MoVEInt: Mixture of Variational Experts for Learning HRI from Demonstrations

Vignesh Prasad, Alap Kshirsagar, Dorothea Koert, Ruth Stock-Homburg, Jan Peters, Georgia Chalvatzaki
Technical University of Darmstadt

***Abstract*—In this work, we propose a novel approach for learning a shared latent space representation for HRIs from demonstrations in a Mixture of Experts fashion for reactively generating robot actions from human observations. We train a Variational Autoencoder (VAE) to learn robot motions regularized using an informative latent space prior that captures the multimodality of the human observations via a Mixture Density Network (MDN). We show how our formulation derives from a Gaussian Mixture Regression formulation that is typically used approaches for learning HRI from demonstrations such as using an HMM/GMM for learning a joint distribution over the actions of the human and the robot. We further incorporate an additional regularization to prevent "mode collapse", a common phenomenon when using latent space mixture models with VAEs. We find that our approach of using an informative MDN prior from human observations for a VAE generates more accurate robot motions compared to previous HMM-based or recurrent approaches of learning shared latent representations, which we validate on various HRI datasets involving interactions such as handshakes, fistbumps, waving, and handovers. For further information, code, and videos, please visit https://bit.ly/MoVEInt.**

I. Introduction

Ensuring a timely response to an interaction can enable a feeling of connectedness to a partner [25] making it an important aspect of Human-Robot Interaction (HRI). One way to do so is by learning a shared representation space between the human and the robot [7, 17, 18, 29, 30, 38, 39]. An important aspect of such approaches, for learning HRI from demonstrations, is accurately capturing the multimodality of the underlying data to effectively learn the various underlying skills and generate accurate response motions for them.

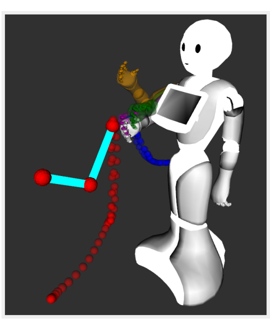


Fig. 1: Example of multiple policies for a handshake interaction generated by MoVEInt based on human observations combined to generate suitable robot motions.

In our work, rather than learning just a single task, we further explore how underlying latent strategies can be learned from different tasks in a dataset by using a mixture distribution to predict different latent policies, which are then combined in a Mixture of Experts fashion. An example of this can be seen in Fig. 1, where we show a handshake interaction with the Pepper robot. Trained on a dataset of different physical interactions like waving, handshakes, and fistbumps, we see the different policies that get predicted (shown by the different colored arms) which are then combined in the latent space yielding a suitable response motion (shown in white).

To learn multiple latent policies and effectively combine them, we employ Mixture Density Networks (MDNs) [6] to capture the multimodality of the demonstrations. MDNs predict a mixture of Gaussians and the corresponding mixture coefficients yielding a multimodal prediction, rather than a unimodal distribution or a single output.

In this paper, we propose "MoVEInt", a novel framework that employs a Mixture of Variational Experts for learning Human-Robot Interactions from demonstrations through a shared latent representation of a human and a robot. We learn latent space policies in a Mixture of Experts fashion via a Mixture Density Network (MDN) to encode the latent trajectory of a human partner, regularize the robot embeddings, and subsequently, predict the robot motions reactively.

Through our experiments, we see that our approach successfully captures the best of recurrent, multi-modal, and reactive representations for learning short-horizon Human-Robot Interactions from demonstrations. We find that MoVEInt generates highly accurate robot behaviors without explicit action labels, as done in previous works, which is more natural as humans also internally infer what our interaction partner is doing and adapt to it without explicitly communicating the action being done. We validate our predictive performance on a variety of physical HRI scenarios such as handshakes, fistbumps, and robot-to-human handovers. We further demonstrate the efficacy of MoVEInt in a real-world interaction scenario for bimanual (dual-arm) robot-to-human handovers.

