

# Social NCE: Contrastive Learning of Socially-aware Motion Representations

Yuejiang Liu, Qi Yan, Alexandre Alahi  
 École Polytechnique Fédérale de Lausanne (EPFL)  
 {firstname.lastname}@epfl.ch

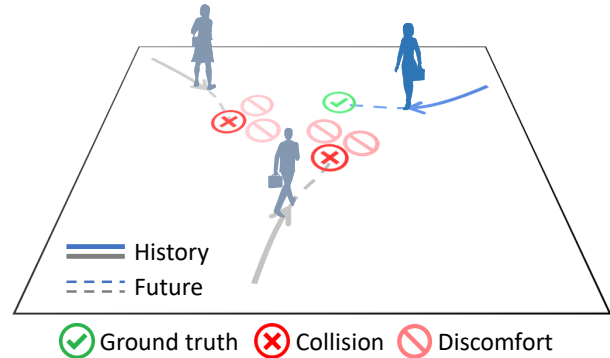
## Abstract

Learning socially-aware motion representations is at the core of recent advances in human trajectory forecasting and robot navigation in crowded spaces. Despite promising progress, existing neural motion models often struggle to generalize in closed-loop operations (e.g., output colliding trajectories), when the training set lacks examples collected from dangerous scenarios. In this work, we propose to address this issue via contrastive learning with negative data augmentation. Concretely, we introduce a social contrastive loss that encourages the encoded motion representation to preserve sufficient information for distinguishing a positive future event from a set of negative ones. We explicitly draw these negative samples based on our domain knowledge of unfavorable circumstances in the multi-agent context. Experimental results show that the proposed method dramatically reduces the collision rates of recent trajectory forecasting, behavioral cloning and reinforcement learning algorithms, outperforming current state-of-the-art models on several benchmarks. Our method makes few assumptions about neural architecture designs, and hence can be used as a generic way to promote the robustness of neural motion models.

## 1. Introduction

Humans have an instinctive ability to anticipate the future motions of other people while navigating in crowded spaces. This ability allows us to not only keep a comfortable distance from the others but also identify potential dangers or discomforts ahead of time. However, building predictive models capable of doing so is challenging. Recent works have proposed a plethora of neural network-based models [6, 21, 34, 36, 41, 50, 76, 83, 88] to learn socially-aware motion representations and demonstrated their potential for human trajectory forecasting [3, 27, 47, 77] or autonomous robot navigation [12, 14, 15] in crowded spaces. Yet existing methods still occasionally output unacceptable solutions (e.g., trajectories lead to collisions), which raises significant safety concerns for real-world applications.

One key challenge for learning robust neural motion models stems from the common shortage of critical data. Very often, the training data is gathered from only a subset

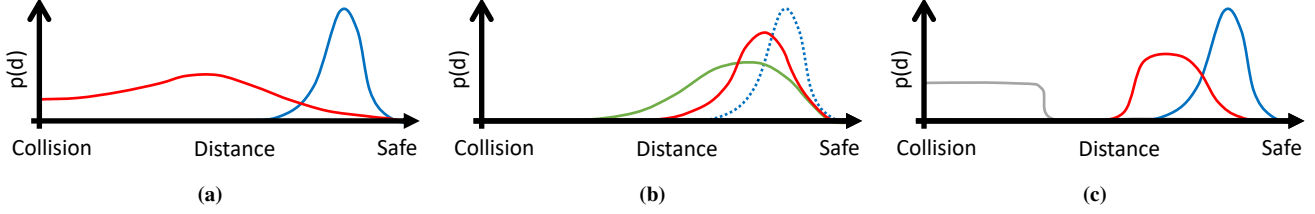


**Figure 1:** Illustration of social contrastive learning. Our method promotes the social awareness of neural motion models through the combination of contrastive representation learning and negative data augmentation in the multi-agent context.

of the state space rather than the entire one, e.g., human trajectories observed in safe scenarios without any dangerous occurrences. The lack of near-hit examples poses a significant difficulty for learning methods to truly capture the underlying social norms and function well in novel scenarios. As a consequence, prediction errors made by the learned models may rapidly accumulate over time, create a discrepancy between the training and test state distributions and eventually cause catastrophic errors [17, 19, 73].

Previous methods attempt to mitigate this issue mainly through *interactive data collections*, such as expert queries [20, 46, 52, 74, 85] and extra interactions between the learning agent and the environment [10, 33, 42, 70, 89]. Unfortunately, these methods are not only costly and tedious but often infeasible for forecasting problems, where interventions in human behaviors cannot easily be executed by another learning system for data collection purposes. These shortcomings motivate us to explore an alternative approach: *given a fixed training dataset, can we learn a robust model by exploiting our prior knowledge of socially unfavorable events?*

To this end, we propose a social contrastive learning method to promote the robustness of neural motion models (Figure 1). Contrastive learning [28, 29, 63, 81] has been a powerful technique for learning useful representations of complex signals, such as images and texts [13, 16, 26, 30, 38, 39, 53, 57, 58, 65, 90], in self-supervised



**Figure 2:** Illustration of different learning approaches to socially-aware sequential predictions from a distributional perspective. Due to the dependence of the state distribution (*e.g.*, separation distances between agents) on the sequential model in closed-loop operation [19, 73], (a) the vanilla supervised learning approach often suffers from covariate shift between the training (blue) and test (red) data. (b) Interactive data collection methods [20, 33, 46, 74] expand the training distribution to a wider range (green) through additional experiments, which however are expensive and even infeasible for forecasting problems. (c) Our social contrastive learning method augments negative data based on prior knowledge, explicitly informing neural motion models about socially unfavorable states (gray) for improved robustness.

and semi-supervised settings. In this work, we adapt this approach in the multi-agent context and introduce Social-NCE. Serving as a complement to the conventional predictive loss, *e.g.*, minimizing the distance between model’s outputs and the labeled behaviors, our Social-NCE encourages the extracted motion representation to preserve sufficient information for distinguishing a positive future event from a set of synthetic negative ones.

One crucial factor for our method lies in the design of positive and negative samples. Existing contrastive methods in vision and language domains often uniformly sample a large set of negatives from training data [13, 26, 30, 57, 67, 90]. However, this common choice does not provide much extra knowledge of spatio-temporal interactions among multiple agents. Here, we propose a more dedicated negative sampling strategy that explicitly constructs negative events from the regions of other agents in the future, given that it is typically forbidden or uncomfortable for multiple agents to visit the same or adjacent places simultaneously. As illustrated in Figure 2, our method can be viewed as a form of negative data augmentation [80]. It fulfills the goal of building robust neural motion models by intentionally informing the model about low likelihood states based on prior knowledge, as opposed to laboriously collecting state-action pairs in dangerous scenarios.

We evaluate our method on three tasks: human trajectory forecasting, behavioral cloning and reinforcement learning for robot navigation in crowds. Experimental results show that the proposed Social-NCE consistently reduce the collision rates and yields new state-of-the-art results on several benchmarks. Our method is model-agnostic and hence can be used as a generic component to promote the robustness of neural motion models. Our code is available at <https://github.com/vita-epfl/social-nce>.

## 2. Related Work

**Socially-aware Motion Representations.** Human motion in the social context has been traditionally studied based on relative distances and specific rules [4, 31, 56, 87].

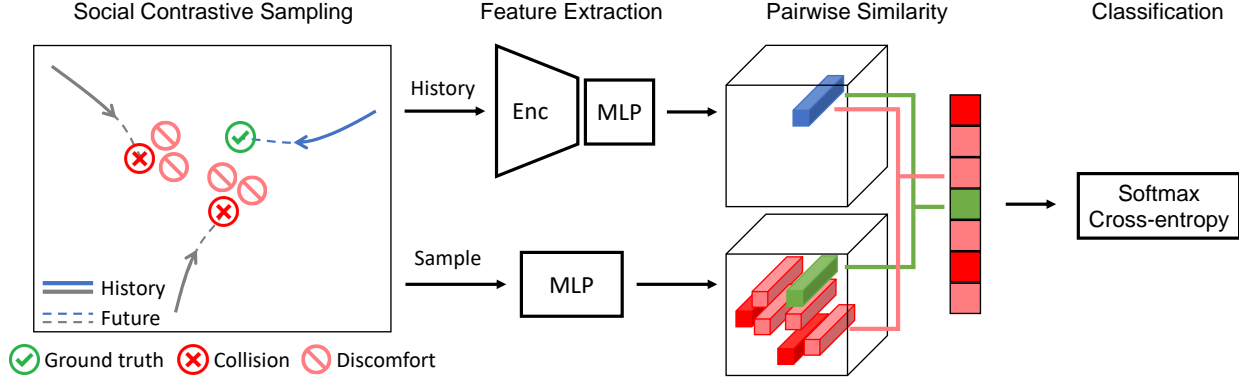
While these hand-crafted models have been successfully applied to various tasks [18, 24, 54, 91], they often struggle to capture the strong interactions among agents in complex scenes [75]. Some other methods attempt to learn the patterns of social behaviors from data. Yet early work often suffers from considerable performance drop in densely populated spaces due to limited modeling capacity [66, 86].

