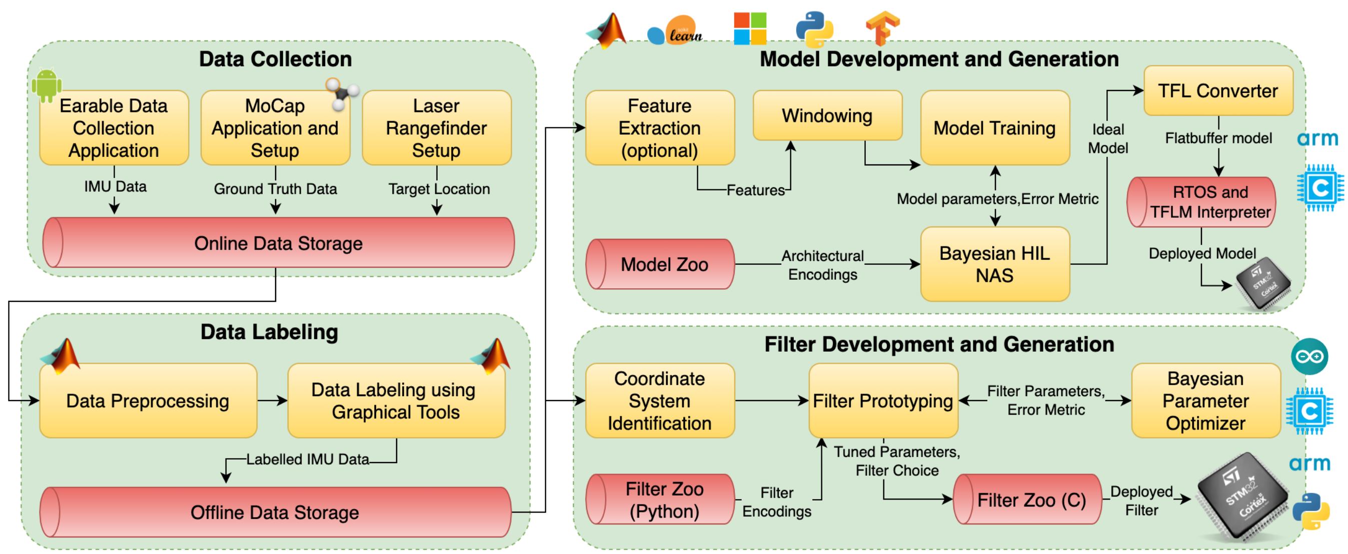

Figure 1: Architecture of Auritus [1]. The first two modules take care of collecting and labeling high resolution earable data interactively. The development and generation modules allow model and filter optimization through automated hardware-in-the-loop Bayesian neural architecture search and optimization, respectively.

ABSTRACT

Auritus is an extendable and open-source optimization toolkit designed to enhance and replicate earable applications. Auritus serves two primary functions. *Firstly*, Auritus handles data collection, pre-processing, and labeling tasks for creating customized earable datasets using graphical tools. The system includes an open-source dataset with 2.43 million inertial samples related to head and full-body movements, consisting of 34 head poses and 9 activities from 45 volunteers. *Secondly*, Auritus provides a tightly-integrated hardware-in-the-loop (HIL) optimizer and TinyML interface to develop lightweight and real-time machine-learning (ML) models for activity detection and filters for head-pose tracking. Auritus recognizes activities with 91% leave 1-out test accuracy (98% test accuracy) using real-time models as small as 6-13 kB. Our models are 98-740× smaller and 3-6% more accurate over the state-of-the-art. We also estimate head pose with absolute errors as low as 5 degrees using 20kB filters, achieving up to 1.6× precision improvement over existing techniques. Auritus is available at https://github.com/nesl/auritus.

UbiComp/ISWC '22 Adjunct, September 11–15, 2022, Cambridge, United Kingdom

ACM ISBN 978-1-4503-9423-9/22/09.
https://doi.org/10.1145/3544793.3563423

CCS CONCEPTS

• **Human-centered computing** → **Ubiquitous and mobile computing systems and tools**; • **Computing methodologies** → *Machine learning*; • **Computer systems organization** → Embedded systems.

KEYWORDS

earable, network architecture search, neural networks, machine learning, datasets, filters, human activity, head-pose, TinyML, optimization, hardware-in-the-loop

ACM Reference Format:
Swapnil Sayan Saha, Sandeep Singh Sandha, Siyou Pei, Vivek Jain, Ziqi Wang, Yuchen Li, Ankur Sarker, and Mani Srivastava. 2022. Excerpt of

