



SRIN: A New Dataset for Social Robot Indoor Navigation

Kamal M Othman and Ahmad B Rad*

School of Mechatronic Systems Engineering, Simon Fraser University, Canada

*Corresponding author: Ahmad B Rad, School of Mechatronic Systems Engineering, Simon Fraser University, Surrey, Canada.

Received Date: February 19, 2020

Published Date: February 26, 2020

Abstract

Generating a cohesive and relevant dataset for a particular robotics application is regarded as a crucial step towards achieving better performance. In this communication, we propose a new dataset referred to as SRIN, which stands for Social Robot Indoor Navigation. This dataset consists of 2D colored images for room classification (termed SRIN-Room) and doorway detection (termed SRIN-Doorway). SRIN-Rooms has 37,288 raw and processed colored images for five main classes: bedrooms, bathrooms, dining rooms, kitchens, and living rooms. The SRIN-Doorway contains 21,947 raw and processed colored images for three main classes: no-door, open-door and closed door. The main feature of SRIN dataset is that its images have been purposefully captured for short robots (around 0.5-meter tall) such as NAO humanoid robots. All images of the first version of SRIN were collected from several houses in Vancouver, BC, Canada. The methodology of collecting SRIN was designed in a way that facilitates generating more samples in the future regardless of where the samples have come from. For a validation purposes, we trained a CNN-based model on SRIN-Room dataset, and then tested it on Nao humanoid robot. The Nao prediction results in this paper are presented and compared with the prediction results using the same model with Places-dataset. The results suggest an improved performance for this class of humanoid robots.

Keywords: Social robots; Nao; SRIN-dataset; CNN; Room classification

Introduction

Providing a seamless and reliable solution to indoor navigation is a central research problem in robotics as resolving this challenge is a precursor for success of many activities of a social robot. Indeed, achieving the ultimate objective of having a social robot in every home depends on a reliable solution to this problem. Social robots will be part of the family as pets are. They interact and assist in chores and will keep company for minors and seniors. Within this context, they must be able to flawlessly roam around the home and be able to identify different locations and their functionalities in a house. The authors of this communication presented a CNN-based model (Convolutional Neural Network) that demonstrated promising results with respect to room classification in an indoor setting [1]. As training a CNN-model requires a significant number of samples, there are many models trained on popular computer visions datasets, such as ImageNet [2] and Places [3]. However, adopting pre-trained CNN models that learned features from

computer vision datasets to be tested in real-time experiments on social robots, e.g. Nao humanoid robot [4] was not overwhelmingly successful [1]. We suggest that a dedicated dataset as opposed to general datasets such as ImageNet or Places could drastically improve the performance in real-time experiments.

Thus, the objective of this work is to report a new dataset called SRIN, which stands for Social Robot Indoor Navigation, for room classification and doorway detection applications in indoor environments – particularly in homes. This dataset has its unique feature for social robots with medium size and height, such as Nao humanoid robot, shown in Figure 1 which indicates the height of the robot with respect to the door handle. SRIN is a 2D RGB images dataset that consists over 75000 raw and augmented images for several rooms and doorways, i.e. SRIN-Rooms and SRIN-Doorway datasets, respectively. The distinct feature of the dataset is that all collected images were captured at the height of 0.5 m, which will be practically a useful dataset for social robots with medium size.

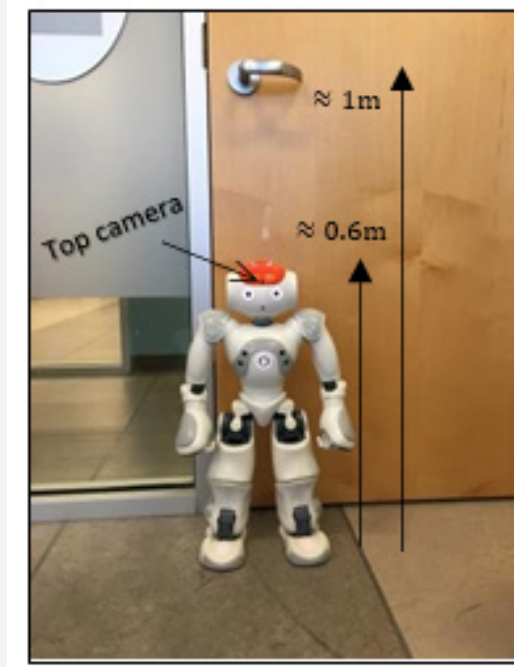


Figure 1: Nao robot's height with respect to the door handle.