More recently, a variety of neural networks have been explored for learning socially-aware motion representations [5, 43]. Some peculiar neural architecture designs, such as feature pooling [3, 21, 27], attention mechanism [12, 34, 76, 88], and spatio-temporal graph [36, 41, 50, 83], have yielded promising results in crowded environments. However, the robustness of these methods remains a central concern. Our work is orthogonal to the design of neural motion models and focuses on the learning aspect to enhance the extracted motion representation.

**Covariate Shift.** The problem of covariate shift was observed back in [68] and has been a persistent challenge for sequential prediction problems [19, 73]. One practical solution is to actively query experts [46, 74, 85], which has been shown effective for behavioral cloning but hardly applicable to forecasting problems [43, 71, 75]. Inverse reinforcement learning methods [1, 10, 33, 42, 61, 89, 93] jointly learn a reward function and the corresponding optimal policy to account for the sequential nature. However, they typically require extensive explorations to solve a reinforcement learning (RL) subproblem [22].

Another line of work introduces additional loss terms penalizing the predictions that lead to undesirable events, such as collisions and off-road trajectories [8, 62]. However, these penalties are dependent on the predicted states during training and become utterly ineffective once the model fits the dataset well in late training stages.

Closely related to our work, [55, 92] propose to learn a robust value (or cost) function by making use of negative samples. Our method differs from theirs in two key aspects: [55, 92] change the task loss that directly affects (and potentially biases) the model output, whereas our goal



**Figure 3:** Social contrastive learning in the multi-agent context. Given a scenario that contains a primary agent of our interest (blue) and multiple neighboring agents in its surroundings (gray), our Social-NCE loss encourages the extracted motion representation to be close to a positive future event in an embedding space and apart from some synthetic negative events that could have caused collisions or discomforts.

is to enhance the extracted motion representation without modifications in the main task; they draw negative samples randomly, in contrast, we design a more informed sampling strategy using prior knowledge.

**Contrastive Learning.** Contrastive learning was proposed in [29] to learn an embedding space such that a simple similarity measure can approximate the preferred neighborhood relationships. This approach has recently achieved stunning results in a broad spectrum of areas, including computer vision [13, 30], natural language understanding [7, 53, 64], image synthesis [65] and robotics [79]. Some detailed design choices, such as positive and negative sampling, often play a critical role in the practical success of contrastive methods [16, 38, 69, 72]. To the best of our knowledge, we are the first to adapt contrastive learning in the multi-agent motion context and explore the sampling method which is unique and critical to socially-aware motion representation learning.

### 3. Method

The robustness of neural motion models has been a long-standing concern for safety-critical applications like autonomous vehicles and social robots. In many practical scenarios, the training data may only contain quality examples from safe normal states but lack dangerous occurrences. The incomplete state coverage poses a significant challenge for learning algorithms to truly capture the underlying social norms and generalize to challenging scenarios.

In this section, we present a learning method that aims to tackle this challenge by means of contrastive representation learning. We will first briefly introduce the background of contrastive learning and then present a social contrastive loss for motion representations. We will finally introduce a sampling strategy tailored for the multi-agent context.

#### 3.1. Contrastive Representation Learning

Representation learning typically consists in learning a parametric function (*i.e.*, encoder) that maps the raw data into a feature space to extract abstract and useful information for downstream tasks [9]. Recent contrastive learning methods often adopt the principle of noise contrastive estimation [23, 28] in an embedding space, namely the InfoNCE loss [63], to train an encoder:

$$\mathcal{L}_{\text{NCE}} = -\log \frac{\exp(\text{sim}(q, k^+)/\tau)}{\sum_{n=0}^N \exp(\text{sim}(q, k_n)/\tau)}, \quad (1)$$

where the encoded query  $q$  is brought close to one positive key  $k_0 = k^+$  and pushed apart from  $N$  negative keys  $\{k_1, \dots, k_N\}$ ,  $\tau$  is a temperature hyperparameter, and  $\text{sim}(u, v) = u^\top v / (\|u\| \|v\|)$  is the cosine similarity between two feature vectors. It has been shown that minimizing the InfoNCE loss is equivalent to maximize the lower bound on the mutual information between the raw input and the latent representation [63]. Moreover, the representations learned by this approach have provable performance guarantees on downstream tasks [7]. **The empirical success of this approach often highly relies on the informativeness of positive and negative samples** [13, 16, 37, 38, 69, 72, 82].

#### 3.2. Social NCE

Consider a trajectory forecasting problem in crowded spaces as an example. Let  $s_t^i = (x_t^i, y_t^i)$  denote the position of agent  $i$  at time  $t$  and  $s_t = \{s_t^1, \dots, s_t^M\}$  denote the joint state of  $M$  agents in the scene. Given a sequence of history observations  $\{s_1, \dots, s_t\}$ , the task is to predict future trajectories of all agents  $\{s_{t+1}, \dots, s_T\}$  until time  $T$ . Many recent forecasting models are designed as encoder-decoder neural networks, where the motion encoder  $f(\cdot)$  first extracts a compact representation  $h_t^i$  with respect to agent  $i$  and the decoder  $g(\cdot)$  subsequently rolls out its future trajec-

tory  $\hat{s}_{t+1:T}^i$ :

$$\begin{aligned} h_t^i &= f(s_{1:t}, i), \\ \hat{s}_{t+1:T}^i &= g(h_t^i). \end{aligned} \quad (2)$$

To model social interactions among agents,  $f(\cdot)$  typically contains two sub-modules: a sequential module  $f_S(\cdot)$  that encodes each individual sequence and an interaction module  $f_I(\cdot)$  that shares information among agents, *e.g.*,

$$\begin{aligned} z_t^i &= f_S(h_{t-1}^i, s_t^i), \\ h_t^i &= f_I(z_t, i). \end{aligned} \quad (3)$$

where  $z_t^i$  is the latent representation of agent  $i$  given the observation of its own state at time  $t$  and  $z_t = \{z_t^1, \dots, z_t^M\}$ . A variety of architectures have been explored for each module and validated on accuracy measures [3, 50, 51]. Nevertheless, their robustness remains an open issue. Several recent works [8, 43] have shown that trajectories predicted by existing models often output socially unacceptable solutions (*e.g.*, collisions), indicating a lack of common sense about social norms.

To tackle this challenge, we propose Social-NCE, a variant of InfoNCE tailored for socially-aware motion representation learning. As illustrated in Figure 3, we construct the encoded query and key vectors for the primary agent  $i$  at time  $t$  as follows:

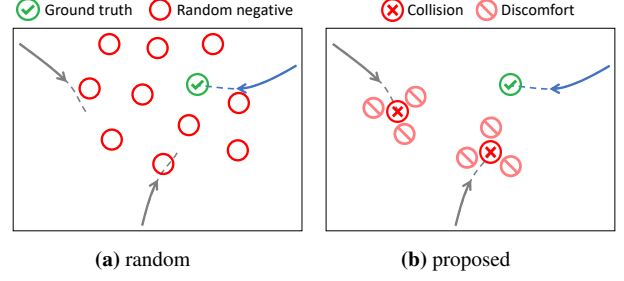
- query: embedding of history observations  $q = \psi(h_t^i)$ , where  $\psi(\cdot)$  is an MLP projection head.
- key: embedding of a future event  $k = \phi(s_{t+\delta t}^i, \delta t)$ , where  $\phi(\cdot)$  is an event encoder modeled by an MLP,  $s_{t+\delta t}^i$  is a sampled spatial location and  $\delta t > 0$  is the sampling horizon.

By tuning  $\delta t \in \Lambda$  in a range, *e.g.*,  $\Lambda = \{1, \dots, 4\}$ , we can take into account future events in the next few steps simultaneously. Nevertheless, when  $\delta t$  is a fixed value,  $\phi(\cdot)$  can be simplified as a location encoder, *i.e.*,  $\phi(s_{t+\delta t}^i)$ .

In each frame, we draw one positive key and multiple negative keys based on future trajectories in the scene, which we will describe in the next Section 3.3. Following [13, 30, 90], we normalize the embedding vectors onto a unit sphere and train the parametric models  $f(\cdot)$ ,  $\phi(\cdot)$ ,  $\psi(\cdot)$  jointly with the objective of mapping the positive pair of query and key to similar points, relative to the other negative pairs, in the embedding space:

$$\mathcal{L}_{\text{SocialNCE}} = -\log \frac{\exp(\psi(h_t^i) \cdot \phi(s_{t+\delta t}^{i,+}, \delta t)/\tau)}{\sum_{\delta t \in \Lambda} \sum_{n=0}^N \exp(\psi(h_t^i) \cdot \phi(s_{t+\delta t}^{i,n}, \delta t)/\tau)}. \quad (4)$$

The full training objective is a weighted combination of the conventional task loss, *e.g.*, mean squared error (MSE)



**Figure 4:** Different negative sampling methods in the multi-agent context. (a) The conventional random sampling method draws negative samples homogeneously scattered in the space, which does not provide much information about social norms. (b) Our safety-driven sampling strategy seeks more informative negative samples from the neighborhood of other agents in the future.

or negative log-likelihood (NLL) for trajectory forecasting, and the proposed social contrastive loss:

$$\mathcal{L}(f, g, \psi, \phi) = \mathcal{L}_{\text{task}}(f, g) + \lambda \mathcal{L}_{\text{SocialNCE}}(f, \psi, \phi), \quad (5)$$

where  $\lambda$  is a hyper-parameter controlling the emphasis on Social-NCE.

