Tactile-based human-robot collaboration: a performance analysis*

Francesco Grella[1], Roberto Canale[1], Francesco Giovinazzo[1], Alessandro Albini[2], and Giorgio Cannata[1]

[1] Università degli Studi di Genova, Genova, Italy
[2] Oxford Robotics Institute, Oxford, United Kingdom

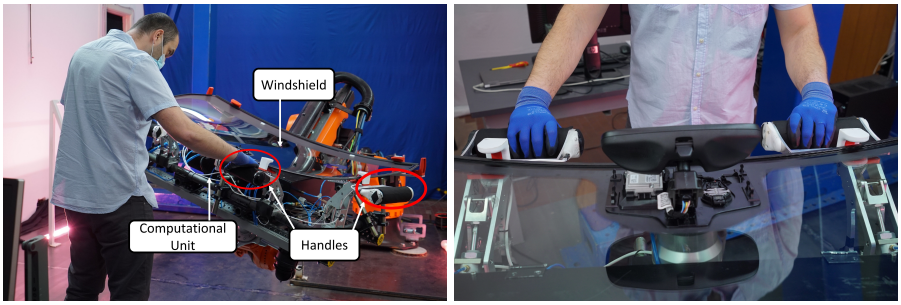
Abstract. In this paper we present a performance analysis of a perceptual architecture for industrial robots based on tactile sensing. We have developed a human-robot interface based on sensing devices which can be fixed on the robot body or the gripper. The devices are cylinder-shaped handles covered with tactile sensors, which can be intuitively grasped by an operator. Tactile data are fed into a neural network in order to recognize human touch during grasp, thus providing an enabling command for the control system. We provide an inference time comparison between two computing architectures: a desktop workstation with an high-performance GPU and an embedded solution based on the NVIDIA Jetson Nano board. We also compare inference time obtained from three instances of the neural network, compiled with three different engines: Keras, TensorRT floating-point 16 and TensorRT floating-point 32, showing that numerically optimized models allow to perform inference also on the embedded board without violating timing constraints imposed by data acquisition. Moreover, we assess the robustness of touch recognition when the user is wearing work gloves, showing that the difference from the bare-hand case is negligible.

Keywords: Tactile Sensing · Human-Robot Interaction · Deep Learning.

1 Introduction

Industry 4.0 paradigm [1] foresees the co-existence of humans and robots in applied manufacturing scenarios. Human's experience and dexterity should be leveraged to handle new situations and solve complex problems, while robot's strength, high speed and accuracy should be used to accomplish monotonous, dangerous and repetitive tasks. Compared to more traditional manufacturing layouts which expect the robot to be confined within a restricted workspace, *human-robot collaboration* (HRC) aims towards the realization of tasks performed in a shared workspace. This novel working configuration allows to define and perform operations that can exploit capabilities both from the human and

* Supported by the European Community's Framework Programme Horizon 2020 (grant agreement number 820767, project CoLLaboratE.)



(a) (b)

Fig. 1. Two views of the use-case scenario. Image (a) shows the gripper which holds the windshield. Data acquisition is performed by devices contained in the black box below the vacuum which holds the glass. In image (b) the two front handles and the work gloves of the operator are clearly visible. The model is similar to the glove used for robustness experiment.

the robot, but also raises challenges in terms of safety and communication between the operator and the machine. The analysis described in this paper is contextualized in the car windshield assembly use-case of the CoLLaboratE H2020 [2] project. We designed a cell to achieve HRC by means of physical Human-Robot Interaction (pHRI), a specific class of interaction situation in which an operator is expected to make contact on one or more points of the robot body. In the depicted scenario (presented in fig. 1), the industrial manipulator is used as a weightlifting system which absorbs the whole effort of moving an high-inertia object around the workspace. The capability of effortlessly moving a windshield by hand also improves the flexibility of maintenance operations, since the user can now tailor and adjust the position and orientation of the object according to his/her needs. If the inspection and manufacturing operations would occur on a fixed object, the worker should instead adapt to the position of the object with the risk of being penalized by proper anatomic characteristics such as height.