The rest of the paper is organized as follows: Section 2 introduces related studies and presents different examples of robotics datasets as well as different examples of social robots with medium size. Next, we introduce the process of collecting SRIN dataset in section 3. We include some experiments and results in section 4 that demonstrates the superiority of SRIN to train CNN-model for robotics application and compare them with our previous reported results in [1]. We conclude the paper with additional remarks in section 5.

Related Studies

By and large, the performance of indoor robotic navigation tasks that employ general computer vision datasets has been rather poor [5,6]. To circumvent this shortcoming, robotics researchers have started developing their own dataset based on the robot's platform and its onboard sensors. We suggest that creating a robot-specific dataset improves the performance of navigation tasks. In [7], a dataset was acquired from a specific environment (only two locations of labs) using Pioneer and Virtual Me robots in which the camera is mounted in the height of 88 cm and 117cm, respectively. Although it is reported that 100-500 images per class for 17 classes were collected, the variety of the images was limited for the two main locations. The dataset of Robot@Home [5] is a collection of 83 time-stamped observations obtained from a mobile robot equipped with 4 RGB-D cameras and a 2D laser for robotics mapping applications in indoor domestic environments. Autonomous Robot Indoor Dataset, ARID [6], is an object-centric dataset that was collected by RGB-D camera on a customized Pioneer mobile robot, in which the camera was mounted at a height of over m. The main objective of that study was to test and compare previous CNN performances to

recognize 51 objects using more than 6000 RGB-D image scenes. Another dataset referred to as AdobeIndoorNav [8] was collected for Deep Reinforcement Learning (DRL) robotics applications from 24 scenes. Each scene consisted of 3D reconstruction point cloud, 360 panoramic view from grid locations, and 4 different images in 4 different directions using sensors mounted on a Turtlebot mobile robot.

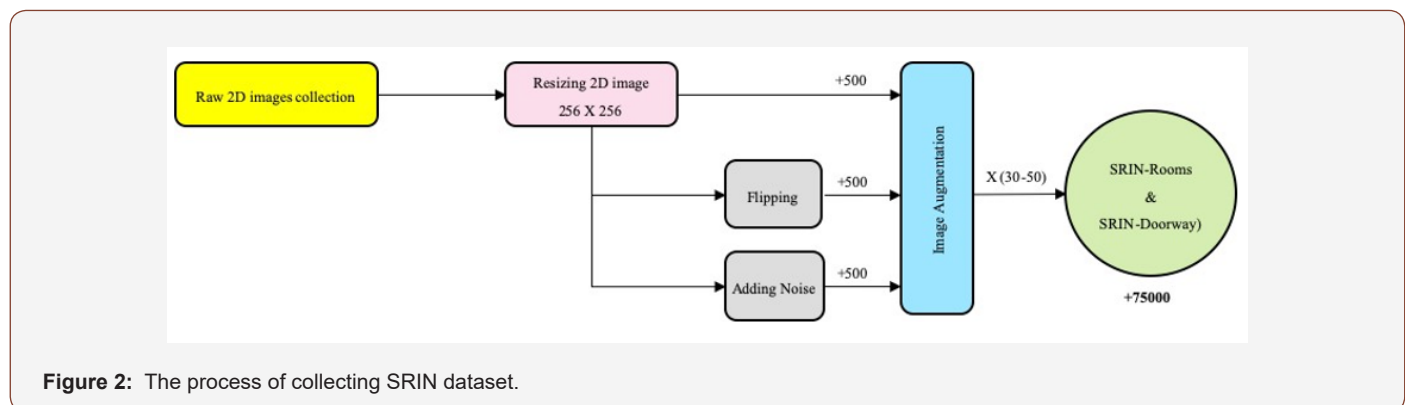
An affordable social robot in every home is one of the robotics researchers dreams in the near future. Such a robot most likely has a limited number of sensors, such as a monocular camera instead of a stereo camera or expensive 3D sensors. Such domestic social robots are probably medium sized and are the same height as domestic pets. There are already several types of mobile robots with a reasonable size in the market, i.e. less than 1 m. Among such robots are those with the average height of ~ 0.6 m, such as Nao by Aldebaran [4], QRio by Sony [9,10], Zeno by Hanson Robotics [11,12], Manava by Diwakar Vaish in the labs of A-SET Training & Research Institutes [13], and DARwIn-OP by Dinnes Hong in Robotics and Mechanisms Laboratory [14,15]. There are also other class of robots which are a bit shorter, i.e. the average of 0.4 m, such as HOVIS [16] and Surena-Mini [17], or a bit taller, i.e. the average of 0.8m, such as Poppy [18]. However, there is no suitable dataset for this kind of robots that can be used to address houses' environments navigation tasks such as recognizing room classes. Furthermore, testing a pure learned CNN-based model that trained on an existing computer vision dataset, such as Places dataset, with a medium-sized robot-like Nao, gives an undesirable performance as shown in [1]. Therefore, we propose a new scene-centric dataset called SRIN, which stands for Social Robot Indoor Navigation