II. Mixture of Variational Experts for Learning Human-Robot Interactions from Demonstrations

In this section, we present MoVEInt, a novel framework that learns latent space policies in a "Mixture of Experts" fashion for modeling the shared dynamics of a human and a robot in HRI tasks. This process can be seen in Fig. 2. We aim to model the dynamics of HRI tasks via shared latent representations of a human and a robot in a way that captures the multimodality of the demonstrations and subsequently predicts the robot's motions in a reactive manner. To do so, we use an MDN that takes the human observations as input and predicts multiple latent policies, thereby enabling a multimodal

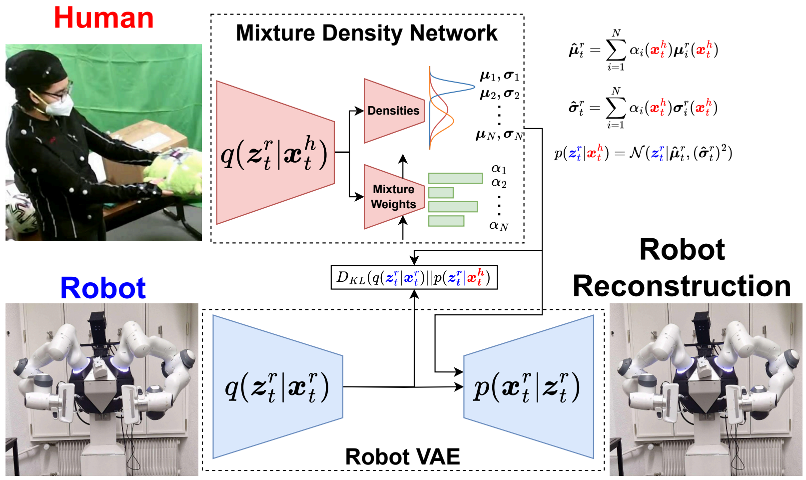


Fig. 2: Overview of our approach "MoVEInt".

output, and subsequently, the relative weights for each policy so that they can be effectively combined. For learning a shared latent representation between the human and the robot, we train a VAE over the robot motions and regularize the VAE with the predicted policy from the MDN, thereby learning the robot embeddings and the subsequent human-conditioned policy predictions in a cohesive manner.

We denote the human variables in red with the superscript h and the robot variables in blue with the superscript r.

A. Learning Interaction Dynamics with MDNs

Learning a joint distribution over the degrees of freedom of a human and a robot has been widely used in learning HRI from demonstrations [8, 9, 11, 12, 24, 29, 30]. With a joint distribution, Gaussian Mixture Regression (GMR) provides a mathematically sound formulation of predicting the conditional distribution of the robot actions. When using a Mixture of N Gaussian components $\{\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\}$ that model a joint distribution of the Human and Robot trajectories, the distribution can be decomposed into the marginals for the human and the robot

$$\boldsymbol{\mu}_i = \begin{bmatrix} \boldsymbol{\mu}_i^h \\ \boldsymbol{\mu}_i^r \end{bmatrix}; \boldsymbol{\Sigma}_i = \begin{bmatrix} \boldsymbol{\Sigma}_i^{hh} & \boldsymbol{\Sigma}_i^{hr} \\ \boldsymbol{\Sigma}_i^{rh} & \boldsymbol{\Sigma}_i^{rr} \end{bmatrix} \tag{1}$$

To consider the temporal aspect of learning such trajectories from demonstrations, the Mixture Model coefficients can be calculated using Hidden Markov Models (HMMs) [8]. Given that there exist parallels between HMMs and RNNs [4, 10, 16], we use a recurrent layer for predicting the mixing coefficients. Using the predictions of the mixture model parameters $(\boldsymbol{\mu}_i^r(\boldsymbol{x}_t^h), \boldsymbol{\sigma}_i^r(\boldsymbol{x}_t^h)^2)$ and coefficients $\alpha_i(\boldsymbol{x}_t^h)$ from the MDN, which is then used as a prior $p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h)$ for regularizing the VAE and subsequently training the decoder to reconstruct latent samples obtained after observing the human partner

