Affective Robot Movement Generation Using CycleGANs

Michael Suguitan
HRC[2] Lab
Cornell University
Ithaca, NY, USA
mjs679@cornell.edu

Mason Bretan
AI Center
Samsung Research America
Mountain View, CA, USA
mason.bretan@samsung.com

Guy Hoffman
HRC[2] Lab
Cornell University
Ithaca, NY, USA
hoffman@cornell.edu

***Abstract*—Social robots use gestures to express internal and affective states, but their interactive capabilities are hindered by relying on preprogrammed or hand-animated behaviors, which can be repetitive and predictable. We propose a method for automatically synthesizing affective robot movements given manually-generated examples. Our approach is based on techniques adapted from deep learning, specifically generative adversarial neural networks (GANs).**

***Index Terms*—social robot; robot movement; deep learning; generative adversarial networks**

I. Introduction

Robots designed for social interaction are becoming more common in spaces such as homes and retail environments. In these contexts, they use movements and gestures to express internal and affective states. However, these behaviors are often difficult to create and are largely preprogrammed or animated by hand, resulting in repetitiveness that can lead to diminished interest over time [1, 3].

There have been some techniques to automatically generate robot movements based on known heuristics and knowledge of the robot's embodiment [4], but these are not easily generalizable to different robots. At the same time, advancements in deep learning have enabled the creation of data-driven models for applications ranging from temporal forecasting to image modulation and generation [6]. While these deep learning methods have seen success in robotic perception, they have remained largely unadopted for generating robot behaviors.

We propose a method to automatically generate robot movements that express affect by using movement exemplars and techniques from deep learning, specifically generative adversarial networks (GANs) [2]. We use a cycle-consistent GAN (CycleGAN) [8] to simultaneously generate robot movements from human movements and evaluate the mapping by using examples of hand-coded robot movements.

II. Implementation

The CycleGAN architecture shown in Figure 1 uses a pair of GANs in a cyclic configuration to perform translation between movement spaces. We use multiple binary CycleGANs, one for each emotion class, due to the difficulty we encountered in creating a single CycleGAN that can generalize across multiple emotion classes.

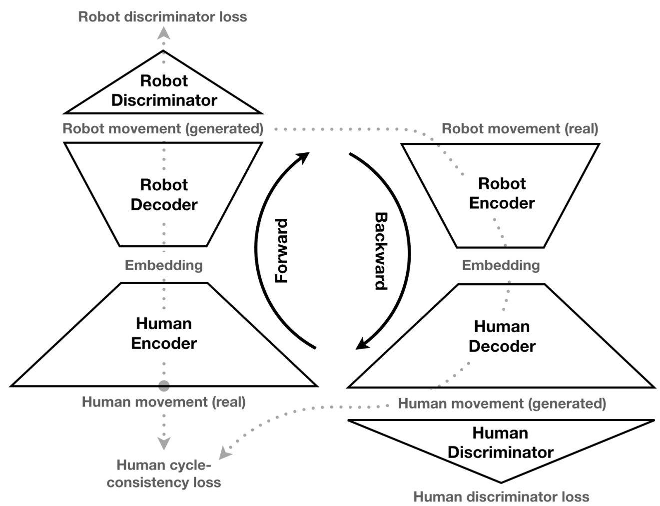

Fig. 1. CycleGAN architecture for automatic affective gesture generation. The path for the forward cycle (human→robot) is outlined by the dotted line. A real human movement sample is passed through the forward generator (human encoder and robot decoder) and generates a corresponding robot movement. This generated robot movement is passed through both the robot discriminator which provides the robot discriminator loss and the backward generator (robot encoder and human decoder) to reconstruct the human movement input which provides the human cycle-consistency loss. The network also performs the backward cycle (human←robot) in parallel.

The inputs to the network are robot and human movements that share common emotion class labels. We currently use three classes: happy, sad, and angry. The robot dataset consists of approximately 10 movement samples per emotion class. The human dataset was sourced from 30 non-professional actors and consists of over 4,000 movement samples of various emotive actions [5].

The objective is for the CycleGAN to use a source human movement and generate a robot movement which matches the affective label of the human movement. This is measured by how convincingly real the generated movements are while ensuring that the network is able to reconstruct the input movements. For the forward cycle (human→robot, outlined in Figure 1 with the dotted line), the forward generator (human encoder and robot decoder) uses a source human movement to generate a robot movement. This generated movement is

978-1-5386-8555-6/19/$31.00 ©2019 IEEE

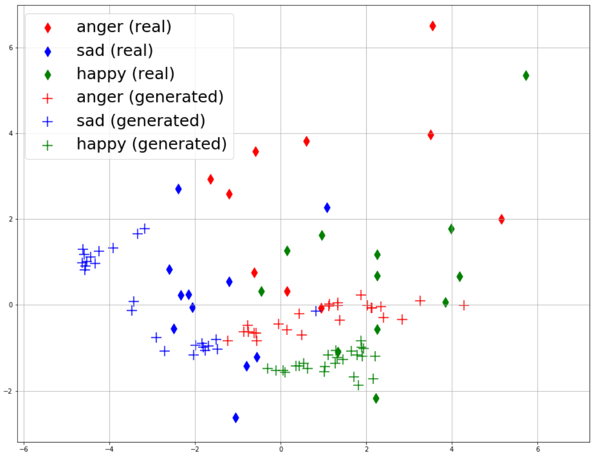

Fig. 2. Comparison of real and generated robot movements. Kinematic features (speed and range for each DOF, posture, height, and yaw-pitch-roll, resulting in a 13-dimension feature vector) for each movement sample are calculated and plotted along their two principal components.

