2021 I-RIM Conference
October 8-10, Rome, Italy
ISBN: 9788894580525
DOI:10.5281/zenodo.6367972

A Generative Model Towards Conditioned Robotic Object Manipulation

Luca Garello[*,1,2], Linda Lastrico[*,1,2], Fulvio Mastrogiovanni[1], Alessandra Sciutti[2], Nicoletta Noceti[1], Francesco Rea[2]

***Abstract*—In a collaborative scenario, robots performing communicative and legible gestures would improve the safety and the naturalness of the interaction. In this study, we introduce a novel conditional Generative Adversarial Network (cGAN) for the specific problem of generating time-series related to human manipulation of objects with different characteristics. A two-steps process involves the generation of new data in a latent features space, then their decoding to the target domain through a pre-trained decoder. Our model allows the control over specific properties of the generated output. The long-term goal of our approach is to use the synthetic time-series to control the end-effector of a robot, to produce motions as communicative and implicitly informative on the object properties as humans ones.**

***Index Terms*—Conditional Generative Adversarial Network, Time-series, Implicit Communication, Object Manipulation**

I. Introduction

When we perform a gesture, such as moving an object, our movement is primarily goal-oriented and therefore the kinematics are optimized to accomplish the movement effectively and safely. However, if we observe another person performing the same action, we can not only interpret the scene, but also extract additional information about the object that is being manipulated, for instance its weight [1], [2]. Social robots should be able to resort to the same communication channel [3], generating movements that are equally informative about the properties of the manipulated object, for example to allow the human partner to prepare for handover, granting a more fluid and dynamic interaction.

In a previous work we used multiple GANs to synthesize velocity profiles modulated by the hidden properties of the manipulated objects, such as their weight and fragility [4]. We showed that the generative approach allows to create new and consistent motion patterns, preserving the informative content of the actions. In this work, we propose to use conditional GANs (cGANs) to have a single model that can generate synthetic velocity profiles reflecting different classes of motion. Considering the carefulness that may be applied in handling an object, we trained our generative model using velocity profiles associated with moving an empty glass (Not Careful movements) or one completely filled with water, which required to be moved with care (Careful). Potentially, the conditioned approach would also allow to generate synthetic data intermediate to the two classes used to train the model. In the future, the appropriately modulated velocity profile can be used to control the robot's end-effector, thus having virtually unlimited movements that are unique, yet consistent with the object's properties and as communicative as human ones.

[*] Equal contribution {luca.garello,linda.lastrico}@iit.it
[1] DIBRIS, Università degli Studi di Genova, Genoa, Italy
[2] Istituto Italiano di Tecnologia, Genoa, Italy

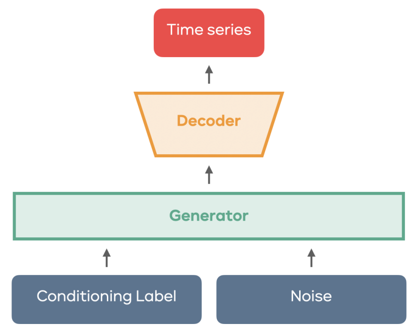

Fig. 1. **Proposed architecture:** our model takes as input random noise and a conditioning label, then the generator outputs in an embedding space that represents the characteristics of the signal we want to generate. Finally previously trained decoder translates the desired embedded properties into a time series.

II. Materials and Methods

We used data recorded in a previous experiment to train the generative model and the detailed methods of data acquisition are described in [5]. In this study we involved sixteen volunteers, whose right hand kinematics was recorded through an active infrared marker of the motion capture system Optotrak Certus®, NDI. The volunteers performed the same set of transportation movements of transparent glasses back and forth from a table in front of them, towards two shelves on their right and left, with an upper and bottom level, so that the transportation movements had a great variability. In the current work we consider the transportation of two glasses with the same weight (167 gr): one empty, the other full with water, therefore requiring a careful handling. The dataset consists of 248 transportation for the empty glass (Not Careful movement) and 254 for the glass with water (Careful movement). As in [4], to train the model, we used for each transport movement the norm of the velocity, derived from the three velocity components of a hand marker, with a sampling frequency of 22 Hz.

The goal of this work is to obtain synthetic data which preserve the temporal dynamics of the original time-series.

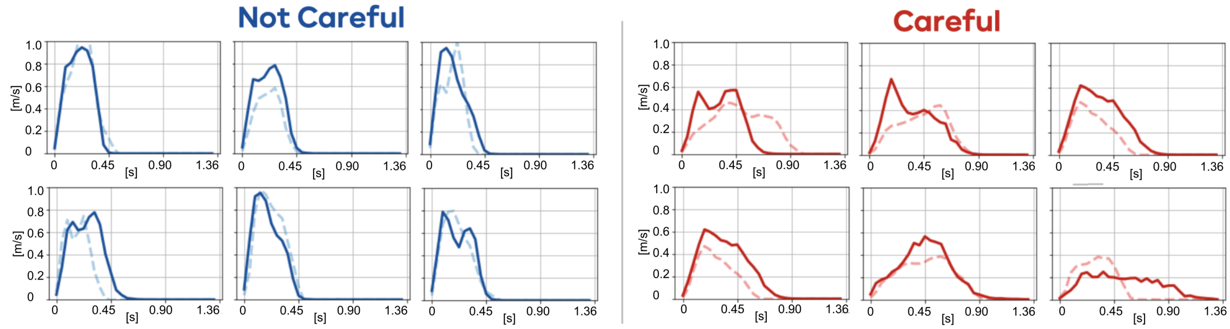

Fig. 2. **Generated velocity profiles:** we generate multiple classes of data by using different conditional inputs to a single model. The dashed lines represent random examples of the original velocity profiles from human participants, while the continuous lines are synthetic profiles for the same class of action.