$$\begin{aligned} \hat{\boldsymbol{\mu}}_t^r &= \sum_{i=1}^{N} \alpha_i(\boldsymbol{x}_t^h)\boldsymbol{\mu}_i^r(\boldsymbol{x}_t^h) \\ (\hat{\boldsymbol{\sigma}}_t^r)^2 &= \sum_{i=1}^{N} \alpha_i(\boldsymbol{x}_t^h)\boldsymbol{\sigma}_i^r(\boldsymbol{x}_t^h)^2 \\ p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h) &= \mathcal{N}(\boldsymbol{z}_t^r|\hat{\boldsymbol{\mu}}_t^r, (\hat{\boldsymbol{\sigma}}_t^r)^2) \end{aligned} \tag{2}$$

where $\boldsymbol{z}_t^r$ denotes the latent space of the robot.

B. Learning Robot Motion Embeddings for Reactive Motion Generation

To learn a meaningful representation of the robot's actions, we train a VAE to reconstruct the robot's actions at each timestep. Typically, in VAEs, a standard normal distribution is used as the latent space prior $p(\boldsymbol{z}) = \mathcal{N}(\mathbf{0}, \boldsymbol{I})$. Rather than forcing an uninformative standard normal prior as in Eq. 11, we use the reactive policy predicted from the human observations by the MDN (Eq. 7) to regularize the VAE's posterior $KL(q(\boldsymbol{z}_t^r|\boldsymbol{x}_t^r)||p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h))$, thereby learning a task-oriented latent space. Our ELBO can then be written as

$$\begin{aligned} ELBO_t^r = \; &\mathbb{E}_{q(\boldsymbol{z}_t^r|\boldsymbol{x}_t^r)}[\log p(\boldsymbol{x}_t^r|\boldsymbol{z}_t^r)] \\ &- \beta KL(q(\boldsymbol{z}_t^r|\boldsymbol{x}_t^r)||p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h)) \end{aligned} \tag{3}$$

where β is a relative weight used to ensure numerical stability between the KL divergence term and the image reconstruction term [15].

We aim to learn a policy for reactively generating the robot's latent trajectory based on human observations $p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h)$. We do so in a Behavior Cloning Paradigm by maximizing the probability of the observed trajectories w.r.t. the current policy $\mathcal{L}_t^{BC} = -\mathbb{E}_{\boldsymbol{z}_t^r \sim p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h)} p(\boldsymbol{x}_t^r|\boldsymbol{z}_t^r)$ wherein we first draw samples from the current policy $\boldsymbol{z}_t^r \sim p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h)$ which we then reconstruct $p(\boldsymbol{x}_t^r|\boldsymbol{z}_t^r)$, thereby enabling the decoder to reconstruct latent samples obtained after observing the human, as done during test time.

However, as highlighted in [41], MDN policy representations are prone to mode collapses. Therefore, to ensure