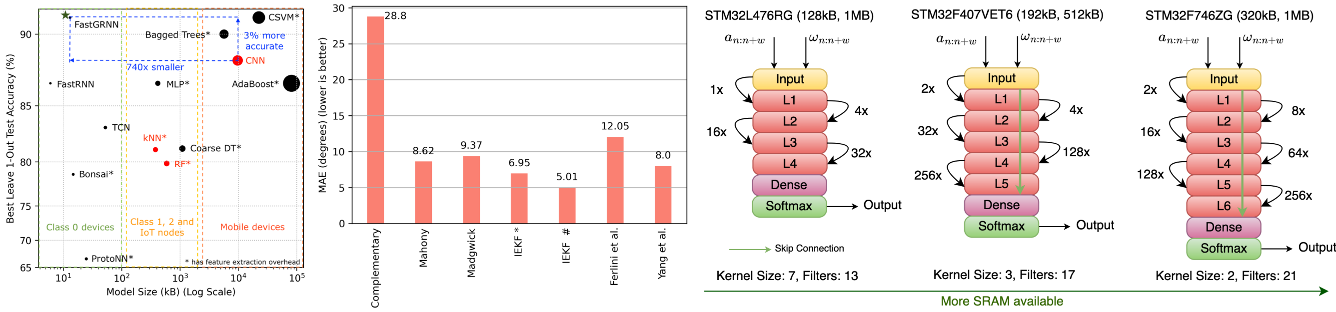

Figure 2: (Left) Accuracy vs. model size of Auritus models (black) vs. state-of-the-art models (red). (Middle) Error of Auritus filters vs. proposed earable head-pose filters. (Right) Our NAS adapts the same model to exploit full device capabilities.

Auritus: An Open-Source Optimization Toolkit for Training and Development of Human Movement Models and Filters Using Earables. In *Proceedings of the 2022 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp/ISWC '22 Adjunct), September 11–15, 2022, Cambridge, United Kingdom.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3544793.3563423

1 INTRODUCTION

The bulk of emerging innovations in earables builds upon ML advances for wearable activity detection. However, the data-hungry nature of ML training demands access to large-scale earable datasets, which are hard to come by. Moreover, given the tight memory, power, and compute constraints of earables (e.g., 56 kB SRAM, 16 MB flash), deploying on-device AI-driven activity detection and head-pose estimation is challenging. Auritus addresses the scarcity of earable datasets and software tools by providing:

- An open-source earable dataset from 45 volunteers containing 34 distinct head-poses and 9 classes of full-body activities with 2.43 million samples.
- Tools to enable data collection, processing and labeling.
- A zoo of 5 lightweight models, 5 conventional models, and 4 headpose filters that significantly outperform the state-of-the-art in terms of accuracy and resource usage.
- A HIL Bayesian neural architecture search (NAS) framework for training and deploying said models and filters on earables.

2 DATA COLLECTION AND LABELLING

For data logging, we used the eSense earable device from Nokia Bell Labs. The 6-channel inertial data was broadcasted at ∼100 Hz to an in-house smartphone application we developed using eSense Android middleware backend. For sub-mm resolution ground truth collection, the participants wore a hat with OptiTrack Prime 17W MoCap infrared markers. The motion data of the head and the marker visual cues were tracked using Motive:Tracker and screen recorder applications, respectively. For the head-pose dataset, we collected 34 distinct head-poses from 27 targets per participant, totalling 45 participants. For the activity dataset, 9 classes of actions were recorded, namely walking, jogging, jumping, standing, turning left, turning right, sitting, lying, and falling. A total 6.75 hours of human movement data is available. To ease labeling the data collected in continuous chunks, we designed a graphical-user-interface to allow head-pose and activity data annotation via plots. The application developer selects points directly on a plot signifying the start and endpoints of regions of interests. After specifying all the endpoints and making any numerical adjustments to the data, the developer runs a script to perform automatic segmentation and labeling based on the endpoints.

3 MODEL AND FILTER DEVELOPMENT

To enable real-time activity classification on resource-constrained devices, we included several lightweight classifiers and filters suitable for onboard activity inference in the model and filter zoo. The models include temporal convolutional network, fast gated RNN, fast RNN, Bonsai and ProtoNN. The filters include complementary filter, Madgwick filter, Mahony filter and indirect extended Kalman filter. To find the ideal activity detection model candidate from the model search space for limited flash, RAM, and latency requirements, or to optimize filter parameters, we included a HIL Bayesian optimizer. The goal is to find a model/filter that maximizes the hardware SRAM and flash usage within the device capabilities while minimizing latency and error on the validation set. Our optimizer communicates with the target hardware during the optimization process to guarantee deployability and architectural adaptation based on resource constraints. We use Gaussian process as the surrogate model and gradient-free Monte Carlo sampling with Upper-Confidence Bounds as the acquisition function.

4 KEY RESULTS

Auritus generates models that are 98-740× smaller, yet 3-6% more accurate over the state-of-the-art. Our models are as small as 6-13kB with 91-98% accuracy, and our fall detection models are as small as 2kB with 98% accuracy. Our filters have ∼5 degrees of error with 20 kB of code, providing 1.6× improvement over the state-of-the-art. Our NAS framework performs intelligent architectural adaptation and device capability exploitation based on resource availability.

REFERENCES

[1] Swapnil Sayan Saha, Sandeep Singh Sandha, Siyou Pei, Vivek Jain, Ziqi Wang, Yuchen Li, Ankur Sarker, and Mani Srivastava. 2022. Auritus: An Open-Source Optimization Toolkit for Training and Development of Human Movement Models and Filters Using Earables. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies* 6, 2 (2022), 1–34.