then passed through a discriminator to determine if it is real or fake. This discriminator computes a loss function $\mathcal{L}_{D,fwd}$. The generated movement is also passed through the backward generator to reconstruct the human movement (computing a human cycle-consistency loss, $\mathcal{L}_{CC,fwd}$). $\mathcal{L}_{CC,fwd}$ is weighed by a factor λ_{fwd} and added to $\mathcal{L}_{D,fwd}$ to define the combined forward cycle loss $\mathcal{L}_{fwd} = \mathcal{L}_{D,fwd} + \lambda_{fwd}\mathcal{L}_{CC,fwd}$. The backward cycle (human←robot) is trained concurrently using a similarly structured loss $\mathcal{L}_{bwd}$, and the forward and backward cycle losses are combined to define the entire network loss $\mathcal{L}$.

$$\begin{aligned}\mathcal{L} &= \mathcal{L}_{fwd} + \mathcal{L}_{bwd} \\ &= \mathcal{L}_{D,fwd} + \lambda_{fwd}\mathcal{L}_{CC,fwd} + \mathcal{L}_{D,bwd} + \lambda_{bwd}\mathcal{L}_{CC,bwd}\end{aligned}$$

The network is trained with categorical cross-entropy and mean squared error loss functions for the discriminator and cycle-consistency losses, respectively. After training, robot movements are generated by passing human movement samples through the forward generator.

III. Results and Future Work

To evaluate this approach, we used a Blossom robot [7], an open-source social robot which features 4 degrees-of-freedom (DOFs) through its head platform: pitch, roll, yaw, and vertical translation. Blossom has few DOFs compared to other robots but is still capable of expressive movements, making it suitable as a testbed for this method.

The affective quality of the generated robot movements can be objectively evaluated through heuristic kinematic measures such as movement speed and range [4]. As shown in the principal component analysis of these features in Figure 2, the generated happy and sad movements cluster fairly well along the real movements, but the generated angry movements are clustered more closely towards happy than anger. Notably, the real angry gestures were also widely dispersed.

Perceptual metrics can be used to subjectively evaluate the generated movements. As shown in Figure 3, the intended

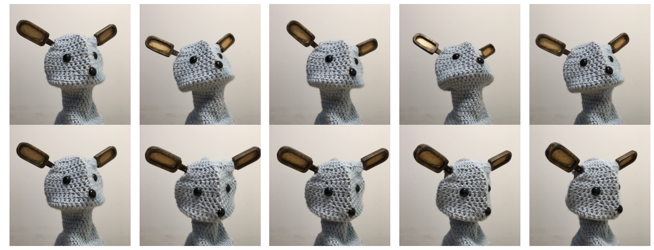

Fig. 3. Sequences of generated happy (top) and sad (bottom) movements.

emotions are fairly discernible between classes. We plan to evaluate the results through a user study that compares the real and generated movements in terms of naturalness and how well the intended emotion is conveyed.

Future plans for this work are focused on improving the network implementation and generalizability. Rather than an ensemble of binary CycleGANs with labeled data, the ideal implementation would be a single CycleGAN that could translate with either multiclass or unlabeled data. We also plan to explore different modalities for translation. For example, a given musical sequence could be used to generate a robot movement, while a user-generated robot movement could be used to create an appropriate musical cue. Finally, this method can be evaluated on other robots with different embodiments.

References

[1] Rachel Gockley et al. "Designing robots for long-term social interaction". In: *Intelligent Robots and Systems, IROS*. IEEE. (2005), pp. 1338–1343.

[2] Ian Goodfellow et al. "Generative adversarial nets". In: *Advances in neural information processing systems*. (2014), pp. 2672–2680.

[3] Takayuki Kanda et al. "Interactive robots as social partners and peer tutors for children: A field trial". In: *Human–Computer Interaction* 19.1-2 (2004), pp. 61–84.

[4] Michelle Karg et al. "Body movements for affective expression: A survey of automatic recognition and generation". In: *IEEE Transactions on Affective Computing* 4.4 (2013), pp. 341–359.

[5] Yingliang Ma, Helena M Paterson, and Frank E Pollick. "A motion capture library for the study of identity, gender, and emotion perception from biological motion". In: *Behavior research methods* 38.1 (2006), pp. 134–141.

[6] Jürgen Schmidhuber. "Deep learning in neural networks: An overview". In: *Neural networks* 61 (2015), pp. 85–117.

[7] Michael Suguitan and Guy Hoffman. "Blossom: A Tensile Social Robot Design with a Handcrafted Shell". In: *Companion of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM. (2018), pp. 383–383.

[8] Jun-Yan Zhu et al. "Unpaired image-to-image translation using cycle-consistent adversarial networks". In: *arXiv preprint* (2017).