Dataset (units)	Action	MILD [30]	Bütepage et al. [7]	MoVEInt
HHI (Bütepage et al. [7]) (cm)	Hand Wave	0.788 ± 1.226	4.121 ± 2.252	**0.448 ± 0.630**
	Handshake	1.654 ± 1.549	1.181 ± 0.859	**0.196 ± 0.153**
	Rocket Fistbump	0.370 ± 0.682	0.544 ± 1.249	**0.123 ± 0.175**
	Parachute Fistbump	0.537 ± 0.579	0.977 ± 1.141	**0.314 ± 0.348**
HRI-Pepper (Bütepage et al. [7]) (rad)	Hand Wave	0.103 ± 0.103	0.664 ± 0.277	**0.087 ± 0.089**
	Handshake	0.056 ± 0.041	0.184 ± 0.141	**0.015 ± 0.014**
	Rocket Fistbump	0.018 ± 0.035	0.033 ± 0.045	**0.007 ± 0.015**
	Parachute Fistbump	0.088 ± 0.148	0.189 ± 0.196	**0.048 ± 0.112**
HRI-Yumi (Bütepage et al. [7]) (rad)	Hand Wave	1.033 ± 1.204	0.225 ± 0.302	**0.147 ± 0.072**
	Handshake	0.068 ± 0.052	0.133 ± 0.214	**0.057 ± 0.044**
	Rocket Fistbump	0.128 ± 0.071	0.147 ± 0.119	**0.093 ± 0.045**
	Parachute Fistbump	**0.028 ± 0.034**	0.181 ± 0.155	0.081 ± 0.082
HHI (NuiSI [30]) (cm)	Hand Wave	0.408 ± 0.538	3.168 ± 3.392	**0.298 ± 0.274**
	Handshake	0.311 ± 0.259	1.489 ± 3.327	**0.149 ± 0.120**
	Rocket Fistbump	1.142 ± 1.375	3.576 ± 3.082	**0.673 ± 0.679**
	Parachute Fistbump	0.453 ± 0.578	2.008 ± 2.024	**0.291 ± 0.199**
HRI-Pepper (NuiSI [30]) (rad)	Hand Wave	0.046 ± 0.059	0.057 ± 0.093	**0.044 ± 0.048**
	Handshake	0.020 ± 0.014	0.083 ± 0.075	**0.011 ± 0.008**
	Rocket Fistbump	0.077 ± 0.067	0.101 ± 0.086	**0.045 ± 0.045**
	Parachute Fistbump	0.022 ± 0.027	0.049 ± 0.040	**0.017 ± 0.014**
HHI-Handovers (Kshirsagar et al. [21]) (cm)	Unimanual	0.441 ± 0.280	1.133 ± 0.721	**0.441 ± 0.221**
	Bimanual	0.869 ± 0.964	0.990 ± 0.764	**0.685 ± 0.643**

TABLE I: Prediction MSE for robot trajectories after observing the human partner averaged over all joints and timesteps. Results for the HHI scenarios are in cm and for the HRI scenarios are in radians. (Lower is better)

adequate separation between the modes so that we can learn a diverse range of actions, we employ a contrastive loss at each timestep. The contrastive loss pushes the means of each mixture component further away, while maintaining temporal similarity by pushing embeddings that are closer in time nearer to each other. Further, as done in [41], we add entropy cost to ensure a balanced prediction of the mixture coefficients. Our separation loss can be written as

$$\mathcal{L}_t^{sep} = \underbrace{\sum_{i=1}^{N-1}\sum_{j=i+1}^{N} exp(-\|\boldsymbol{\mu}_i^r(\boldsymbol{x}_t^h) - \boldsymbol{\mu}_j^r(\boldsymbol{x}_t^h)\|^2)}_{\text{separation of means}}$$
$$+ \underbrace{1 - \frac{1}{N}\sum_{i=1}^{N} exp(-\|\boldsymbol{\mu}_i^r(\boldsymbol{x}_t^h) - \boldsymbol{\mu}_i^r(\boldsymbol{x}_{t-1}^h)\|^2)}_{\text{temporal closeness}} \quad (4)$$
$$+ \underbrace{\sum_{i=1}^{N} \alpha_i(\boldsymbol{x}_t^h)\ln\alpha_i(\boldsymbol{x}_t^h)}_{\textit{entropy cost}}$$

Our final loss consists of the Behavior Cloning loss, the ELBO of the robot VAE, and the separation loss $\sum_{t=1}^{T}\left[\mathcal{L}_t^{BC} - ELBO_t^r + \beta\mathcal{L}_t^{sep}\right]$ where β is the same KL weight factor used in Eq. 3.

During test time, given human observations $\boldsymbol{x}_t^h$, we compute the latent policy from the MDN $\boldsymbol{z}_t^r = p(\boldsymbol{z}_t^r|\boldsymbol{x}_t^h)$ which is then decoded to obtain the robot action $p(\boldsymbol{x}_t^r|\boldsymbol{z}_t^r)$.