dataset. This dataset has been collected to be employed for indoor navigation tasks on Nao humanoid robot or any other robot with a similar height.

Collecting Dataset Methodology

The main feature of the RGB images of SRIN dataset is that all raw images were taken at a height of 0.5m above ground from several houses in Vancouver, BC. In order to generalize our work for any social robot similar to Nao as well as to simplify the process

of increasing dataset in future without depending on taking Nao to different houses, it was preferable to start taking images by personal cameras restricted to the condition of the 0.5m height. These images were used to train and validate a CNN-based model, then the trained model was tested on Nao robot in new environments. SRIN dataset contains +75000 of the raw and augmented images for several rooms and doorways, i.e. SRIN-Rooms and SRIN-Doorway datasets. The procedure of collecting and increasing this dataset has been done through three main steps as shown in Figure 2:



- Over 500 raw images of five rooms classes (bedroom, bathroom, dining room, kitchen and living room) and three doorway classes (open-door, closed door or no-door) were captured from seven different houses.

- From each room, 5-8 images were collected with different scenes from different angles.

- In front of each door, 5-9 images were collected with different scenes from different distances and angles.

- If the room had a window, then more images were taken during the day, based on the sun light, and the night, based on the house's lights.

- All raw images were resized to a resolution of 256×256 .

- Each image was flipped and changed with Gaussian noise separately using OpenCV libraries. Therefore, the number of images was tripled.

- An aggressive augmentation process using Keras API [19] was applied on all images, creating the new dataset. This augmentation includes a random combination of rotation, width shift, height shift, shear range, zoom range and channel shift range.

SRIN dataset has several unique characteristics among which are: images are captured from the view of short robots as opposed to humans' view; and the dataset allows other researchers add images from indoor homes of different countries, cultures, etc. The criteria are also basic as it required only RGB images from different layouts from different rooms in a home. Figure 3 provides clarification of how the SRIN's content has been further processed by showing a raw image and its associated processed and augmented images per category.

Experiments and Results

In this section, we present the superiority of training the CNN-based model using SRIN dataset, i.e. let's call it CNN-SRIN, over Places dataset for indoor robotic applications. We present the validation accuracy, then test the trained model on real-time experiments with Nao robot. For appropriate validation and comparison, we applied the same CNN training process, via the concept of transfer learning as shown in Figure 4 and explained in detail in [1], with only SRIN-Rooms dataset for training the model. Then both CNN-SRIN results and CNN with places results in [1] have been compared with each other. The number of room samples in SRIN-Room dataset was 37288 images divided as follow: [bathrooms: 7538, bedrooms: 7634, dining rooms: 7561, kitchen: 7185, living rooms: 7370]. 20% of images of each class have been used for validation. Thus, the total images for training were 29832 samples, while 7456 samples were employed for validation.

As mentioned in [1], training the CNN model using places dataset for room classification problem was accomplished through two stages, see Figure 4. First stage was the training of the classifier part, i.e. fully connected network, while the feature extractor part, i.e. VGG16, was frozen, which implies that all their weights were non-trainable. The average of validation accuracy of this stage could not increase beyond 87%. In order to improve the performance, we added the second stage whereby the whole model was trained, i.e. both the feature extractor and the classifier parts. The validation accuracy reached to 93.29% after around 8:35 hours of training process on Graham cluster provided by Compute Canada Database [20]. Interestingly, training CNN-SRIN reached a validation accuracy of 97.3% after 1:32 hours of applying only the first stage of CNN transfer learning that is shown in Figure 4. Therefore, there was no need to go through the second stage.

