Demo: Learning to Perceive Long-Range Obstacles Using Self-Supervision from Short-Range Sensors

Mirko Nava,[1] Jérôme Guzzi,[1] R. Omar Chavez-Garcia,[1] Luca M. Gambardella,[1] Alessandro Giusti[1]
[1]All authors are with the Dalle Molle Institute for Artificial Intelligence (IDSIA), USI-SUPSI, Lugano, Switzerland.
This work is supported by the Swiss National Science Foundation (SNSF) through the NCCR Robotics.
mirko@idsia.ch

Abstract

We demonstrate a self-supervised approach which learns to detect long-range obstacles from video: it automatically obtains training labels by associating the camera frames acquired at a given pose to short-range sensor readings acquired at a different pose.

Introduction

We consider a robot capable of detecting obstacles only at short range (e.g. using proximity sensors). We demonstrate an approach for automatically learning to perceive long-range obstacles, given as input an image acquired by a forward-facing uncalibrated camera.

Self-supervised approaches have been frequently used for ground robot navigation, detecting obstacles or traversable regions (Dahlkamp et al. 2006; Brooks and Iagnemma 2012; Pinto and Gupta 2016). We use self-supervision to automatically collect training data while the robot randomly explores the environment. For a frame acquired at time t, ground truth information concerning the presence of an obstacle at a given distance (e.g., 20 cm) in front of the robot is obtained from the binary outputs of the proximity sensors observed *at a different time* t', i.e. when the robot is at a specific, different pose (in our example, 20 cm in front of the pose it had at t). The resulting information is collected in a dataset, and used for training a model mapping the camera frame to the presence of obstacles at a set of predefined distances.

This approach is an instantiation of a general idea applicable in other contexts (Nava et al. 2019), and extends our previous work (demo at AAAI 2018 (Toniolo et al. 2018)), where we only predicted the *current* outputs of the proximity sensors given the current camera frame; in that case, the practical applicability of the system was limited as the range of the camera-based obstacle detection could not exceed the range of the proximity sensors (about 8 cm).

System

Platform

We demonstrate the approach on a Mighty Thymio (Guzzi et al. 2018), a differential drive robot equipped with 5 forward-looking infra-red proximity sensors with a range of approximately $5\,\mathrm{cm}$ to $10\,\mathrm{cm}$, depending on the color and size of the obstacle. The robot is also equipped with a forward-looking 720p webcam with an horizontal field of view of $68°$, used as a long-range sensor (see Figure 1).

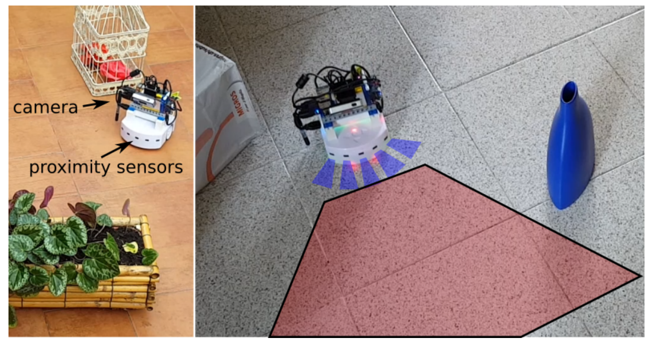

Figure 1: The Mighty Thymio robot. In blue, the proximity sensors' range. In red, the camera's field of view.

Data Acquisition

Robot odometry, proximity sensor readings and camera frames are gathered using an ad-hoc controller which randomly explores the environment; data was automatically collected from 10 different scenarios with very different floors and obstacles, for a total of 90 minutes.

Self-supervised Learning

We train a model which, given a camera frame, predicts 31×5 binary labels. Those are defined as the output that each of the 5 front-looking proximity sensors would have if the robot had advanced straight for a given distance (0 to 30 cm in steps of 1 cm).

Let $p(t)$ denote the pose of the robot at time t; a training instance is composed by the camera image acquired at time t, and a set of 31×5 automatically generated binary labels, which may be unknown for some of the 31 distances. In particular, for a given distance d_i between 0 and 30 cm, we look in the dataset for a time t_i such that $p(t_i)$ is at a distance d_i in front of $p(t)$. If such a t_i is found, the corresponding 5 proximity sensor outputs are associated to the 5 labels for d_i; if such a t_i does not exist, these labels are unknown.