III. Experiments and Results

A. Reactive Motion Generation Results

We compare MoVEInt with Bütepage et al. [7], who use an LSTM-based approach as a latent regularization for reactive motion generation, which is close to a unimodal version of our approach. Further, we compare MoVEInt to MILD [30] which uses HMMs to capture the multimodal latent dynamics of interactive tasks. The efficacy of MoVEInt can be seen via the low error of the predicted robot motions (Table I).

We perform better than both MILD [30] and Bütepage et al. [7] on almost all interaction scenarios. We additionally want to highlight that on the robot scenarios, unlike MILD [30] and Bütepage et al. [7] where the pre-trained model from the Human-Human scenario is used, we train our model completely from scratch and still achieve better performance. Moreover, it is worth noting that both [7] and MILD are trained in a partially supervised manner using the interaction labels. In [7], a one-hot label denoting the interaction being performed is given as an input to the network for generalizing to different interactions, whereas in MILD, a separate HMM is trained for each interaction. In contrast, MoVEInt is trained on all the tasks in a given dataset without any labels in a purely unsupervised manner and still achieves competitive results on the different datasets.

We additionally show some qualitative results of MoVEInt. We train a Handover model with just the hand trajectories whose predictions are used for reactive motion generation on a Bimanual Franka Emika Panda robot setup, "Kobo", as shown in Fig. 3. Some additional examples of the trajectories generated by MoVEInt for bimanual and unimanual handovers from the HHI-Handovers dataset [21] are shown in Fig. 4a and 4b respectively. Since MoVEInt is trained on all the interactions in a corresponding dataset, which, when coupled with the separation loss, learns a diverse and widespread set of components that cover the various demonstrations, as can be seen by the reconstruction of the individual components. Combining the components in the latent space subsequently leads to an accurate and suitable motion (shown in blue).

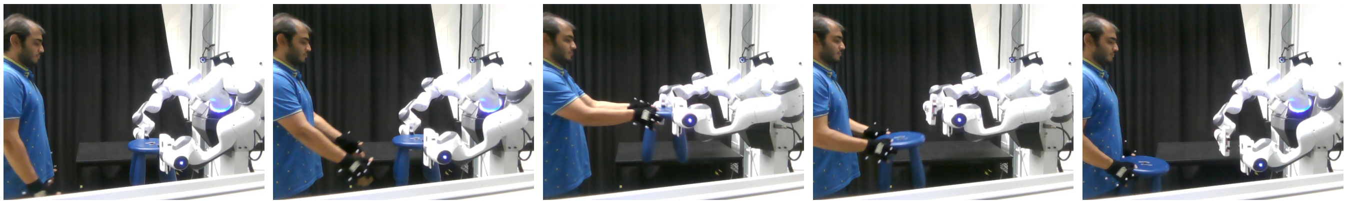
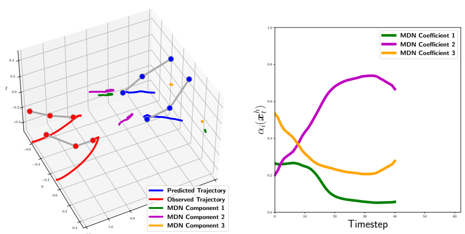
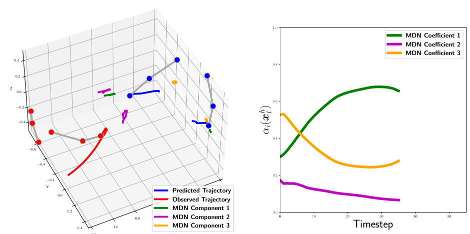


Fig. 3: Sample Human-Robot Interactions generated with the reactive motions generated by MoVEInt for a Bimanual Handover scenario.



