Semi-Supervised Learning for Efficient Perception of Human-Robot Walking Environments

Dmytro Kuzmenko, Oleksii Tsepa, Garrett Kurbis, Alex Mihailidis, and Brokoslaw Laschowski

Abstract - **Research has shown that convolutional neural networks trained using supervised learning can improve vision-based stair recognition for human-robot walking. Such advances were made possible using large-scale datasets like ExoNet and StairNet - the largest open-source image datasets of real-world walking environments. However, these datasets required vast amounts of manually annotated data, the development of which is time consuming and labor intensive. Here we present a new semi-supervised learning system (ExoNet-SSL) that uses over 1.2 million unlabelled images from ExoNet to improve training efficiency. We developed a deep learning model based on mobile vision transformers and trained the model using semi-supervised learning for image classification. Compared to standard supervised learning (98.4%), our ExoNet-SSL system was able to maintain high prediction accuracy (98.8%) when tested on new walking environments, while requiring 35% fewer labelled images during training. These results show that semi-supervised learning can improve training efficiency by leveraging large amounts of unlabelled data and minimize the size requirements for manually annotated images. Future work will focus on model deployment for onboard real-time inference and control of human-robot walking.**

I. Introduction

Research has shown that convolutional neural networks trained using supervised learning can improve vision-based automated stair recognition for control of wearable robotics. Such advances were made possible because of datasets such as ExoNet and StairNet, the largest open-source image datasets of real-world human-robot walking environments. However, these datasets required large amounts of manually annotated data, the development of which is time consuming and labor intensive. Here we developed a new semi-supervised learning model to improve training efficiency by significantly minimizing the number of required labelled images while maintaining high prediction accuracy comparable to state-of-the-art models for automated stair recognition [1].

II. Methods

Our research made use of unlabelled images from ExoNet (~4.5 million RGB images) to improve performance on the StairNet dataset. The class distribution of StairNet is greatly unbalanced such that the steady-state classes, LG and IS, comprise over 95% of the sampled images, whereas the transition classes, IS-LG and LG-IS, form the remaining 5%. We used the FixMatch semi-supervised learning algorithm as a proof-of-concept since it is relatively straightforward to implement. We developed a vision transformer (ViT) model using the base model of MobileViT, which is a transformer-based deep learning model that uses mechanisms of attention and depth-wise dilated convolutions.

Table 1. Comparison between supervised (MobileNetV2) and semi-supervised (MobileViT XS) learning models in terms of prediction accuracy and annotated image requirements.

Training Method	Accuracy	F1	Precision	Recall	Labelled Images
Supervised	98.4	98.4	98.5	98.4	461,328
Semi Supervised	98.8	98.9	98.9	98.8	300,000

III. Results and Discussion

Our new automated stair recognition system powered by semi-supervised learning uses large amounts of unlabelled data to improve training efficiency while maintaining high prediction accuracy. Compared to supervised learning (98.4% accuracy), our new semi-supervised learning model using mobile vision transformers achieved high classification accuracy during inference (98.8% accuracy) while requiring ~35% less annotated data, therein improving training efficiency. These results can help make deep learning systems for computer vision more accessible to researchers in wearable robotics and support the development of new autonomous controllers for human-robot walking.

References

[1] Kurbis, A. G., Laschowski, B., and Mihailidis, A. (2022). "Stair recognition for robotic exoskeleton control using computer vision and deep learning" in IEEE International Conference on Rehabilitation Robotics (ICORR).

*Research supported by AGE-WELL Networks of Centres of Excellence (NCE) program, Canada.

D. Kuzmenko is with the Department of Mathematics, National University of Kyiv-Mohyla Academy, Kyiv, Ukraine (e-mail: kuzmenko@ukma.edu.ua).

O. Tsepa is with the Department of Mathematical Methods of System Analysis, Igor Sikorsky Kyiv Polytechnic Institute, Kyiv, Ukraine, on placement in the Department of Computer Science, University of Toronto, Toronto, Canada (e-mail: tsepa.oleksii@lll.kpi.ua).

A. G. Kurbis is with the Department of Systems Design Engineering, University of Waterloo, Waterloo, Canada, on placement in the Temerty Faculty of Medicine, University of Toronto, Toronto, Canada (e-mail: agzkurbis@uwaterloo.ca).

A. Mihailidis is with the Institute of Biomedical Engineering, University of Toronto, Toronto, Canada, and the Toronto Rehabilitation Institute, Toronto, Canada (e-mail: alex.mihailidis@utoronto.ca).

B. Laschowski is with the Department of Mechanical and Industrial Engineering, University of Toronto, Toronto, Canada, and the Toronto Rehabilitation Institute, Toronto, Canada (e-mail: brokoslaw.laschowski@utoronto.ca).