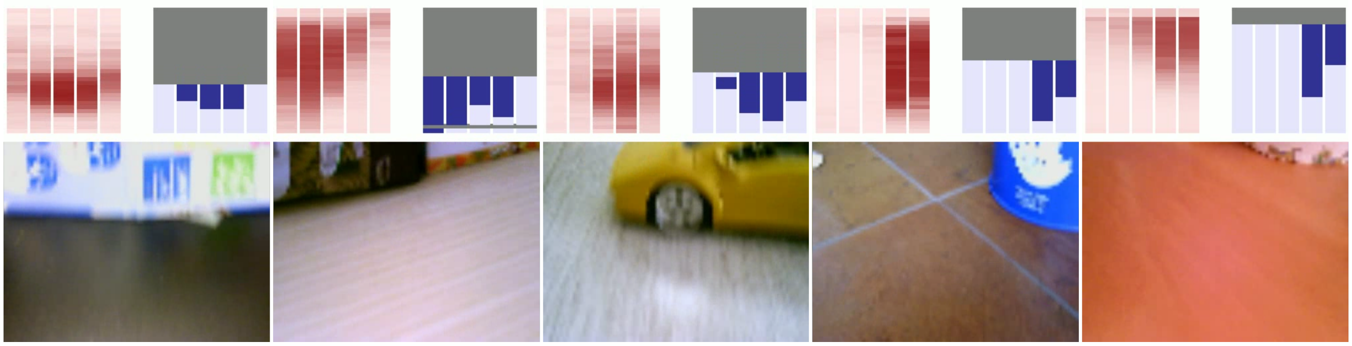

Figure 2: Five instances from the acquired dataset, each coming from a different scenario. Bottom: camera frame. Top left (red): predictions of a model trained on the other scenarios (one column per sensor; one row per distance: 0 to 30cm). Top right (blue): automatically generated ground truth labels. Gray rectangles denote missing labels.

The resulting dataset (with partially-known labels) is used to train a convolutional neural network, similar in structure to the one used in (Toniolo et al. 2018), which is trained using a masked loss to account for partially unknown labels.

Figure 3 reports the area under the receiver operating characteristic curve (AUC) values obtained for each sensor-distance pair on testing data from different scenarios. We observe that the quality of the prediction decreases with distance; also, obstacles which are on the side or lie very close to the robot are harder to detect (because they may be partly or completely out of the camera view).

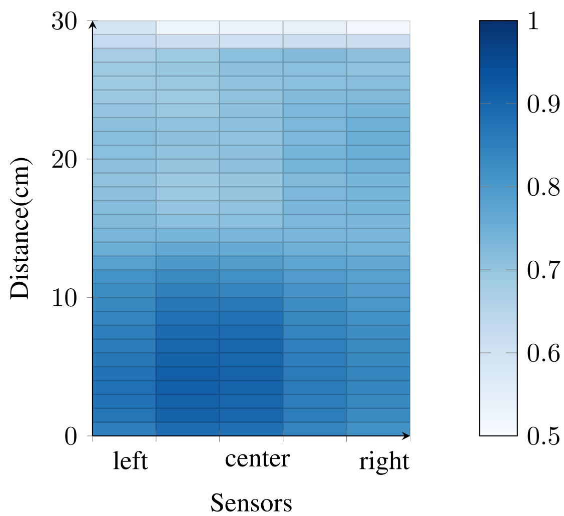

Figure 3: Mean AUC value obtained for each sensor (column) and distance (row).

Demonstration

We demonstrate the approach with a real Mighty Thymio robot implementing a pre-trained model onboard: the model generates predictions in real-time, which are used as the only input to a controller which steers away from detected obstacles. As a result, the robot reacts to obstacles much earlier than it would if it was only using proximity sensors. We also visualize the robot view and its predictions in real-time on a laptop. We can also demonstrate self-supervised learning by acquiring data on the spot to adapt the model.

Trained models generalize well to different scenarios: we can demonstrate usage on different robots (e.g. a TurtleBot) or even as an obstacle detector for the blind, when given data from a belt-mounted camera.

Videos, datasets, and code to reproduce our results are available at: https://github.com/idsia-robotics/learning-long-range-perception/.

References

Brooks, C. A., and Iagnemma, K. 2012. Self-supervised terrain classification for planetary surface exploration rovers. *Journal of Field Robotics* 29(3):445–468.

Dahlkamp, H.; Kaehler, A.; Stavens, D.; Thrun, S.; and Bradski, G. 2006. Self-supervised monocular road detection in desert terrain. In *Proceedings of Robotics: Science and Systems (RSS)*.

Guzzi, J.; Giusti, A.; Di Caro, G. A.; and Gambardella, L. M. 2018. Mighty thymio for higher-level robotics education. In *Proceedings of the 32nd AAAI Conference on Artificial Intelligence (The Eighth Symposium on Educational Advances in Artificial Intelligence, EAAI-18)*.

Nava, M.; Guzzi, J.; Chavez-Garcia, R. O.; Gambardella, L. M.; and Giusti, A. 2019. Learning Long-range Perception using Self-Supervision from Short-Range Sensors and Odometry. *ArXiv e-prints*.

Pinto, L., and Gupta, A. 2016. Supersizing self-supervision: Learning to grasp from 50k tries and 700 robot hours. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, 3406–3413.

Toniolo, S.; Guzzi, J.; Giusti, A.; and Gambardella, L. M. 2018. Learning an image-based obstacle detector with automatic acquisition of training data. In *Proceedings of the Thirty-Second AAAI Conference on Artificial Intelligence (AAAI-18 Demo Track)*.