(a) Example of a generated Bimanual Handovers



(b) Example of a generated Unimanual Handovers

Fig. 4: Sample trajectories generated by MoVEInt for the Bimanual and Unimanual Handovers in the HHI-Handovers dataset in [21]. The 3D plots show the reconstructed trajectories and the 2D plots show the corresponding progression of $\alpha_i(\boldsymbol{x}_t^h)$ for the different components of the MDN. In the 3D plots, the observed trajectory of the receiver is shown in red and the generated trajectory of the giver is shown in blue and the giver's corresponding ground truth is shown in black. The reconstruction of the individual latent components of the MDN are shown in green, magenta, and orange. It can be seen that the learned components correspond to different parts of the task space. For example, green denotes the hand locations for a unimanual handover, magenta denotes the hand locations for a bimanual handover, and orange denotes the static hand locations for the starting and ending neutral poses. In the 2D plot, it can be seen how the coefficients for components corresponding to bimanual (magenta) and unimanual (green) get activated based on the interaction being performed, while the component corresponding to a neutral pose (orange) gets activated at the beginning of the interaction while both partners are static.

B. User study

To study the effectiveness of MoVEInt in the real world, we perform a feasibility study as a proof-of-concept with five users who perform bimanual robot-to-human handovers with the Kobo robot. We evaluate the ability of MoVEInt to successfully generate a handover motion with three different objects where each participant interacts with the robot five times for each object (a total of 15 runs per participant). To maintain the object-centric nature of the interaction, we use a controller that tracks the mid-point of both end-effectors, thereby resembling tracking the object's trajectory.

As shown in Table II, our approach can generate successful handover trajectories for different users and different objects. We observed some failure cases due to sudden jumps in the predicted robot motions resulting from inaccuracies in perceiving the human, which would overshoot the robot's dynamic limits. However, this failure could be avoided by incorporating additional filters over the input and output data to MoVEInt. Some failures occurred because the object did not reach the exact vicinity of the human's hand location. This failure could be avoided by incorporating object-related information such as the size or weight, allowing the robot to gauge better when the handover is executed. Sometimes, the robot would retreat before the human could grasp the object if sufficient time had passed. One reason for this hasty retreating behavior could be that the recurrent network's hidden inputs overpower the observational input, causing the robot to follow the general motion of the handover seen during training and retreat. Such a failure could be mitigated by incorporating the robot state as part of the input.

User ID / Object	#1	#2	#3	#4	Total (per object)
Stool	5	4	5	5	19/20
Box	4	5	4	4	17/20
Bedsheet	4	5	3	3	15/20
Total (per user)	13/15	14/15	12/15	12/15	**51/60**

TABLE II: Number of successful handovers of each object by the Kobo robot to each user (total of 5 per object per user i.e. total of 20 per object and 15 per user).

IV. Conclusion and Future Work

In this work, we presented "MoVEInt", a novel deep generative Imitation Learning approach for learning Human-Robot Interaction from demonstrations in a Mixture of Experts fashion. We demonstrated the use of Mixture Density Networks (MDNs) as a multimodal policy representation in a shared latent space of the human and the robot. We showed how MoVEInt stems from the GMR-based formulation of predicting interaction dynamics used in HMM-based approaches to learning HRI. We showed how our MDN policy can predict multiple underlying policies and combine them to effectively generate response motions for the robot. We verified the efficacy of MoVEInt across a variety of interactive tasks, where we found that MoVEInt mostly outperformed other baselines that either use explicitly modular representations like an HMM or simple recurrent policy representations. Our evaluations showcases the versatility of MoVEInt, which effectively combines explicitly modular distributions with recurrent policy representations for learning interaction dynamics.

References

[1] 3DiVi. Nuitrack. URL https://nuitrack.com/.

[2] Pooya Abolghasemi, Amir Mazaheri, Mubarak Shah, and Ladislau Boloni. Pay attention!-robustifying a deep visuomotor policy through task-focused visual attention. In *IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 2019.

[3] Ali Baheri. Safe reinforcement learning with mixture density network, with application to autonomous driving. *Results in Control and Optimization*, 2022.

[4] Matt Baucum, Anahita Khojandi, and Theodore Papamarkou. Hidden markov models as recurrent neural networks: An application to alzheimer's disease. In *IEEE International Conference on Bioinformatics and Bioengineering (BIBE)*, 2021.