In order to be guided by an operator, the robot motion must be governed by an admittance controller which takes as input an external force/torque and outputs a proper velocity command which makes the robot follow the applied wrench. Such motion behavior anyway does not distinguish who or what is actually applying the force: if the operator hits the robot accidentally an undesired and dangerous movement will be performed. Also an object inattentively laid on the payload or the gripper could cause a constant force applied on the robot, thus a constant velocity which can bring the robot to collide with the ground, other equipment and people. These considerations justify and contextualize the need of a way to discriminate between accidental and voluntary contact. In this scenario human contact is used as a novel communication interface through which the human can directly manifest his/her intention to interact with the robot by classifying the data obtained from the tactile sensors. Voluntary interaction is

detected when the shape of a human hand is recognized by the classifier, meaning that the operator is deliberately grasping a handle in order to move the robot. Endowing an industrial manipulator with an external sensing architecture whose feedback actively affects motion control yields implementation issues which need specific design choices. In pHRI one major requirement is safety, which in reactive control architectures also depends on system's response. In the proposed application the activation/deactivation of the robot motion based on hand detection should not introduce latency to the control system, so data processing should be performed within the time window of each new acquisition step. Thus, given the rate bottleneck of the sensors and knowing that inference from deep learning models is a computationally expensive processing task, proper hardware should be used for online inference. In literature, pHRI is addressed with different sensing architectures. Authors in [3] and [4] address physical interaction by means of admittance control strategies relying only on the feedback provided by joint level torque sensors on small-sized robots. These solutions are unfeasible when dealing with heavy industrial manipulators, since joint level sensors are not a common design choice. Moreover heavy-duty robots lack the necessary sensing capability to detect and localize a light contact due to the weight of their links.

To endow industrial manipulators with physically interactive behaviors purely reactive compliance-based approaches could not be enough to ensure safety standards described in works such as [5]. In [6] a survey on human-robot collaboration in industrial settings is presented, however only few works fall in the area of 'shared-payload' tasks like the one described throughout this work. In [7] an industrial robot as flexible weightlifting system is presented, but no direct measurement feedback between the machine and the operator is considered. Therefore a distributed tactile interface providing a higher level abstraction about the tactile information should be leveraged. In [8] tactile image processing is used to semantically interpret grasp poses to produce discrete motions on a lightweight manipulator. Among the cited papers authors anyway do not consider use-case scenario and therefore no analysis on implementation details is provided. Moreover, in [8] which addresses human hand detection no considerations are provided with respect to robustness in presence of protective clothing such as gloves, which are mandatory in manufacturing environments. An application based on tactile sensing to detect user grasp should consider the fact that the operator is very likely to wear protective gloves which can alter the sensors response. To the best of our knowledge, no such analysis is available in literature. Thus, the main contributions of this paper are the following:

- ◇ a performance analysis of the learning model in terms of inference time. In particular we carried out a comparative analysis between three instances of the same model: a Keras [9] model and two models compiled with the optimization framework *TensorRT* [10] with different floating point precision. The tests have been iterated on two different computational architectures based on a RTX 3070 GPU and a Jetson Nano board.
- ◇ a robustness analysis on hand detection while the user is wearing a protective glove, to assess its impact on the performance of the classifier.

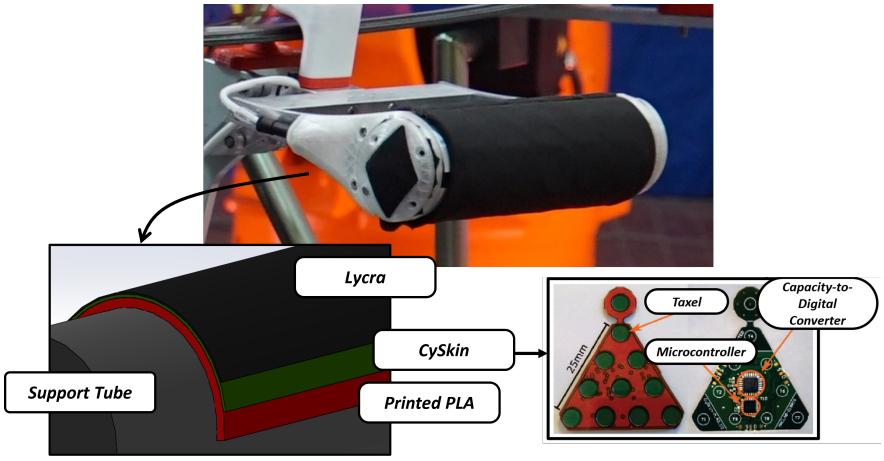


Fig. 2. Detail on the device structure and CySkin architecture.

The rest of the paper is organized as follows:
Section 2 provides an overview of the sensing architecture implemented for the use-case and details regarding the fine-tuning procedure we adopted to improve accuracy performance. Section 3 describes inference time evaluation, while section 4 presents results related to robustness analysis. In section 5 we provide our conclusions as well as insights for an extension of the research.