### 3.3. Multi-agent Contrastive Sampling

One critical design choice of the proposed social contrastive learning lies in the sampling strategy. The recent successes in contrastive learning of visual representations are heavily tied to the use of a large set of negative samples uniformly drawn from the training dataset [13, 26, 30, 57, 67, 90]. Unfortunately, this common practice is not suitable for socially-aware motion representation learning. As the main predictive loss already encourages the model to replicate the socially good behaviors from training examples, adding another discrimination task between the correct solution and other randomly scattered negatives cannot provide much extra information about social norms. Worse yet, the random negative sampling, like Figure 4a, may contradict the nature of the multimodal distribution of the future trajectories and incorrectly penalize plausible solutions.

To effectively incorporate our domain knowledge of socially unfavorable events in the multi-agent context, we propose a safety-driven sampling strategy. As shown in Figure 4b, we draw a set of negative samples from the neighborhood of other agents in the future at time  $t + \delta t$ ,

$$s_{t+\delta t}^{i,n-} = s_{t+\delta t}^j + \Delta s_p + \epsilon, \quad (6)$$

where  $j \in \{1, 2, \dots, M\} \setminus i$  is the index of other agents, and  $\Delta s_p = (\rho \cos \theta_p, \rho \sin \theta_p)$  is a local displacement to account for the social comfort area. For instance,  $\rho$  can be the minimum physical distance between two agents and  $\theta_p = 0.25p\pi$ ,  $p \in \{0, 1, \dots, 7\}$ . Thus, a total of  $N = 8(M - 1)$

negative samples are synthesized. We also add random perturbations to each sampled location  $\epsilon \sim \mathcal{N}(0; c_\epsilon \mathbf{I})$ , where  $c_\epsilon$  is a small constant, *e.g.*, 0.05 [m], to prevent over-fitting. For positive sampling, we pick a location from the ground truth region of the primary agent  $i$  at time  $t + \delta t$ ,

$$s_{t+\delta t}^{i,+} = s_{t+\delta t}^i + \epsilon. \quad (7)$$

The key intuition behind our method is that conventional predictive learning only requires the model to replicate good motion behaviors in normal states without the need to understand the consequence of undesirable outputs. In contrast, by using Social-NCE in tandem with safety-driven negative sampling, we actively enforce the extracted motion representation  $h_t^i$  to contain the necessary information for identifying unfavorable events that could have led to catastrophic outcomes. This subtle but essential difference enables us to learn a significantly more robust model from the fixed training dataset.

## 4. Experiment

We empirically validate the proposed Social-NCE on three different tasks: (i) human trajectory forecasting, (ii) imitation learning and (iii) reinforcement learning for robot navigation in multi-agent environments.

On each task, we compare the models obtained by three training methods:

- Vanilla: models trained without contrastive loss.
- Random: models trained with contrastive loss and the random negative sampling (Figure 4a).
- Social-NCE (ours): models trained with contrastive loss and our safety-driven negative sampling strategy (Figure 4b).

### 4.1. Implementation Details

In our experiments, we use two different 2-layer MLPs as the projection head  $\psi(\cdot)$  and the event encoder  $\phi(\cdot)$ . We encode the history observations and future events into 8-dimensional embedding vectors. The distance hyperparameter  $\rho$  is set as 0.2 [m] for trajectory forecasting tasks and 0.6 [m] for robot navigation according to the geometric size of agents in environments. By default, the sampling horizon  $\delta t$  is set up to 4 and the temperature  $\tau$  is set as 0.1. All models are trained with the Adam optimizer [40].

### 4.2. Trajectory Forecasting

We first evaluate our method on the human trajectory forecasting task. We compare the performances of several forecasting models trained with and without the proposed Social-NCE on the ETH & UCY benchmark [48, 66] and the Trajnet++ interaction-centric benchmark [43]. For a fair

and direct comparison with prior work, we implement our learning method in the official public code of each model without any modifications on architectures.

All forecasting models are trained and evaluated in the following setting: given 8 time steps (3.2 seconds) of observations as input, predict future trajectories for 12 time steps (4.8 seconds) for all human agents in the scene. Similar to previous work [27, 43, 92], we measure the performance of each model with two metrics:

- Final displacement error (FDE): the Euclidean distance between the predicted output and the ground truth at the last time step.
- Collision rate (COL): the percentage of test cases where the predicted trajectories of different agents run into collisions.

More experimental details are outlined in the Appendix A.1.

#### 4.2.1 ETH & UCY Benchmark

We start the evaluation of our method on the ETH and UCY datasets, which contain 5 subsets of real-world pedestrian trajectories widely used in previous work [3, 27, 60, 77]. We consider the following two state-of-the-art multi-modal forecasting models as baselines:

- Social-STGCNN [60]: a spatio-temporal convolutional network with graph representations of pedestrian trajectories.
- Trajectron++ [77]: a VAE-based recurrent model using element-wise sum to aggregate interactions.

Table 1 shows the experimental results in terms of Top-20 FDE and collision rate. Compared with the conventional predictive learning, our Social-NCE reduces the collision rate by 37.0% on average for the Social-STGCNN and 45.7% for the Trajectron++, while remaining competitive with respect to the final displacement error. These results suggest the benefits of our method for boosting the robustness of modern trajectory forecasting models without hurting the accuracy and diversity of endpoint predictions.

#### 4.2.2 Trajnet++ Benchmark

We further validate our method on the Trajnet++ [43], an emerging open challenge with an emphasis on interaction modeling. Our evaluation is performed on the following interacting sub-categories:

- Avoidance: the sub-category where the primary pedestrian avoids others coming from the opposite direction.
- Group: the sub-category where the primary pedestrian keeps a close and approximately constant distance with one or more moving neighbors.



Dataset	Social-STGCNN [60]					Trajectron++ [77]				
	Vanilla		Ours		Gain	Vanilla		Ours		Gain
	FDE ↓	COL ↓	FDE ↓	COL ↓	COL ↑	FDE ↓	COL ↓	FDE ↓	COL ↓	COL ↑
ETH	1.223	1.33	1.224	0.61	<b>54.1%</b>	0.810	1.16	0.791	0.00	<b>100.0%</b>
Hotel	0.687	3.82	0.678	3.35	<b>12.3%</b>	0.184	0.84	0.177	0.38	<b>54.6%</b>
Univ	0.912	9.11	0.879	6.44	<b>29.3%</b>	0.450	3.38	0.435	3.08	<b>8.9%</b>
Zara1	0.525	2.27	0.515	1.02	<b>55.1%</b>	0.320	0.46	0.330	0.18	<b>61.5%</b>
Zara2	0.480	6.86	0.482	3.37	<b>50.9%</b>	0.250	1.03	0.255	0.99	<b>3.3%</b>
Average	0.765	4.70	0.756	2.96	<b>37.0%</b>	0.403	1.37	0.398	0.93	<b>45.7%</b>

**Table 1:** Comparison between Social-NCE and the vanilla predictive learning for two recent multi-modal trajectory forecasting models, Social-STGCNN [60] and Trajectron++ [77], on the ETH [66] / UCY [48] datasets. Our method reduces the collision rates of these state-of-the-art models by more than 37%, while being on par with the vanilla training counterparts in terms of the top-20 final displacement error. Note that we run the official public code to obtain the baseline results, which are subject to mild differences ( $< 2\%$  on average) from the corresponding papers.

Category	Social-LSTM [3]					Directional-LSTM [43]				
	Vanilla		Ours		Gain	Vanilla		Ours		Gain
	FDE ↓	COL ↓	FDE ↓	COL ↓	COL ↑	FDE ↓	COL ↓	FDE ↓	COL ↓	COL ↑
Avoidance	1.23	13.12	1.23	10.83	<b>17.5%</b>	1.33	12.92	1.33	10.42	<b>19.4%</b>
Group	0.97	7.44	0.97	4.65	<b>37.5%</b>	1.05	6.51	1.05	5.89	<b>9.5%</b>
Overall	1.14	6.44	1.14	5.31	<b>17.6%</b>	1.22	5.49	1.22	4.59	<b>16.4%</b>

**Table 2:** Comparison between Social-NCE and the vanilla predictive learning for two top-ranking uni-modal forecasting models on different interacting categories of the Trajnet++ benchmark [43]. Our method reduces the collision rate of the previously most robust model by a clear margin. Note that we obtain all results from the benchmark reports, where FDEs are rounded to two decimal places.

- Overall: all the scenes with a presence of social interactions, including the ones that are hard to categorize.

These interacting categories exclude both *linear* trajectories that often dilute the evaluations of interaction modeling and *non-linear yet non-agent-interacting* trajectories that are strongly affected by the scene context or other unobservable variables, and are therefore particularly suitable for assessing the social awareness of forecasting models.

We use two top-ranking models on the Trajnet++ benchmark as baselines:

- Social-LSTM [3]: a LSTM-based model with an interaction module over hidden states of nearby agents.
- Directional-LSTM [43]: a LSTM-based model with an interaction module sharing relative velocities of nearby agents.