[5] Loris Bazzani, Hugo Larochelle, and Lorenzo Torresani. Recurrent mixture density network for spatiotemporal visual attention. In *International Conference on Learning Representations (ICLR)*, 2016.

[6] Christopher M Bishop. Mixture density networks. 1994.

[7] Judith Bütepage, Ali Ghadirzadeh, Özge Öztimur Karadağ, Mårten Björkman, and Danica Kragic. Imitating by generating: Deep generative models for imitation of interactive tasks. *Frontiers in Robotics and AI*, 2020.

[8] Sylvain Calinon. A tutorial on task-parameterized movement learning and retrieval. *Intelligent service robotics*, 2016.

[9] Sylvain Calinon, Paul Evrard, Elena Gribovskaya, Aude Billard, and Abderrahmane Kheddar. Learning collaborative manipulation tasks by demonstration using a haptic interface. In *International Conference on Advanced Robotics (ICAR)*, 2009.

[10] Justin Chiu and Alexander M Rush. Scaling hidden markov language models. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*, 2020.

[11] Paul Evrard, Elena Gribovskaya, Sylvain Calinon, Aude Billard, and Abderrahmane Kheddar. Teaching physical collaborative tasks: object-lifting case study with a humanoid. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2009.

[12] Marco Ewerton, Gerhard Neumann, Rudolf Lioutikov, Heni Ben Amor, Jan Peters, and Guilherme Maeda. Learning multiple collaborative tasks with a mixture of interaction primitives. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2015.

[13] Lars Fritsche, Felix Unverzag, Jan Peters, and Roberto Calandra. First-person tele-operation of a humanoid robot. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2015.

[14] David Ha and Jürgen Schmidhuber. Recurrent world models facilitate policy evolution. *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[15] Irina Higgins, Loic Matthey, Arka Pal, Christopher Burgess, Xavier Glorot, Matthew Botvinick, Shakir Mohamed, and Alexander Lerchner. beta-vae: Learning basic visual concepts with a constrained variational framework. In *International Conference on Learning Representations (ICLR)*, 2016.

[16] Tatsuya Hiraoka, Sho Takase, Kei Uchiumi, Atsushi Keyaki, and Naoaki Okazaki. Recurrent neural hidden markov model for high-order transition. *Transactions on Asian and Low-Resource Language Information Processing*, 2021.

[17] Ananth Jonnavittula and Dylan P Losey. Learning to share autonomy across repeated interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2021.

[18] Ravi Prakash Joshi, Jayant Prasad Tarapure, and Tomohiro Shibata. Electric wheelchair-humanoid robot collaboration for clothing assistance of the elderly. In *IEEE International Conference on Human System Interaction (HSI)*, 2020.

[19] Heecheol Kim, Yoshiyuki Ohmura, and Yasuo Kuniyoshi. Memory-based gaze prediction in deep imitation learning for robot manipulation. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2022.

[20] Diederik P Kingma and Max Welling. Auto-encoding variational bayes. In *International Conference on Learning Representations (ICLR)*, 2014.

[21] Alap Kshirsagar, Raphael Fortuna, Zhiming Xie, and Guy Hoffman. Dataset of bimanual human-to-human object handovers. *Data in Brief*, 2023.

[22] Sampo Kuutti, Saber Fallah, and Richard Bowden. Adversarial mixture density networks: Learning to drive safely from collision data. In *IEEE International Intelligent Transportation Systems Conference (ITSC)*, 2021.

[23] Kyungjae Lee, Sungjoon Choi, and Songhwai Oh. Maximum causal tsallis entropy imitation learning. *Advances in Neural Information Processing Systems (NeurIPS)*, 2018.

[24] Guilherme Maeda, Marco Ewerton, Rudolf Lioutikov, Heni Ben Amor, Jan Peters, and Gerhard Neumann. Learning interaction for collaborative tasks with probabilistic movement primitives. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2014.