2 Voluntary Interaction Detection Architecture

This paper presents a performance analysis of an architecture which allows a human operator to physically interact with an industrial robot. The system architecture and methodology are already presented in detail in [11] and [12], so it will only be briefly reviewed in this section. Our pipeline builds a 2D representation identified as *tactile image* from raw tactile data, which is then processed by a binary classifier based on a convolutional neural network which outputs the probability of both *hand* and *non-hand* classes. The output value of the *hand* class is then used to enable robot motion control. The device designed for this application is a cylindrical handle covered by distributed tactile sensors arranged in flexible patches which allow to cover curved surfaces. The handle is covered with the CySkin technology, presented in [13]. It consist of an artificial skin composed of distributed tactile sensing elements (i.e. *taxels*) measuring the pressure distribution generated by the physical interaction. The skin is composed by flexible interconnected modules that can be cut and adapted to curved surfaces. Each modules can contain up to 11 capacitive pressure transducers. New data are sampled at a rate of 20Hz which represents the main timing constraint for our application. It's desirable to meet an inference time which is below 0.05s in order to classify voluntary interaction within the acquisition step. Fig. 2 shows

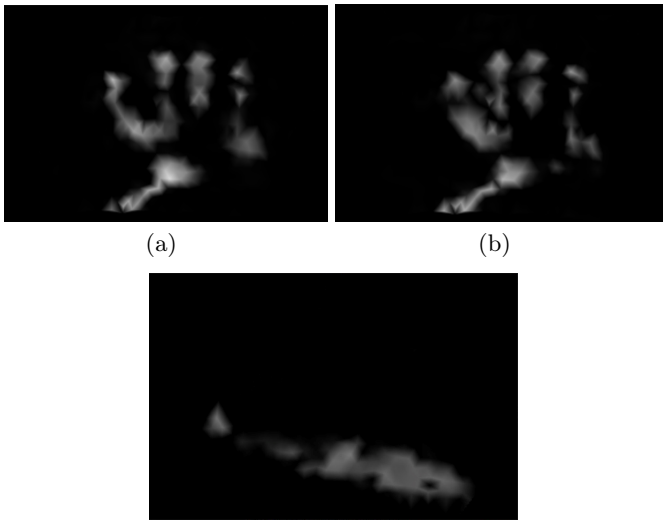


Fig. 3. Three instances of tactile images during interaction events: (a) grasp with bare hand, (b) grasp with working gloves, (c) forearm contact. It can be noticed that when the user is wearing the glove the region between the index and middle fingers is detected as pressed due to the thickness added by the tissue.

the structure of the handle and the CySkin technology. Given the discrete set of pressure sensors, a single sensing element is referred as *taxel* with an associated pressure response. By following a geometric pipeline described in [14] the set of pressure values is interpolated over a triangular mesh approximating the covered surface. A further flattening step leads to a *2D* surface which is transformed in a greyscale image. Fig. 3 shows three examples of tactile images acquired during the experiments for the presented analysis. At each acquisition step a tactile image allows to acquire information on the contact shape. We use a convolutional neural network (*HandsNet*) presented in [11] in order to predict the probability of having a human hand touching the device given the image. The training set was acquired from human subjects touching and grasping the skin-covered forearm of a robot. In order to improve accuracy on images acquired from the handle we performed a fine-tuning of the network which is presented in detail in [12]. The training procedure led to and accuracy of 98.08% and 94.61% on the validation and test sets respectively. The work depicted in [12] also presents an early performance analysis of the detection system on the Jetson Nano board. Results showed a slightly better performance of *fp16* against *fp32*, but the analysis was more conceived as a proof-of-concept for the system and has several differences from the study proposed in this article. In particular a lower amount of data was used and no online acquisition was performed, but frames were taken from a pre-recorded dataset. Moreover the timing analysis also included preliminary