Table 2 shows that, similar to the results on the ETH & UCY benchmark, our social contrastive method yields lower collision rates than the vanilla predictive learning by a clear margin (9.5%-37.5%) across all considered sub-categories. In particular, the Directional-LSTM trained with our method clearly outperforms its counterpart, which was previously the most robust model on the public benchmark. It is worth noting that, again, our method does not have a

significant impact on prediction accuracy. We conjecture that this is because our method tends to adjust the output trajectories locally instead of changing their global patterns, as reflected by the qualitative examples in the Appendix C.

### 4.3. Imitation Learning

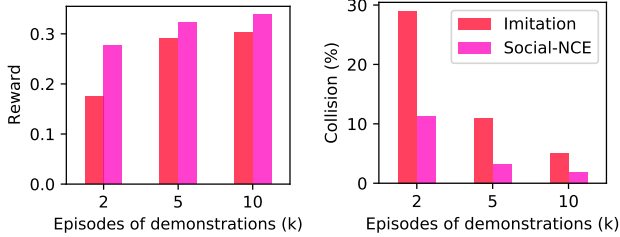
Next, we examine the effectiveness of Social-NCE applied to imitation learning for robot navigation in dense crowds [12, 14, 15]. We use an open-sourced simulator of crowd navigation [12], where the task for the robot is to navigate through 5 simulated pedestrians and arrive at the target destination with time efficiency. In each time step, the robot receives the observable states of other agents and outputs an action. We follow the evaluation protocol in [12], which quantifies the performance of a policy using three metrics: navigation time, collision rate, and the accumulated reward as follows:

$$r(s_t, a_t) = \begin{cases} -0.25 & \text{if } d_m^t < 0 \\ -0.1 + d_m^t/2 & \text{else if } d_m^t < 0.2 \\ 1 & \text{else if goal is reached} \\ 0 & \text{otherwise} \end{cases} \quad (8)$$

where  $d_m^t$  is the minimum separation distance between the robot and the humans during the time interval  $[t - \Delta t, t]$ .

Method	Reward $\uparrow$	Time (s) $\downarrow$	Collision (%) $\downarrow$
Vanilla	$0.28 \pm 0.01$	$10.31 \pm 0.07$	$11.11 \pm 1.45$
Random	$0.24 \pm 0.02$	$10.32 \pm 0.12$	$18.60 \pm 4.69$
Ours	<b><math>0.32 \pm 0.01</math></b>	$10.33 \pm 0.07$	<b><math>3.40 \pm 1.36</math></b>

**Table 3:** Quantitative results of imitation learning with different methods on a 5k demonstration dataset. Higher is better for reward, and lower is better for the other metrics. Compared with the vanilla baseline, our method brings down the collision rate by approximately 69%.



**Figure 5:** Social-NCE for imitation learning with different amounts of demonstrations. The conventional behavioral cloning method suffers from a significant performance drop in the low data regime, whereas our method is able to retain much better results thanks to the additional information absorbed from the social contrastive task.

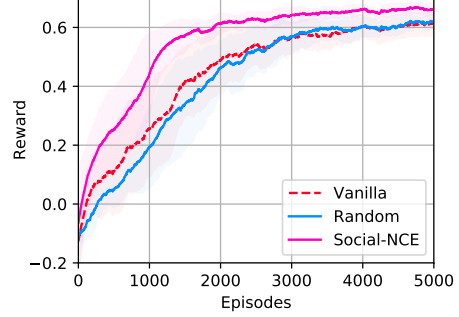
#### 4.3.1 Behavioral Cloning

For imitation learning, we collect a demonstration dataset that consists of 5000 simulation episodes using the pre-trained SARL policy in [12] as an expert. We train an imitator on the collected data for 200 epochs and evaluate the average performance of the last 10 models saved every 5 training epochs.

As shown in Table 3, the imitation learning algorithm trained with random negative sampling fails to outperform the vanilla baseline. In fact, it even worsens the learned policy. In contrast, our method (with weight  $\lambda = 0.1$ ) leads to consistently higher reward and lower collision rate. Specifically, our method reduces the collision rate by approximately 69% compared with the vanilla baseline and attains an average reward of 0.323, which is highly close to the result of the demonstrator in [12].

#### 4.3.2 Low-data Regime

The performance of the standard behavioral cloning approach often degrades substantially when provided with limited demonstrations. We examine the potential of the proposed Social-NCE for data-efficient imitation learning by comparing policies trained on datasets of different sizes. As shown in Figure 5, with decreasing amounts of demonstrations, the performance of the vanilla method drops sharply. Notably, the baseline model trained on 2k episodes of demonstrations causes collisions in 29% of test cases. In



**Figure 6:** Learning curves of Rainbow DQN [32] with different methods for crowd navigation. Results are averaged across 8 random seeds. The shaded area spans one standard deviation. In contrast to the vanilla and random negative counterparts, the agent with Social-NCE is significantly more sample efficient and achieves higher final reward.

Method	Reward w.r.t. fraction of dataset			
	100%	50%	25%	10%
Vanilla	80.1%	75.2%	53.2%	14.4%
Random	81.3%	71.5%	51.7%	7.0%
Ours	<b>91.6%</b>	<b>84.6%</b>	<b>79.2%</b>	<b>69.0%</b>

**Table 4:** Offline RL normalized scores attained by the vanilla rainbow and Social-NCE agents (higher is better). Normalized score is calculated as:  $100 \times (\text{agent score} - \text{random play score}) / (\text{optimal agent score} - \text{random play score})$ , as in [59]. Our Social-NCE consistently facilitates the recovery of a near-optimal policy and is particularly advantageous in the low-data regime.

contrast, our method succeeds in retaining a much higher reward and safety in the low-data regime. For instance, the collision rate of our method with 2k demonstrations is comparable to the baseline with 5k demonstrations. Similarly, our method using 5k training data obtains a higher reward than the counterpart using 10k training data. This result suggests that the learner can absorb a considerable amount of useful information from the designed negative data augmentation, greatly alleviating the information shortage in small training sets.

#### 4.4 Reinforcement Learning

Finally, we evaluate the proposed Social-NCE for reinforcement learning (RL) algorithms on the crowd navigation task. We adopt the Rainbow DQN [32], a state-of-the-art model-free RL method, as baseline and follow the architecture of value-based SARL policy [12] to build the encoder  $f(\cdot)$ . To effectively apply Social-NCE to the Rainbow agent, we add a linear layer after the interaction and pooling modules. We also replace the planning module in SARL by dueling and categorical layers, as in the standard Rainbow agent [32]. In order to isolate the impact of Social-NCE, we use the dense reward function proposed in [78], which elim-

Horizon	Vanilla	1	2	3	4	1-4
Reward $\uparrow$	$0.283 \pm 0.008$	$0.281 \pm 0.019$	$0.296 \pm 0.009$	<b><math>0.311 \pm 0.009</math></b>	$0.307 \pm 0.012$	<b><math>0.323 \pm 0.005</math></b>
Time (s) $\downarrow$	$10.306 \pm 0.068$	$10.345 \pm 0.065$	$10.281 \pm 0.141$	$10.322 \pm 0.134$	$10.348 \pm 0.107$	$10.334 \pm 0.072$
Collision (%) $\downarrow$	$11.11 \pm 1.45$	$11.24 \pm 3.46$	$9.13 \pm 2.02$	<b><math>5.83 \pm 1.62</math></b>	$6.09 \pm 2.26$	<b><math>3.40 \pm 1.36</math></b>

**Table 5:** Social-NCE for imitation learning with different choices of sampling horizon. Higher is better for reward, and lower is better for the other metrics. Taking multiple time steps (1-4) into account simultaneously yields better results than a fixed horizon of one time step.

inates the necessity of imitation pre-training in [12] (more details in the Appendix A.2).

#### 4.4.1 Off-policy Reinforcement Learning

We first validate our Social-NCE method in the standard off-policy setting, where an RL agent learns from the replay buffer data gathered over the learning process. We set the temperature as  $\tau = 0.2$  and weight as  $\lambda = 1.0$  for the Social-NCE loss. Figure 6 shows the experimental results of each method averaged over 8 random seeds.

The vanilla Rainbow agent reaches a reward value of 0.6 using more than 4000 episodes. In comparison, the agent equipped with our method demonstrates a much higher sample efficiency. It attains the same level of reward in less than 2000 episodes and quickly obtains a collision-free policy thanks to the prior knowledge from the social contrastive task. Additionally, our method also offers a slight improvement in the final performance. On the contrary, the random negative sampling is not able to provide any significant performance gain, similar to the experimental results above.

#### 4.4.2 Offline Reinforcement Learning

Lastly, we explore the potential of our method in the offline RL setting, in which an agent learns from a static dataset of logged experiences without additional interactions with the environment [49]. The offline RL setting has attracted rapidly growing attentions due to its tremendous promise for making good use of immerse experience datasets. Nevertheless, most deep reinforcement learning algorithms are highly vulnerable to the distribution mismatch between the policy being trained and the ones used for data collection [2, 25, 35, 44, 45].