However, to produce velocity profiles that reflect the two original classes C and NC, we exploit a two-steps training process. At first, we train an autoencoder to learn embedded representations of the original time-series. Then, such embeddings are used to train our cGAN, which generates new "synthetic" embeddings. The generator can be conditioned to output embeddings belonging to one of the two classes by giving as input a vector of labels and a random time-series noise (see Figure 1). Finally, a decoder is employed to reconstruct the time-series from the generated embedded representations. The main advantage of a generator which directly outputs in the embedded space is that in this way we encourage the model to focus on the relevant features of the dataset rather than on the samples themselves [6].

III. Results

The evaluation of GANs is difficult, as a qualitative analysis is often required. Figure 2 shows a comparison between the original speed profiles and those generated by our architecture. As can be seen, the synthetic data prove to be similar to the real ones. Also when projecting the data in a lower dimensional space, as it happens applying a t-SNE in Figure 3, the distributions of real and synthetic data are comparable. It should be noted how generated data are not a trivial copy of the input, but they give a coherent representation of the original classes.

IV. Conclusions

We showed how it is possible to generate different velocity profiles that are coherent with the properties of the object being manipulated, by conditioning the output of a single model. This approach could be extended in the future by training the model with a greater number of classes derived from human movements, for instance relative to the manipulation of objects with other intrinsic characteristics or to movements performed with a particular emotional attitude (rude, gentle, disgusted and so on). This would allow to generate a varied repertoire of communicative robotic motions. Moreover, the possibility to condition the network to return intermediate outputs between the two extremes used in the training phase, should be explored in detail and the outcomes quantitatively assessed through human-robot interaction experiments.

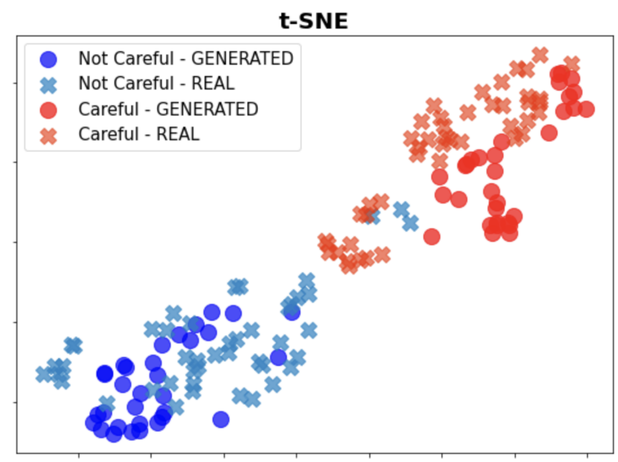

Fig. 3. **T-distributed Stochastic Neighbor Embedding of Real Samples vs Generated Samples:** both the generated classes overlap with the original distributions (Careful and Not Careful velocity profiles).

References

[1] A. Hamilton, D. Joyce, J. Flanagan, C. Frith, and D. Wolpert, "Kinematic cues in perceptual weight judgment and their origins in box lifting," *Psychological research*, vol. 71, pp. 13–21, Feb 2007.

[2] A. Sciutti, L. Patanè, F. Nori, and G. Sandini, "Understanding object weight from human and humanoid lifting actions," *Autonomous Mental Development, IEEE Transactions*, vol. 6, pp. 80–92, Jun 2014.

[3] A. D. Dragan, K. C. T. Lee, and S. S. Srinivasa, "Legibility and predictability of robot motion," in *Proceedings of the 8th ACM/IEEE International Conference on Human-Robot Interaction*, Tokyo, Japan, Mar 2013.

[4] L. Garello, L. Lastrico, F. Rea, F. Mastrogiovanni, N. Noceti, and A. Sciutti, "Property-aware robot object manipulation: a generative approach," in *2021 IEEE International Conference on Development and Learning (ICDL)*, 2021, pp. 1–7.

[5] L. Lastrico, A. Carfì, A. Vignolo, A. Sciutti, F. Mastrogiovanni, and F. Rea, "Careful with that! Observation of human movements to estimate objects properties," in *Human-Friendly Robotics 2020*, 2021, pp. 127–141.

[6] J. Yoon, D. Jarrett, and M. van der Schaar, "Time-series generative adversarial networks," in *Proceedings of 33rd Conference on Neural Information Processing Systems*, Vancouver, Canada, Oct 2019.