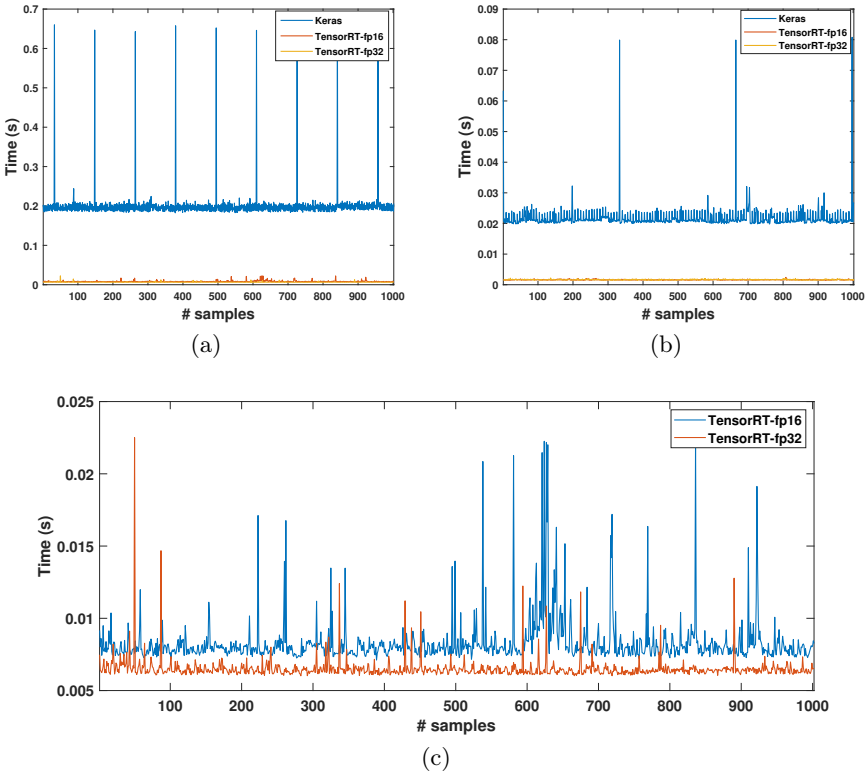


Fig. 4. Inference time plots on 1000 samples. (a) shows the inference time obtained on Jetson Nano, (b) the inference time obtained on RTX 3070. (c) is an insight on fp16 and fp32 obtained on Jetson Nano experiments

processing steps such as image resizing for the input layer of the network, since it is focused on the architecture feasibility more than pure inference performances.

3 Inference Time Performance Evaluation

We propose a comparison between two devices: a desktop workstation with a *RTX 3070 GPU* and a *Jetson Nano* board, both manufactured by Nvidia. Assuming that data are acquired by devices which are placed on the robot gripper, the adoption of a workstation for computing inference requires a more complex communication architecture since the size and weight of the machine do not allow direct placement on the robot. A local network should be used to send data from the on-board sensors to the computing device and then the prediction back to the robot controller, eventually introducing a source of latency. On the other

Table 1. Results of the inference time analysis, expressed in seconds (s).

	Keras			**TensorRT (fp16)**			**TensorRT (fp32)**		
	mean	max	std	mean	max	std	mean	max	std
RTX 3070	0.0216	0.0808	0.0038	0.0016	0.0024	$9.8\ 10^{-5}$	0.0016	0.0022	$1.13\ 10^{-4}$
Jetson Nano	0.1991	0.6786	0.0438	0.0083	0.0224	0.0017	0.0065	0.0225	$8.01\ 10^{-4}$

hand, having a small GPU board placed directly on the gripper allows to perform data acquisition and model inference on the same device, leaving only one communication channel towards the controller. We have performed an inference time analysis by running the model on real data acquired by CySkin in order to assess the feasibility of both solutions. Since the presented analysis is independent from the robot motion, we have used a single sensorized handle fixed on a table. Since this study presents an analysis purely focused on detection performances, data have been acquired from a single subject interacting with the device, and a set of 1000 frames has been considered for each experiment. An in-depth user study which also involves robot motion performances is currently being performed. We also have considered three different instances of HandsNet for each architecture as a complement for our analysis. We have collected inference time from a plain Keras model directly obtained by the framework used for building the network and from two models obtained by compilation with TensorRT (TRT), a toolchain which provides optimization for GPU architectures. The difference between the two optimized models is the floating-point precision used for representing the network parameters: here we discuss the behavior of *floating-point 16* and *floating-point 32*.

Fig. 4 (a) and (b) and Table 1 show that the Keras model exhibits poor time performance compared to TRT models on both devices. Only on the workstation the time constraint is met by observing the mean inference time. Anyway, it must be noticed that peaks reaching $0.0808s$ are detected, so there is no guarantee that inference can be always performed within the same acquisition step. On both architectures TRT models have obtained acceptable results in terms of inference time. On RTX 3070 the difference in performance between fp16 and fp32 is negligible, instead on the Jetson board the fp32 model has a mean inference time improvement of $1.8ms$. The fp32 model on Jetson also shows lower standard deviation with respect to the fp16 counterpart. The better quantitative performance and stability of the fp32 model is also shown in Fig. 4 (c).