To verify our method in the offline RL setting, we collect a dataset using the vanilla rainbow agent in the following process: (i) 10k episodes are collected during online RL training from scratch, (ii) multiple free explorations are carried out using online RL checkpoint models trained after  $K$  episodes, where  $K \in \{500, 1000, 3000, 5000\}$ . Each free exploration contains 5k episodes, and the full dataset is made up of 30k episodes of trials in total. We train an offline Rainbow policy on  $N\%$  of the experiences randomly sampled from the dataset, where  $N \in \{10, 25, 50, 100\}$ . We use temperature  $\tau = 0.2$  and weight  $\lambda = 0.1$ .

Table 4 reports the average performance of different offline methods across 10 random seeds. Due to the aforementioned distributional shift, no offline methods attain reward scores as high as the online RL algorithm. Nevertheless, our Social-NCE substantially narrows the performance gap and consistently delivers better results than the vanilla rainbow. Notably, the offline policy with our method achieves comparable performance to the best of vanilla baseline using only 25% of the collected data.

#### 4.5. Ablation: Event Encoder

To validate the benefits of event encoder  $\phi(s_{t+\delta t}^i, \delta t)$ , we compare the performance of Social-NCE for imitation learning with different sampling horizons. When the sampling horizon is a fixed value, we use a simplified location encoder  $\phi(s_{t+\delta t}^i)$  that only takes the location of a sample as input. Table 5 reports the results obtained by using contrastive samples either at a single fixed horizon in a range from 1 to 4 or across all four steps simultaneously. Among the single-step choices,  $\delta t = 3$  yields significant performance gains on both reward and collision metrics. On the contrary,  $\delta t = 1$  does not provide any improvements in comparison to the baseline due to its short-sightedness. When taking all four steps into account, our method attains the best result, suggesting the importance of establishing social contrastive tasks at multiple horizons.

## 5. Conclusion

In this work, we present a contrastive learning method for socially-aware motion representation learning in the multi-agent context. The proposed Social-NCE loss, combined with informative negative sampling, consistently boosts the performance of recent human trajectory forecasting and crowd navigation algorithms in various settings. The strength of our method suggests that exploiting negative data augmentations by means of contrastive learning can be a promising alternative to traditional interactive data collections for building robust neural motion models. We hope this approach will also be useful for other sequential decision problems that involve complex spatio-temporal dynamics but are short of critical data.



## Acknowledgments

This work is supported by the Swiss National Science Foundation under the Grant 200021-L92326. We thank Parth Kothari for helpful suggestions on human trajectory forecasting experiments. We also thank Yifan Sun, Taylor Mordan, Mohammadhossein Bahari, Lorenzo Bertoni, Sven Kreiss and other VITA members for valuable feedback on early drafts.

## References

- [1] Pieter Abbeel and Andrew Y. Ng. Apprenticeship Learning via Inverse Reinforcement Learning. In *Proceedings of the Twenty-first International Conference on Machine Learning, ICML '04*. ACM, 2004. 2
- [2] Rishabh Agarwal, Dale Schuurmans, and Mohammad Norouzi. An Optimistic Perspective on Offline Reinforcement Learning. In *International Conference on Machine Learning*, pages 104–114. PMLR, Nov. 2020. ISSN: 2640-3498. 8
- [3] Alexandre Alahi, Kratarth Goel, Vignesh Ramanathan, Alexandre Robicquet, Li Fei-Fei, and Silvio Savarese. Social LSTM: Human Trajectory Prediction in Crowded Spaces. In *2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 961–971, June 2016. ISSN: 1063-6919. 1, 2, 4, 5, 6
- [4] Alexandre Alahi, Vignesh Ramanathan, and Li Fei-Fei. Socially-Aware Large-Scale Crowd Forecasting. In *2014 IEEE Conference on Computer Vision and Pattern Recognition*, pages 2211–2218, June 2014. ISSN: 1063-6919. 2, 13
- [5] Alexandre Alahi, Vignesh Ramanathan, Kratarth Goel, Alexandre Robicquet, Amir A Sadeghian, Li Fei-Fei, and Silvio Savarese. Learning to predict human behavior in crowded scenes. In *Group and Crowd Behavior for Computer Vision*, pages 183–207. Elsevier, 2017. 2
- [6] Javad Amirian, Jean-Bernard Hayet, and Julien Pettre. Social Ways: Learning Multi-Modal Distributions of Pedestrian Trajectories With GANs. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops*, pages 0–0, 2019. 1
- [7] Sanjeev Arora, Hrishikesh Khandeparkar, Mikhail Khodak, Orestis Plevrakis, and Nikunj Saunshi. A Theoretical Analysis of Contrastive Unsupervised Representation Learning. *arXiv:1902.09229 [cs, stat]*, Feb. 2019. arXiv: 1902.09229. 3
- [8] Mayank Bansal, Alex Krizhevsky, and Abhijit Ogale. ChauffeurNet: Learning to Drive by Imitating the Best and Synthesizing the Worst. *arXiv:1812.03079 [cs]*, Dec. 2018. arXiv: 1812.03079. 2, 4
- [9] Y. Bengio, A. Courville, and P. Vincent. Representation Learning: A Review and New Perspectives. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 35(8):1798–1828, Aug. 2013. Conference Name: IEEE Transactions on Pattern Analysis and Machine Intelligence. 3
- [10] Kianté Brantley, Wen Sun, and Mikael Henaff. Disagreement-Regularized Imitation Learning. In *International Conference on Learning Representations*, Sept. 2019. 1, 2
- [11] T. Chavdarova, P. Baqué, S. Bouquet, A. Maksai, C. Jose, T. Bagautdinov, L. Lettry, P. Fua, L. Van Gool, and F. Fleuret. WILDTRACK: A Multi-camera HD Dataset for Dense Unscripted Pedestrian Detection. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 5030–5039, June 2018. ISSN: 2575-7075. 13
- [12] Changan Chen, Yuejiang Liu, Sven Kreiss, and Alexandre Alahi. Crowd-Robot Interaction: Crowd-Aware Robot Navigation With Attention-Based Deep Reinforcement Learning. In *2019 International Conference on Robotics and Automation (ICRA)*, pages 6015–6022, May 2019. ISSN: 2577-087X. 1, 2, 6, 7, 8, 13
- [13] Ting Chen, Simon Kornblith, Mohammad Norouzi, and Geoffrey Hinton. A Simple Framework for Contrastive Learning of Visual Representations. *arXiv:2002.05709 [cs, stat]*, Mar. 2020. arXiv: 2002.05709. 1, 2, 3, 4
- [14] Yu Fan Chen, Michael Everett, Miao Liu, and Jonathan P. How. Socially aware motion planning with deep reinforcement learning. In *2017 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 1343–1350, Sept. 2017. ISSN: 2153-0866. 1, 6
- [15] Yu Fan Chen, Miao Liu, Michael Everett, and Jonathan P. How. Decentralized Non-communicating Multiagent Collision Avoidance with Deep Reinforcement Learning. *arXiv:1609.07845 [cs]*, Sept. 2016. arXiv: 1609.07845. 1, 6
- [16] Ching-Yao Chuang, Joshua Robinson, Lin Yen-Chen, Antonio Torralba, and Stefanie Jegelka. Debaised Contrastive Learning. *arXiv:2007.00224 [cs, stat]*, July 2020. arXiv: 2007.00224. 1, 3
- [17] Felipe Codevilla, Eder Santana, Antonio M. Lopez, and Adrien Gaidon. Exploring the Limitations of Behavior Cloning for Autonomous Driving. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 9329–9338, 2019. 1
- [18] Pasquale Coscia, Francesco Castaldo, Francesco AN Palmieri, Lamberto Ballan, Alexandre Alahi, and Silvio Savarese. Point-based path prediction from polar histograms. In *2016 19th International Conference on Information Fusion (FUSION)*, pages 1961–1967. IEEE, 2016. 2
- [19] Hal Daumé, John Langford, and Daniel Marcu. Search-based structured prediction. *Machine Learning*, 75(3):297–325, June 2009. 1, 2
- [20] Pim de Haan, Dinesh Jayaraman, and Sergey Levine. Causal Confusion in Imitation Learning. *arXiv:1905.11979 [cs, stat]*, Nov. 2019. arXiv: 1905.11979. 1, 2
- [21] Nachiket Deo and Mohan M. Trivedi. Convolutional Social Pooling for Vehicle Trajectory Prediction. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pages 1549–15498, June 2018. ISSN: 2160-7516. 1, 2
- [22] Gabriel Dulac-Arnold, Daniel Mankowitz, and Todd Hester. Challenges of Real-World Reinforcement Learning. *arXiv:1904.12901 [cs, stat]*, Apr. 2019. arXiv: 1904.12901. 2
- [23] Chris Dyer. Notes on Noise Contrastive Estimation and Negative Sampling. *arXiv:1410.8251 [cs]*, Oct. 2014. arXiv: 1410.8251. 3