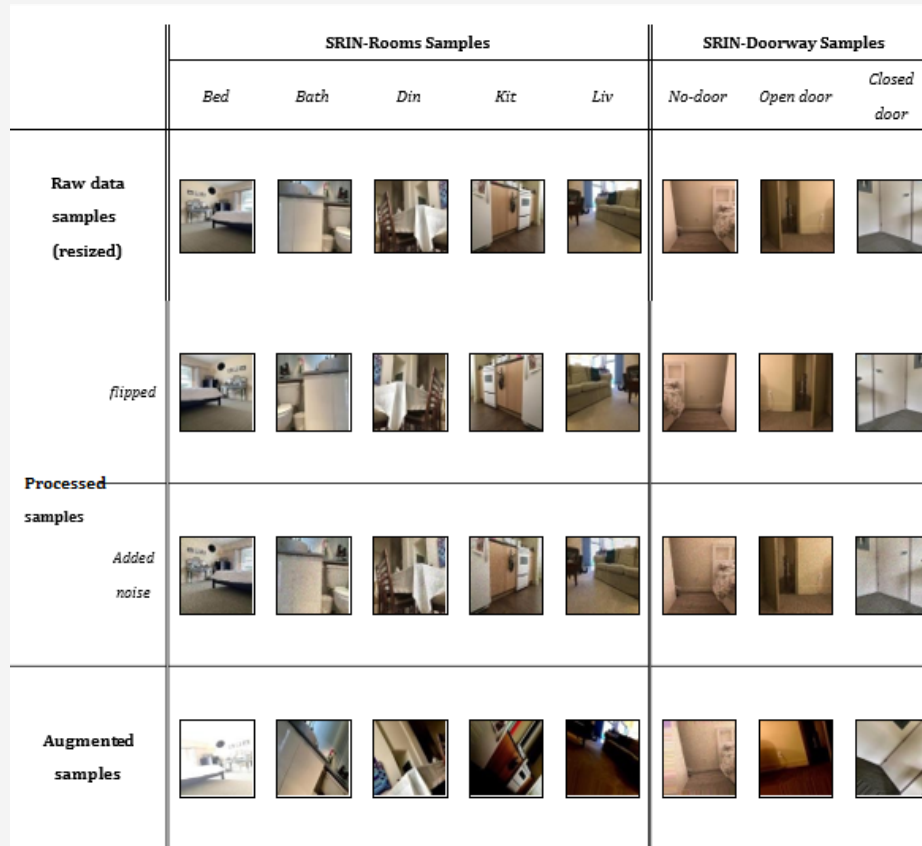


Figure 3: Scene samples from SRIN dataset for each category¹.

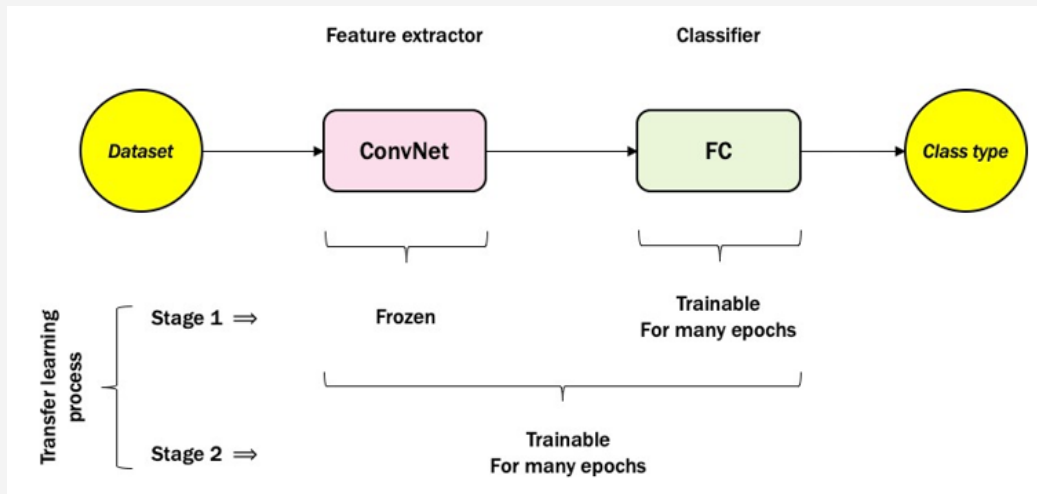


Figure 4: The concept of transfer learning of CNN. ConvNet: any pre-trained Convolutional Network & FC: Fully Connected Network.