[25] Kerry L Marsh, Michael J Richardson, and Richard C Schmidt. Social connection through joint action and interpersonal coordination. *Topics in cognitive science*, 1(2):320–339, 2009.

[26] Geoffrey J McLachlan and Kaye E Basford. *Mixture models: Inference and applications to clustering*. M. Dekker New York, 1988.

[27] Amit Kumar Pandey and Rodolphe Gelin. A mass-produced sociable humanoid robot: Pepper: The first machine of its kind. *IEEE Robotics & Automation Magazine*, 2018.

[28] Vignesh Prasad, Ruth Stock-Homburg, and Jan Peters. Learning human-like hand reaching for human-robot handshaking. In *2021 IEEE International Conference*

on Robotics and Automation (ICRA), 2021.

[29] Vignesh Prasad, Dorothea Koert, Ruth Stock-Homburg, Jan Peters, and Georgia Chalvatzaki. Mild: Multimodal interactive latent dynamics for learning human-robot interaction. In *IEEE-RAS International Conference on Humanoid Robots (Humanoids)*, 2022.

[30] Vignesh Prasad, Lea Heitlinger, Dorothea Koert, Ruth Stock-Homburg, Jan Peters, and Georgia Chalvatzaki. Learning multimodal latent dynamics for human-robot interaction. *arXiv preprint arXiv:2311.16380*, 2023.

[31] Rouhollah Rahmatizadeh, Pooya Abolghasemi, Aman Behal, and Ladislau Bölöni. From virtual demonstration to real-world manipulation using lstm and mdn. In *AAAI Conference on Artificial Intelligence (AAAI)*, 2018.

[32] Danilo Jimenez Rezende, Shakir Mohamed, and Daan Wierstra. Stochastic backpropagation and approximate inference in deep generative models. In *International Conference on Machine Learning (ICML)*, 2014.

[33] Oliver C Schrempf and Uwe D Hanebeck. A generic model for estimating user intentions in human-robot cooperation. In *International Conference on Informatics in Control, Automation and Robotics*. SCITEPRESS, 2005.

[34] Prasanth Sengadu Suresh, Yikang Gui, and Prashant Doshi. Dec-airl: Decentralized adversarial irl for human-robot teaming. In *International Conference on Autonomous Agents and Multiagent Systems (AAMAS)*, 2023.

[35] Hsi Guang Sung. *Gaussian Mixture Regression and Classification*. PhD thesis, RICE UNIVERSITY, 2004.

[36] Mohammad Thabet, Massimiliano Patacchiola, and Angelo Cangelosi. Sample-efficient deep reinforcement learning with imaginary rollouts for human-robot interaction. In *IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2019.

[37] Adrián Tormos Llorente, Víctor Giménez Ábalos, Marc Domènech Vila, Dmitry Gnatyshak, Sergio Álvarez Napagao, and Javier Vázquez Salceda. Explainable agents adapt to human behaviour. In *International Workshop on Citizen-Centric Multi-Agent Systems (CMAS)*, 2023.

[38] David Vogt, Simon Stepputtis, Steve Grehl, Bernhard Jung, and Heni Ben Amor. A system for learning continuous human-robot interactions from human-human demonstrations. In *IEEE International Conference on Robotics and Automation (ICRA)*, 2017.

[39] Chen Wang, Claudia Pérez-D'Arpino, Danfei Xu, Li Fei-Fei, Karen Liu, and Silvio Savarese. Co-gail: Learning diverse strategies for human-robot collaboration. In *Conference on Robot Learning (CoRL)*, 2022.

[40] Hejia Zhang, Eric Heiden, Stefanos Nikolaidis, Joseph J Lim, and Gaurav S Sukhatme. Auto-conditioned recurrent mixture density networks for learning generalizable robot skills. *arXiv preprint arXiv:1810.00146*, 2018.

[41] You Zhou, Jianfeng Gao, and Tamim Asfour. Movement primitive learning and generalization: Using mixture density networks. *IEEE Robotics & Automation Magazine*, 2020.