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

Dmytro Kuzmenko[1,5], Oleksii Tsepa[2,5], Andrew Garrett Kurbis[3,5], Alex Mihailidis[4,5], and Brokoslaw Laschowski[4,5]

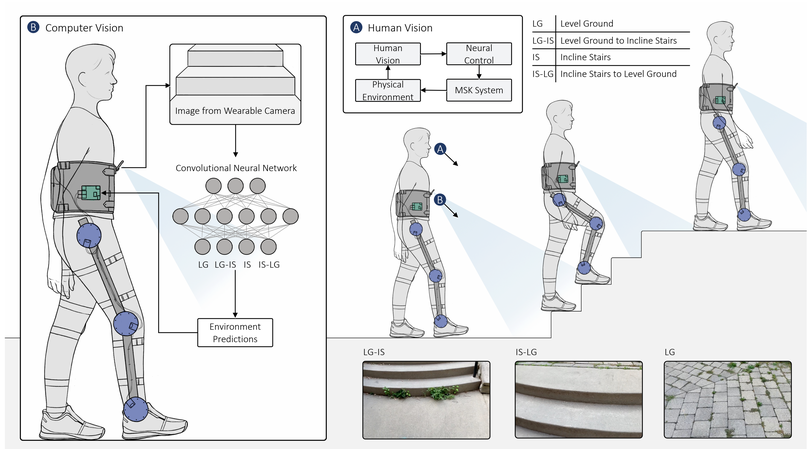
[1]National University of Kyiv-Mohyla Academy, Ukraine; [2]Igor Sikorsky Kyiv Polytechnic Institute, Ukraine; [3]University of Waterloo, Canada; [4]University of Toronto, Canada; [5]KITE-Toronto Rehabilitation Institute, Canada

Introduction

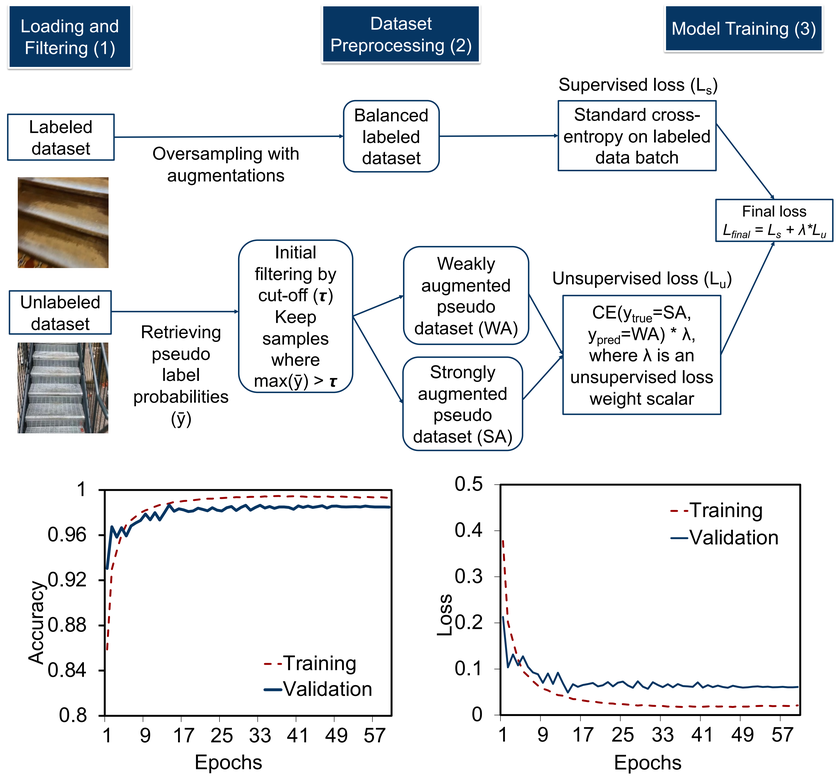
Research has shown that convolutional neural networks trained using supervised learning can improve vision-based automated stair recognition for control of wearable robotics. Such advances were made possible because of datasets such as *ExoNet* and *StairNet,* the largest open-source image datasets of 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.

Objectives

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 existing state-of-the-art models for automated stair recognition [1].



Methods



Results

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-score	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

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.

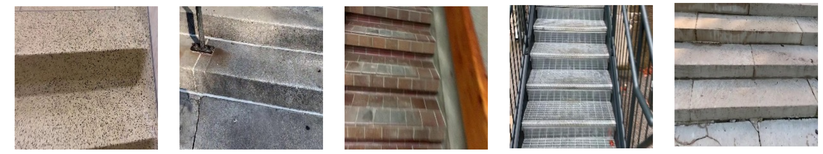


Table 2. Confusion matrix for the MobileNetV2 supervised model.

IS	96.9	1.5	0.3	1.3
IS-LG	6.5	90.5	2.4	0.6
LG	0.1	0.3	99	0.7
LG-IS	4.3	0.5	3.4	91.7
	IS	IS-LG	LG	LG-IS

Table 3. Confusion matrix for our MobileViT semi-supervised model.

IS	96.9	1.1	0.5	1.5
IS-LG	5.2	90.4	3.9	0.5
LG	0.1	0.1	99.5	0.3
LG-IS	3.4	0.4	5.6	90.6
	IS	IS-LG	LG	LG-IS

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. 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 in real-world environments.

References

1. Kurbis, et al. (2022). "Stair recognition for robotic exoskeleton control using computer vision and deep learning" IEEE International Conference on Rehabilitation Robotics (ICORR).

Contact: brokoslaw.Laschowski@uhn.ca