As real-time experiments with Nao was our main concern, we tested the trained CNN-SRIN model on the same Nao images from [1]. Figure 5 shows some examples of Nao images from real-time experiments. Table 1 shows the comparison results of rooms predictions on Nao images for both trained models, i.e.

CNN with places dataset and CNN-SRIN. All results are shown as a confusion matrix for each class. Each model gives the top-1 and top-2 predictions, in which the top-2 is the results of the false prediction from top-1 results. As it is obvious in the last column of Table 1, the overall correct prediction from top-1 and top-2 of CNN-

¹All samples and python codes can be downloaded from the author’s GitHub page via this link: <https://github.com/KamalOthman/SRIN-Dataset.git>

SRIN outperformed the CNN model with Places. Four room classes attained 100% correct prediction in top-2 results. We believe that

with increasing the SRIN dataset in future, the prediction of top-1 for real-time experiments on Nao shall be improved as well.



Figure 5: Scene examples of Nao during real-time experiments².

Table 1: Comparison Results: confusion matrix of each class with top1 & top2 predictions for both models (CNN with places & CNN-SRIN) on a Nao Robot.

CNN-Room Classification with Places Dataset (from previous work [1])																	
		# images	Bath	Bed	Din	Kit	Liv	%		# FALSE	Bath	Bed	Din	Kit	Liv	%	Overall %
	Bath	12	12					100		0						-	100
	Bed	13	4	9				69		4		4				100	100
	Din	6	2		4			67		2		1	1			50	83.3
	Kit	13	10			2	1	15		11	1	4		6		55	61.5
	Liv	12	4	2			6	50		6		2	1		3	50	75
CNN-Room Classification with SRIN Dataset																	
		# images	Bath	Bed	Din	Kit	Liv	%		# FALSE	Bath	Bed	Din	Kit	Liv	%	Overall %
	Bath	12	12					100		0						-	100
	Bed	13		12			1	92		1		1				100	100
	Din	6			2		4	33		4			4			100	100
	Kit	13	6	1	1	3	2	23		10		1	1	5	3	50	61.5
	Liv	12	1	6			5	42		7					7	100	100

Discussion

As this work was concerned about the practical performance on a medium-sized mobile robot, i.e. Nao, Table 1 provided the promising practical results of SRIN dataset. Let us discuss the CNN-SRIN results based on the type of rooms. The “bathroom” is the most class with its unique features in its scenes in which the CNN-Room models (with Places and SRIN) predict 100% correctly in top1. Similarly, the “bedroom” is considered as a unique class that both CNN-models were able to get the 100 % correct predictions in top2. However, CNN-SRIN was much better in top1 results as it predicts 92.3% comparing to the 69.2% of CNN with Places. Next, despite the low percentage of CNN-SRIN prediction for the “dining room” and “living room” in top1, both classes were reached up to 100% in top2 results which is considered a huge improvement

compared to their results of CNN with Places. The most arguable class is the “kitchen”. As discussed in [1], the robot captured the cabinets and drawers rather than capturing the top view of stoves or other appliances. Thus, first predictions of CNN with Places were mostly bathrooms instead of kitchens. However, we consider this case was improved with CNN-SRIN from two perspectives. The first perspective is that the top1 percentage raised from 15.4% to 23.1% although they are quite low. The second perspective is that the number of first predictions as “bathroom” were less than CNN with places, which means the CNN-SRIN model learned different features. While, both models were reached the same percentage in the top2 results. The questions now are how can practically top1 results be improved and how can the robot make a decision based on top1 results? First, we believe that with increasing the SRIN

² All Captured images by Nao were saved and added to the same project in this link: <https://github.com/KamalOthman/SRIN-Dataset.git>

dataset in future, the prediction of top-1 for real-time experiments on Nao shall be improved as well. Second, as what we noticed with most of the “dining room” predictions that their probabilities of CNN-SRIN as “living room” very close to their probabilities as “dining room”. In this case, a threshold between top1 and top2 probabilities’ values can be assigned to make the right decision as well as capturing different views while navigating to maximize the correct prediction.