- [24] Gonzalo Ferrer, Anais Garrell, and Alberto Sanfeliu. Robot companion: A social-force based approach with human awareness-navigation in crowded environments. In *2013 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 1688–1694, Nov. 2013. ISSN: 2153-0866. [2](#)
- [25] Scott Fujimoto, David Meger, and Doina Precup. Off-Policy Deep Reinforcement Learning without Exploration. In *International Conference on Machine Learning*, pages 2052–2062. PMLR, May 2019. ISSN: 2640-3498. [8](#)
- [26] Yoav Goldberg and Omer Levy. word2vec Explained: deriving Mikolov et al.’s negative-sampling word-embedding method. *arXiv:1402.3722 [cs, stat]*, Feb. 2014. arXiv: 1402.3722. [1](#), [2](#), [4](#)
- [27] A. Gupta, J. Johnson, L. Fei-Fei, S. Savarese, and A. Alahi. Social GAN: Socially Acceptable Trajectories with Generative Adversarial Networks. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 2255–2264, June 2018. ISSN: 2575-7075. [1](#), [2](#), [5](#)
- [28] Michael Gutmann and Aapo Hyvärinen. Noise-contrastive estimation: A new estimation principle for unnormalized statistical models. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 297–304. JMLR Workshop and Conference Proceedings, Mar. 2010. ISSN: 1938-7228. [1](#), [3](#)
- [29] Raia Hadsell, Sumit Chopra, and Yann LeCun. Dimensionality Reduction by Learning an Invariant Mapping. In *2006 IEEE Computer Society Conference on Computer Vision and Pattern Recognition - Volume 2 (CVPR’06)*, volume 2, pages 1735–1742, New York, NY, USA, 2006. IEEE. [1](#), [3](#)
- [30] Kaiming He, Haoqi Fan, Yuxin Wu, Saining Xie, and Ross Girshick. Momentum Contrast for Unsupervised Visual Representation Learning. *arXiv:1911.05722 [cs]*, Mar. 2020. arXiv: 1911.05722. [1](#), [2](#), [3](#), [4](#)
- [31] Dirk Helbing and Peter Molnar. Social Force Model for Pedestrian Dynamics. *Physics Review E*, May 1998. [2](#)
- [32] Matteo Hessel, Joseph Modayil, Hado van Hasselt, Tom Schaul, Georg Ostrovski, Will Dabney, Dan Horgan, Bilal Piot, Mohammad Azar, and David Silver. Rainbow: Combining Improvements in Deep Reinforcement Learning. *arXiv:1710.02298 [cs]*, Oct. 2017. arXiv: 1710.02298. [7](#)
- [33] Jonathan Ho and Stefano Ermon. Generative Adversarial Imitation Learning. *arXiv:1606.03476 [cs]*, June 2016. tex.ids: hoGenerativeAdversarialImitation2016a arXiv: 1606.03476. [1](#), [2](#)
- [34] Yingfan Huang, Huikun Bi, Zhaoxin Li, Tianlu Mao, and Zhaoqi Wang. STGAT: Modeling Spatial-Temporal Interactions for Human Trajectory Prediction. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, pages 6272–6281, 2019. tex.ids: huangSTGATModelingSpatialTemporal2019a ISSN: 2380-7504. [1](#), [2](#)
- [35] Riashat Islam, Komal K. Teru, Deepak Sharma, and Joelle Pineau. Off-Policy Policy Gradient Algorithms by Constraining the State Distribution Shift. *arXiv:1911.06970 [cs, stat]*, Dec. 2019. arXiv: 1911.06970. [8](#)
- [36] Boris Ivanovic and Marco Pavone. The Trajectron: Probabilistic Multi-Agent Trajectory Modeling With Dynamic Spatiotemporal Graphs. In *2019 IEEE/CVF International Conference on Computer Vision (ICCV)*, pages 2375–2384, Oct. 2019. ISSN: 2380-7504. [1](#), [2](#)
- [37] Allan Jabri, Andrew Owens, and Alexei Efros. Space-Time Correspondence as a Contrastive Random Walk. *Advances in Neural Information Processing Systems*, 33, 2020. [3](#)
- [38] Yannis Kalantidis, Mert Bulent Sariyildiz, Noe Pion, Philippe Weinzaepfel, and Diane Larlus. Hard Negative Mixing for Contrastive Learning. *arXiv:2010.01028 [cs]*, Oct. 2020. arXiv: 2010.01028. [1](#), [3](#)
- [39] Prannay Khosla, Piotr Teterwak, Chen Wang, Aaron Sarna, Yonglong Tian, Phillip Isola, Aaron Maschinot, Ce Liu, and Dilip Krishnan. Supervised Contrastive Learning. *arXiv:2004.11362 [cs, stat]*, Apr. 2020. arXiv: 2004.11362. [1](#)
- [40] Diederik P. Kingma and Jimmy Ba. Adam: A Method for Stochastic Optimization. *arXiv:1412.6980 [cs]*, Dec. 2014. arXiv: 1412.6980. [5](#)
- [41] Vineet Kosaraju, Amir Sadeghian, Roberto Martín-Martín, Ian Reid, Hamid Rezaeifighi, and Silvio Savarese. Social-BiGAT: Multimodal Trajectory Forecasting using Bicycle-GAN and Graph Attention Networks. *Advances in Neural Information Processing Systems*, 32:137–146, 2019. [1](#), [2](#)
- [42] Ilya Kostrikov, Kumar Krishna Agrawal, Debidatta Dwivedi, Sergey Levine, and Jonathan Tompson. Discriminator-Actor-Critic: Addressing Sample Inefficiency and Reward Bias in Adversarial Imitation Learning. *arXiv:1809.02925 [cs, stat]*, Oct. 2018. arXiv: 1809.02925. [1](#), [2](#)
- [43] Parth Kothari, Sven Kreiss, and Alexandre Alahi. Human trajectory forecasting: A deep learning perspective. *IEEE Transactions on Intelligent Transportation Systems*, 2021. [2](#), [4](#), [5](#), [6](#), [13](#), [14](#)
- [44] Aviral Kumar, Justin Fu, Matthew Soh, George Tucker, and Sergey Levine. Stabilizing Off-Policy Q-Learning via Bootstrapping Error Reduction. *Advances in Neural Information Processing Systems*, 32:11784–11794, 2019. [8](#)
- [45] Sascha Lange, Thomas Gabel, and Martin Riedmiller. Batch Reinforcement Learning. In Marco Wiering and Martijn van Otterlo, editors, *Reinforcement Learning: State-of-the-Art, Adaptation, Learning, and Optimization*, pages 45–73. Springer, Berlin, Heidelberg, 2012. [8](#)
- [46] Michael Laskey, Jonathan Lee, Roy Fox, Anca Dragan, and Ken Goldberg. DART: Noise Injection for Robust Imitation Learning. In *Conference on Robot Learning*, pages 143–156. PMLR, Oct. 2017. ISSN: 2640-3498. [1](#), [2](#)
- [47] Namhoon Lee, Wongun Choi, Paul Vernaza, Christopher B. Choy, Philip H. S. Torr, and Manmohan Chandraker. DE-SIRE: Distant Future Prediction in Dynamic Scenes with Interacting Agents. *arXiv:1704.04394 [cs]*, Apr. 2017. arXiv: 1704.04394. [1](#)
- [48] Alon Lerner, Yiorgos Chrysanthou, and Dani Lischinski. Crowds by Example. *Computer Graphics Forum*, 26(3):655–664, 2007. [5](#), [6](#)
- [49] Sergey Levine, Aviral Kumar, George Tucker, and Justin Fu. Offline Reinforcement Learning: Tutorial, Review, and Perspectives on Open Problems. *arXiv:2005.01643 [cs, stat]*, May 2020. arXiv: 2005.01643. [8](#)
- [50] Jiachen Li, Fan Yang, Masayoshi Tomizuka, and Chihoo Choi. EvolveGraph: Multi-Agent Trajectory Prediction with Dynamic Relational Reasoning. *arXiv:2003.13924 [cs]*, Oct. 2020. arXiv: 2003.13924. [1](#), [2](#), [4](#)