4 Inference Accuracy Robustness

The proposed use-case scenario is set in an industrial environment and is focused on physical interaction between the robot and a specialized worker. Voluntary interaction detection is performed by classifying a human hand shape from a tactile image when the human is grasping the devices described in Sec. 2. In assembly departments there are strict safety policies which the workers must follow, especially regarding protective clothing. In the described scenario the

operator is expected to grasp the handles while wearing protective gloves, like shown in Fig. 1. Since the neural network was originally trained with bare hand grasps, we provide an assessment of its classification performance when grasp is performed while wearing gloves. We acquired data from a user grasping one fixed handle and holding the grasp for two minutes while randomly rotating the wrist to emulate the small motions which occur while the robot moves. We analyzed the probability value of the hand class in terms of mean, minimum value and standard deviation. As shown in Table 2 wearing a work glove has a small impact on the prediction value. In particular the presence of the glove slightly increases fluctuations as can be seen by observing the standard deviation and the minimum value, but in terms of classifications these values are negligible even by choosing a tight threshold of 0.9 to enable the controller.

Table 2. Results of comparison between hand class probability in bare hand grasp and work glove grasp. Reported results represent probability values, thus they are dimensionless and limited between 0 and 1.

	mean	min	std
bare hand	1.00	0.99	$5.14\ 10^{-6}$
work glove	0.99	0.95	0.0019

5 Conclusions

In this paper we have presented a performance analysis for a tactile-based sensing architecture designed for physical human-robot interaction in an industrial scenario. We have evaluated the impact of the computational architecture and deep learning model optimization in terms of inference time, showing that an embedded board with a low-cost GPU is capable to provide a prediction for each acquired image without exceeding the sampling period of $0.05s$. Both fp16 and fp32 models allow to obtain performances on the Jetson board which are very close to those observed on the RTX 3070. This shows that model optimization could enable deep learning model inference in applications with tight space or power consumption constraints. We have also tested the robustness of our hand classification model with respect to grasp performed while wearing working gloves. We have obtained results showing that this common safety practice in industrial environment does not affect voluntary interaction detection. Further steps in our research will involve performance evaluation during operations on the full use-case scenario, as well as a user study on multiple subjects. Both studies are currently under way.

References

1. L. D. Xu, E. L. Xu, and L. Li, "Industry 4.0: state of the art and future trends," *International Journal of Production Research*, vol. 56, no. 8, pp. 2941–2962, 2018.

2. "Collaborate project." https://collaborate-project.eu/.
3. E. Magrini, F. Flacco, and A. De Luca, "Control of generalized contact motion and force in physical human-robot interaction," in *2015 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2298–2304, 2015.
4. A. Q. Keemink, H. van der Kooij, and A. H. Stienen, "Admittance control for physical human–robot interaction," *The International Journal of Robotics Research*, vol. 37, no. 11, pp. 1421–1444, 2018.
5. S. Bragança, E. Costa, I. Castellucci, and P. M. Arezes, *A Brief Overview of the Use of Collaborative Robots in Industry 4.0: Human Role and Safety*, pp. 641–650. Cham: Springer International Publishing, 2019.
6. V. Villani, F. Pini, F. Leali, and C. Secchi, "Survey on human–robot collaboration in industrial settings: Safety, intuitive interfaces and applications," *Mechatronics*, vol. 55, pp. 248–266, 2018.
7. V. Gopinath and K. Johansen, "Risk assessment process for collaborative assembly – a job safety analysis approach," *Procedia CIRP*, vol. 44, pp. 199–203, 2016. 6th CIRP Conference on Assembly Technologies and Systems (CATS).
8. T. Li, K. Zheng, J. Liu, X. Jia, and J. Feng, "Research on operation intention based on flexible tactile sensing handle," *IEEE Access*, vol. 9, pp. 12362–12373, 2021.
9. F. Chollet *et al.*, "Keras," 2015.
10. "Nvidia, tensorrt." https://developer.nvidia.com/tensorrt.
11. A. Albini and G. Cannata, "Pressure distribution classification and segmentation of human hands in contact with the robot body," *The International Journal of Robotics Research*, vol. 39, no. 6, pp. 668–687, 2020.
12. F. Grella, G. Baldini, R. Canale, K. Sagar, S. A. Wang, A. Albini, M. Jilich, G. Cannata, and M. Zoppi, "A tactile sensor-based architecture for collaborative assembly tasks with heavy-duty robots," in *2021 20th International Conference on Advanced Robotics (ICAR)*, pp. 1030–1035, 2021.
13. M. Maggiali, G. Cannata, P. Maiolino, G. Metta, M. Randazzo, and G. Sandini, "Embedded distributed capacitive tactile sensor," *Mechatronics*, 07 2008.
14. G. Cannata, S. Denei, and F. Mastrogiovanni, "On internal models for representing tactile information," in *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pp. 1108–1113, 2010.