Conclusion

This communication presents a new Social Robot Indoor Navigation dataset called SRIN. It consists of 2D colored images for both rooms and doorways, i.e. SRIN-Rooms & SRIN-Doorways, respectively. SRIN is a useful dataset for medium sized social robots in indoor environments, specifically houses. SRIN has been validated through training a CNN-based model using SRIN-Rooms dataset, and then tested on the same Nao’s images in [1] for real-time experiments validation. The novelty of this work was illustrated when the validation accuracy of CNN-SRIN for room classification reached to 97.3% in a relatively short time. This was a huge improvement compared to training the same architecture with Places dataset, shown in [1], that reached 93.29% of validation accuracy after a long time. In addition, the significance of this work was also shown through the comparison of the performance of two models on real-time experiments for Nao. It shows a big improvement of predicting bedrooms and slightly different performance of other classes in Top1. However, it reached the 100% in the Top2 of the correct predictions for four classes out of five. We believe that with increasing the SRIN dataset in future, the prediction of top-1 for real-time experiments on Nao shall be improved as well. Future work will focus on using the same model for doorway detection problem using SRIN-Doorway dataset within a robotic navigation scenario. The research reported here is part of a larger project undertaking navigation in home environments. As such, details about mapping does not fall within the scope of this paper but is an important part of another on-going project.

Author Contribution

Kamal M Othman is a PhD candidate undertaking his research under the supervision of Prof. Ahmad Rad.

Acknowledgment

Authors acknowledge the laboratory mates, neighbors and friends for their support of collecting SRIN Data. This research was enabled in part by support provided by WestGrid (<https://www.westgrid.ca/>) and Compute Canada (<https://www.computecanada.ca/>).

westgrid.ca/) and Compute Canada (<https://www.computecanada.ca/>).

Conflicts of Interest

The authors declare no conflict of interest.

References

1. KM Othman and AB Rad (2019) An Indoor Room Classification System for Social Robots via Integration of CNN and ECOC. *Appl Sci* 9(3): 470.
2. JDJ Deng, WDW Dong R Socher, LJ LLJ Li, KL K Li, L FF L Fei-Fei (2009) ImageNet: A large-scale hierarchical image database. 2009 IEEE Conf Comput Vis Pattern Recognit, pp. 2-9.
3. B Zhou, A Lapedriza, A Khosla, A Oliva, A Torralba (2017) Places: A 10 million Image Database for Scene Recognition. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.
4. Nao humanoid robot.
5. JR Ruiz Sarmiento, C Galindo, J Gonzalez Jimenez (2017) Robot@Home, a robotic dataset for semantic mapping of home environments. *Int J Rob Res*.
6. M Reza Loghmani, B Caputo, M Vincze (2018) Recognizing objects in-the-wild: Where do we stand? in *Proceedings - IEEE International Conference on Robotics and Automation*.
7. R Sahdev, JK Tsotsos (2016) Indoor place recognition system for localization of mobile robots. in *Proceedings - 2016 13th Conference on Computer and Robot Vision*.
8. K Mo, H Li, Z Lin, JY Lee (2018) The AdobeIndoorNav Dataset: Towards Deep Reinforcement Learning based Real-world Indoor Robot Visual Navigation.
9. Qrio, the robot that could. *IEEE Spectr* 41(5): 34-37.
10. Qrio (2003)
11. D Hanson et al. (2008) Zeno: A cognitive character. in *AAAI Workshop - Technical Report*.
12. Zeno research robot.
13. Meet Manav, India’s first 3D-printed humanoid robot.
14. I Ha, Y Tamura, H Asama, J Han, DW Hong (2011) Development of open humanoid platform DAR wIn-OP. in *Proceedings of the SICE Annual Conference*.
15. DAR wIn OP: Open Platform Humanoid Robot for Research and Education.
16. HOVIS Guide.
17. A Nikkhah, A Yousefi Koma, R Mirjalili, HM Farimani (2018) Design and Implementation of Small-sized 3D Printed Surena-Mini Humanoid Platform. in *5th RSI International Conference on Robotics and Mechatronics, IcRoM*.
18. M Lapeyre et al. (2014) Poppy Project: Open-Source Fabrication of 3D Printed Humanoid Robot for Science, Education and Art. *Digital Intelligence* 2014.
19. F Chollet (2015) Keras Documentation. *Keras.io*.
20. Compute Canada.