- [51] Lingyun Luke Li, Bin Yang, Ming Liang, Wenyuan Zeng, Mengye Ren, Sean Segal, and Raquel Urtasun. End-to-end Contextual Perception and Prediction with Interaction Transformer. *arXiv:2008.05927 [cs]*, Aug. 2020. arXiv: 2008.05927. [4](#)
- [52] Yuejiang Liu, An Xu, and Zichong Chen. Map-based Deep Imitation Learning for Obstacle Avoidance. In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 8644–8649, Oct. 2018. ISSN: 2153-0866. [1](#)
- [53] Lajanugen Logeswaran and Honglak Lee. An efficient framework for learning sentence representations. *arXiv:1803.02893 [cs]*, Mar. 2018. arXiv: 1803.02893. [1](#), [3](#)
- [54] Matthias Lubner, Johannes A. Stork, Gian Diego Tipaldi, and Kai O Arras. People tracking with human motion predictions from social forces. In *2010 IEEE International Conference on Robotics and Automation*, pages 464–469, May 2010. ISSN: 1050-4729. [2](#)
- [55] Yuping Luo, Huazhe Xu, and Tengyu Ma. Learning Self-Correctable Policies and Value Functions from Demonstrations with Negative Sampling. *arXiv:1907.05634 [cs, stat]*, Oct. 2019. arXiv: 1907.05634. [2](#)
- [56] Ramin Mehran, Alexis Oyama, and Mubarak Shah. Abnormal crowd behavior detection using social force model. In *2009 IEEE Conference on Computer Vision and Pattern Recognition*, pages 935–942, June 2009. ISSN: 1063-6919. [2](#)
- [57] Tomas Mikolov, Ilya Sutskever, Kai Chen, Greg S Corrado, and Jeff Dean. Distributed Representations of Words and Phrases and their Compositionality. In *Advances in Neural Information Processing Systems 26*, pages 3111–3119. Curran Associates, Inc., 2013. [1](#), [2](#), [4](#)
- [58] Ishan Misra and Laurens van der Maaten. Self-Supervised Learning of Pretext-Invariant Representations. *arXiv:1912.01991 [cs]*, Dec. 2019. arXiv: 1912.01991. [1](#)
- [59] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A. Rusu, Joel Veness, Marc G. Bellemare, Alex Graves, Martin Riedmiller, Andreas K. Fiedjeland, Georg Ostrovski, Stig Petersen, Charles Beattie, Amir Sadik, Ioannis Antonoglou, Helen King, Dharshan Kumaran, Daan Wierstra, Shane Legg, and Demis Hassabis. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, Feb. 2015. Number: 7540 Publisher: Nature Publishing Group. [7](#)
- [60] A. Mohamed, K. Qian, M. Elhoseiny, and C. Claudel. Social-STGCNN: A Social Spatio-Temporal Graph Convolutional Neural Network for Human Trajectory Prediction. In *2020 IEEE/CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, pages 14412–14420, June 2020. ISSN: 2575-7075. [5](#), [6](#), [13](#)
- [61] Andrew Y. Ng and Stuart Russell. Algorithms for Inverse Reinforcement Learning. In *in Proc. 17th International Conf. on Machine Learning*, pages 663–670. Morgan Kaufmann, 2000. [2](#)
- [62] M. Niedoba, Henggang Cui, K. Luo, Darshan Hegde, Fang-Chieh Chou, and Nemanja Djuric. Improving Movement Prediction of Traffic Actors using Off-road Loss and Bias Mitigation, 2019. [2](#)
- [63] Aaron van den Oord, Yazhe Li, and Oriol Vinyals. Representation Learning with Contrastive Predictive Coding. *arXiv:1807.03748 [cs, stat]*, Jan. 2019. arXiv: 1807.03748. [1](#), [3](#)
- [64] Matteo Pagliardini, Prakhhar Gupta, and Martin Jaggi. Unsupervised Learning of Sentence Embeddings Using Compositional n-Gram Features. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 528–540, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. [3](#)
- [65] Taesung Park, Alexei A. Efros, Richard Zhang, and Jun-Yan Zhu. Contrastive Learning for Unpaired Image-to-Image Translation. *arXiv:2007.15651 [cs]*, July 2020. arXiv: 2007.15651. [1](#), [3](#)
- [66] Stefano Pellegrini, Andreas Ess, and Luc Van Gool. Improving Data Association by Joint Modeling of Pedestrian Trajectories and Groupings. In *Computer Vision – ECCV 2010*, Lecture Notes in Computer Science, pages 452–465, Berlin, Heidelberg, 2010. Springer. [2](#), [5](#), [6](#)
- [67] Matthew Peters, Mark Neumann, Mohit Iyyer, Matt Gardner, Christopher Clark, Kenton Lee, and Luke Zettlemoyer. Deep Contextualized Word Representations. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2227–2237, New Orleans, Louisiana, June 2018. Association for Computational Linguistics. [2](#), [4](#)
- [68] Dean A. Pomerleau. ALVINN: An Autonomous Land Vehicle in a Neural Network. In D. S. Touretzky, editor, *Advances in Neural Information Processing Systems*, pages 305–313, 1989. [2](#)
- [69] Senthil Purushwalkam and Abhinav Gupta. Demystifying Contrastive Self-Supervised Learning: Invariances, Augmentations and Dataset Biases. *arXiv:2007.13916 [cs]*, July 2020. arXiv: 2007.13916. [3](#)
- [70] Siddharth Reddy, Anca D. Dragan, and Sergey Levine. SQL: Imitation Learning via Reinforcement Learning with Sparse Rewards. In *International Conference on Learning Representations*, Sept. 2019. [1](#)
- [71] Daniela Ridel, Eike Rehder, Martin Lauer, Christoph Stiller, and Denis Wolf. A Literature Review on the Prediction of Pedestrian Behavior in Urban Scenarios. In *2018 21st International Conference on Intelligent Transportation Systems (ITSC)*, pages 3105–3112, Nov. 2018. ISSN: 2153-0017. [2](#)
- [72] Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stefanie Jegelka. Contrastive Learning with Hard Negative Samples. *arXiv:2010.04592 [cs, stat]*, Oct. 2020. arXiv: 2010.04592. [3](#)
- [73] Stephane Ross and Drew Bagnell. Efficient Reductions for Imitation Learning. In *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, pages 661–668. JMLR Workshop and Conference Proceedings, Mar. 2010. ISSN: 1938-7228. [1](#), [2](#)
- [74] Stephane Ross, Geoffrey Gordon, and Drew Bagnell. A Reduction of Imitation Learning and Structured Prediction to No-Regret Online Learning. In *Proceedings of the Fourteenth International Conference on Artificial Intelligence and Statistics*, pages 627–635. JMLR Workshop and Con-

- ference Proceedings, June 2011. ISSN: 1938-7228. [1](#), [2](#)
- [75] Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Darius M Gavrilu, and Kai O Arras. Human motion trajectory prediction: a survey. *The International Journal of Robotics Research*, 39(8):895–935, July 2020. Publisher: SAGE Publications Ltd STM. [2](#)
- [76] Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. SoPhie: An Attentive GAN for Predicting Paths Compliant to Social and Physical Constraints. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 1349–1358, 2019. tex.ids: sadeghianSoPhieAttentive-GAN2019a ISSN: 2575-7075. [1](#), [2](#)
- [77] Tim Salzmann, Boris Ivanovic, Punarjay Chakravarty, and Marco Pavone. Trajectron++: Dynamically-Feasible Trajectory Forecasting with Heterogeneous Data. In Andrea Vedaldi, Horst Bischof, Thomas Brox, and Jan-Michael Frahm, editors, *Computer Vision – ECCV 2020*, Lecture Notes in Computer Science, pages 683–700, Cham, 2020. Springer International Publishing. [1](#), [5](#), [6](#), [13](#)
- [78] Samaneh Hosseini Semnani, Hugh Liu, Michael Everett, Anton de Ruiter, and Jonathan P. How. Multi-Agent Motion Planning for Dense and Dynamic Environments via Deep Reinforcement Learning. *IEEE Robotics and Automation Letters*, 5(2):3221–3226, Apr. 2020. Conference Name: IEEE Robotics and Automation Letters. [7](#), [13](#)
- [79] Pierre Sermanet, Corey Lynch, Yevgen Chebotar, Jasmine Hsu, Eric Jang, Stefan Schaal, and Sergey Levine. Time-Contrastive Networks: Self-Supervised Learning from Video. *arXiv:1704.06888 [cs]*, Mar. 2018. arXiv: 1704.06888. [3](#)
- [80] Abhishek Sinha, Kumar Ayush, Jiaming Song, Burak Uzkent, Hongxia Jin, and Stefano Ermon. Negative Data Augmentation. In *International Conference on Learning Representations*, Sept. 2020. [2](#)
- [81] Kihyuk Sohn. Improved Deep Metric Learning with Multi-class N-pair Loss Objective. In D. D. Lee, M. Sugiyama, U. V. Luxburg, I. Guyon, and R. Garnett, editors, *Advances in Neural Information Processing Systems 29*, pages 1857–1865. Curran Associates, Inc., 2016. [1](#)
- [82] Jiaming Song and Stefano Ermon. Multi-label Contrastive Predictive Coding. *arXiv:2007.09852 [cs, stat]*, July 2020. arXiv: 2007.09852. [3](#)
- [83] Jianhua Sun, Qinhong Jiang, and Cewu Lu. Recursive Social Behavior Graph for Trajectory Prediction. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 660–669, 2020. tex.ids: sunRecursiveSocialBehavior2020a ISSN: 2575-7075. [1](#), [2](#)
- [84] L. Sun, Z. Yan, S. M. Mellado, M. Hanheide, and T. Duckett. 3DOF Pedestrian Trajectory Prediction Learned from Long-Term Autonomous Mobile Robot Deployment Data. In *2018 IEEE International Conference on Robotics and Automation (ICRA)*, pages 5942–5948, May 2018. ISSN: 2577-087X. [13](#)
- [85] Wen Sun, Arun Venkatraman, Geoffrey J. Gordon, Byron Boots, and J. Andrew Bagnell. Deeply AggreVaTeD: Differentiable Imitation Learning for Sequential Prediction. In *International Conference on Machine Learning*, pages 3309–3318. PMLR, July 2017. ISSN: 2640-3498. [1](#), [2](#)
- [86] Peter Trautman and Andreas Krause. Unfreezing the robot: Navigation in dense, interacting crowds. In *2010 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 797–803, Oct. 2010. ISSN: 2153-0866. [2](#)
- [87] Jur van den Berg, Stephen J. Guy, Ming Lin, and Dinesh Manocha. Reciprocal n-Body Collision Avoidance. In Cédric Pradalier, Roland Siegwart, and Gerhard Hirzinger, editors, *Robotics Research*, Springer Tracts in Advanced Robotics, pages 3–19, Berlin, Heidelberg, 2011. Springer. [2](#)
- [88] Anirudh Vemula, Katharina Muelling, and Jean Oh. Social Attention: Modeling Attention in Human Crowds. *arXiv:1710.04689 [cs]*, Oct. 2017. arXiv: 1710.04689. [1](#), [2](#)
- [89] Ruohan Wang, Carlo Ciliberto, Pierluigi Amadori, and Yiannis Demiris. Random Expert Distillation: Imitation Learning via Expert Policy Support Estimation. *arXiv:1905.06750 [cs, stat]*, June 2019. arXiv: 1905.06750. [1](#), [2](#)
- [90] Zhirong Wu, Yuanjun Xiong, Stella Yu, and Dahua Lin. Unsupervised Feature Learning via Non-parametric Instance Discrimination. In *2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 3733–3742, June 2018. ISSN: 2575-7075. [1](#), [2](#), [4](#)
- [91] Francesco Zanlungo, Tetsushi Ikeda, and Takayuki Kanda. Social force model with explicit collision prediction. *EPL (Europhysics Letters)*, 93(6):68005, Mar. 2011. Publisher: IOP Publishing. [2](#)
- [92] Wenyuan Zeng, Shenlong Wang, Renjie Liao, Yun Chen, Bin Yang, and Raquel Urtasun. DSDNet: Deep Structured self-Driving Network. *arXiv:2008.06041 [cs]*, Aug. 2020. arXiv: 2008.06041. [2](#), [5](#)
- [93] Brian D. Ziebart, Andrew Maas, J. Andrew Bagnell, and Anind K. Dey. Maximum entropy inverse reinforcement learning. In *Proceedings of the 23rd national conference on Artificial intelligence - Volume 3, AAAI’08*, pages 1433–1438, Chicago, Illinois, July 2008. AAAI Press. [2](#)



## A. Benchmark Details

### A.1. Trajectory Forecasting Details

We validate our method on two forecasting benchmarks, the ETH & UCY benchmark and the Trajnet++ benchmark. The former is a *general* benchmark containing pedestrian trajectories in a variety of scenarios. The latter one is based on a curated meta-dataset that consists of *interacting* scenarios selected from several publicly available subsets, such as WildTrack [11], L-CAS [84] and CFF [4]. This benchmark has been used in a series of recent competitions.

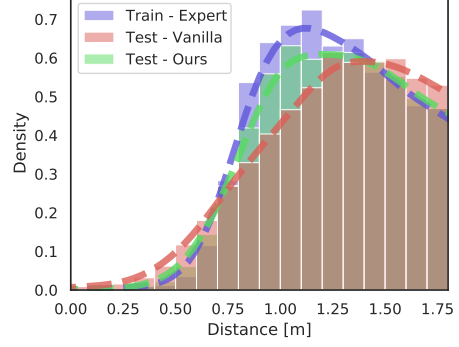
Our evaluation protocol follows the previous work [43, 60, 77]. On the ETH & UCY datasets, we use the leave-one-out approach, where forecasting models are trained on four sub-datasets and tested on the held-out fifth. On the Trajnet++ dataset, we use the official split of the training and test set. One common feature of recent models like Social-STGCNN and Trajectron++ is that the prediction of the primary agent is only conditioned on the states of neighboring agents up to the observation time  $t_o$  but not on any steps from  $t_o$  to  $t_p$  that have already been predicted, *i.e.*,  $s_{t_p+1}^i = f(s_{1:t_p}^i, s_{1:t_o}^{M \setminus i})$ . While this design choice accelerates training and inference, it makes the forecasting model unaware of the latest states of the nearby agents and causes notoriously high collision rate at long horizon. As such, our evaluation of collision rate for the Social-STGCNN and Trajectron++ is focused on the first four prediction steps where the models still have access to relatively up-to-date information of the surrounding neighbors. Yet, for the Trajnet++ models that perform fully joint prediction in a recurrent manner,  $s_{t_p+1}^i = f(s_{1:t_p}^i, s_{1:t_p}^{M \setminus i})$ , we measure the collision rate over the entire prediction horizon.

### A.2. Reinforcement Learning Details

The original SARL policy [12] requires a linear motion model as well as imitation pre-training to accomplish the reinforcement learning task from sparse reward feedback. These hand-crafted components, however, introduce extra assumptions over the crowd navigation task and make it hard to analyze the sample efficiency of an RL algorithm. To tease apart the effect of our proposed method, we adopt the following dense reward function [78] for the model-free Rainbow algorithm,

$$r(s_t, a_t) = \alpha(d_g^{t-1} - d_g^t) + \begin{cases} -1 & \text{if } d_m^t < 0 \\ 10d_m^t - 1 & \text{else if } d_m^t < 0.1 \\ 1 & \text{else if goal is reached} \\ 0 & \text{otherwise} \end{cases} \quad (9)$$

where  $d_g$  is the Euclidean distance between the robot and its goal,  $\alpha = 0.08$  is a control parameter. Other settings are



**Figure 7:** Histogram of the human-robot distance on the crowd navigation task with imitation learning. The vanilla method suffers from the problem of covariate shift in closed-loop sequential predictions, whereas our method results in a much smaller gap between the training and test state distributions.

kept the same as Section 4.3.

## B. Covariate Shift

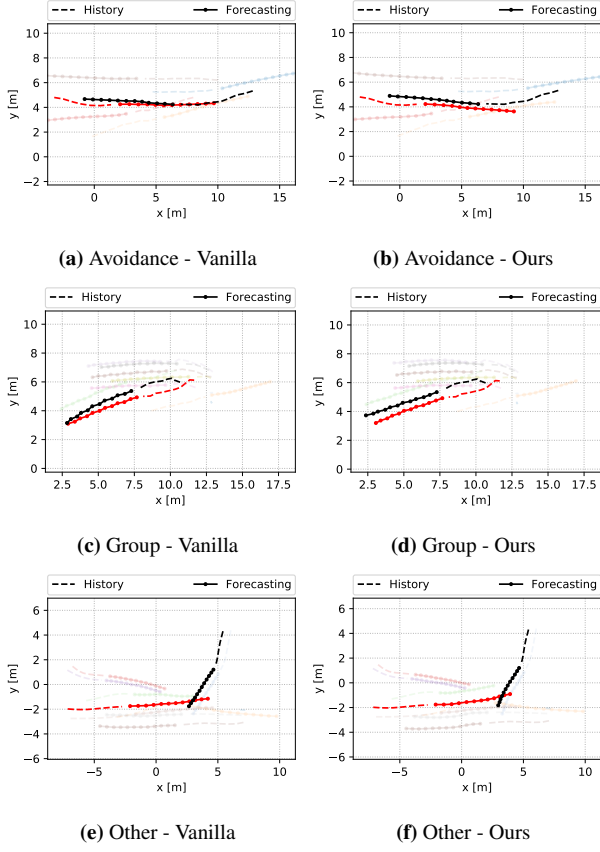
To further understand the effect of our learning method on closed-loop sequential predictions, we conduct a detailed analysis of the state distribution at test time on the crowd navigation task with imitation learning. Here, we focus on minimum distance between the robot and the other surrounding agents at each frame and collect a set of robot-human distances from 500 test episodes.

Figure 7 shows the histogram of human-robot distance with different policies. As expected, the density of short-distance states under the expert demonstrator is close to zero, which reflects the expert’s high degree of social awareness as well as confirms the lack of dangerous occurrences in the demonstration data. Compared with the training distribution, the test distribution induced by the model from the vanilla imitation learning yields a clear distinction. It exhibits lower density at the distance around 1.0 [m], but higher density in the dangerous regime, *e.g.*, distance smaller than 0.5 [m]. On the contrary, our method results in a test state distribution almost overlapping with the training one over the dangerous states. These results verify that, given the same distribution of the initial state, the model trained with our method visits dangerous states much less frequently and functions in closed-loop operation much more robustly.

## C. Qualitative Results

In addition to quantitative comparison, Figure 8 shows the qualitative results of different learning methods in three different interacting scenarios on the Trajnet++ benchmark. The Directional-LSTM [43] trained with the vanilla predictive learning outputs colliding trajectories between the





**Figure 8:** Qualitative results of Directional-LSTM [43] models trained with different methods in three *interacting* test cases on the Trajnet++ benchmark [43]. The vanilla method leads to collisions between the primary (black) and the nearby agent (red) at the 4th, 12th, and 10th predicted step on the *Avoidance*, *Group* and *Other* case respectively, whereas our method outputs collision-free trajectories over the whole prediction horizon.

primary agent and its neighbors in these dense scenes. In contrast, our method outputs more socially compliant solutions: in the *Group* scenario, our predicted trajectory for the primary agent stays in the middle of two other neighbors at all time steps instead of sliding towards either of them. In the *Avoidance* scenario, our method adjusts the trajectories of both the primary and the opposite agent cooperatively. Similarly, in the *Other* scenario where pedestrians come from almost orthogonal directions, our method jointly twists the trajectories of these interactive agents, enabling each of them to pass the crowded spot smoothly.