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 multi-agent problems, such as human motion forecasting and robot navigation in crowds. Despite promising progress, existing representations learned with neural networks still struggle to generalize in closed-loop predictions (e.g., output colliding trajectories). This issue largely arises from the non-i.i.d. nature of sequential prediction in conjunction with ill-distributed training data. Intuitively, if the training data only comes from human behaviors in safe spaces, i.e., from “positive” examples, it is difficult for learning algorithms to capture the notion of “negative” examples like collisions. In this work, we aim to address this issue by explicitly modeling negative examples through self-supervision: (i) we introduce a social contrastive loss that regularizes the extracted motion representation by discerning the ground-truth positive events from synthetic negative ones; (ii) we construct informative negative samples based on our prior knowledge of rare but dangerous circumstances. Our method substantially reduces the collision rates of recent trajectory forecasting, behavioral cloning and reinforcement learning algorithms, outperforming state-of-the-art methods on several benchmarks. Our code is available at https://github.com/vita-epfl/social-nce.

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 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, 27, 41, 43, 47, 57, 84, 90, 96] to learn socially-aware motion representations for multi-agent problems such as trajectory forecasting [3, 33, 54, 85] and robot navigation [15, 18, 19]. Yet existing methods still output unacceptable solutions (e.g., collisions) from time to time in closed-loop predictions, which raises significant safety concerns for real-world deployment.

One key difficulty in learning robust motion representation stems from the common shortage of critical data. Very often, the training data is only collected from human behaviors in safe scenarios. The lack of near-accident examples poses a significant challenge to the discovery of social norms from data. As a consequence, prediction errors made by the learned model may accumulate over time, gradually create a discrepancy between the training and test state distributions, and eventually cause catastrophic errors [22, 25, 81].

Previous methods attempt to mitigate this issue through interactive data collections, such as expert queries [26, 53, 60, 82, 92] and additional experiments [12, 40, 48, 77, 97]. Unfortunately, these methods are not only costly and tedious but often impractical for forecasting problems, since human behaviors can hardly be intervened at scale for data collection. These shortcomings motivate us to explore an alternative approach: given a fixed training dataset, can we learn a robust motion representation by exploiting our prior knowledge through self-supervision?

To this end, we propose a social contrastive learning method built with negative data augmentation (Figure 1). Our main idea is to promote robust representations by learning from the opposites [74]. Intuitively, an effective way to elucidate the social norms that give rise to the “positive” examples is to explicitly portray the opposite “nega-
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, 37, 64, 95]. While these hand-crafted models have been successfully applied to various tasks [23, 24, 30, 62, 99], they often struggle to capture the strong interactions among agents in complex scenes [83]. 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 [72, 94].

Covariate Shift. The problem of covariate shift was observed back in [75] and has been a persistent challenge for sequential prediction problems [25, 81]. One practical solution is to actively query experts [53, 82, 92], which has been shown effective for behavioral cloning but hardly applicable to forecasting problems [49, 79, 83]. Inverse reinforcement learning methods [1, 12, 40, 48, 67, 97, 101] 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 [28].

Another line of work introduces additional loss terms penalizing the predictions that lead to undesirable events, such as collisions and off-road trajectories [10, 68]. 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, [63, 100] propose to learn a robust value (or cost) function by making use of negative samples. Our method differs from theirs in two key aspects: [63, 100] change the task loss that directly affects
(and potentially biases) the model output, whereas our goal 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.

**Contrastive Learning.** Contrastive learning was proposed in [35] to learn an embedding space such that a simple similarity measure can approximate high-level semantic relations. This approach has recently achieved stunning results in a broad spectrum of areas, including computer vision [16, 36], natural language understanding [7, 61, 70], image synthesis [71] and robotics [87]. Some detailed design choices, such as positive and negative sampling, often play a critical role to the empirical success of contrastive methods [21, 45, 76, 80]. 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 longstanding issue [8, 38, 59, 93]. This issue is particularly concerning in sequential predictions such as behavior forecasting and autonomous navigation. Very often, the training data only contains “positive” examples collected from safe states without any dangerous “negative” occurrences. The severely imbalanced training distribution 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 combined with negative data augmentation. We will first briefly describe the basic idea of contrastive learning and then present a social contrastive loss. 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 [11]. Recent contrastive learning methods often adopt the principle of noise contrastive estimation in an embedding space, namely the InfoNCE loss [29, 34, 69], to train an encoder:

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

where the encoded query $q$ is brought close to one positive key $k_0 = k^+$ and pushed apart from $N$ negative keys $\{k_1, \ldots, k_N\}$. $\tau$ 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 maximizing the lower bound on the mutual information between the raw input and the latent representation [69]. 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 [16, 21, 44, 45, 76, 80, 89].

#### 3.2. Social NCE

Consider a trajectory forecasting problem in crowded spaces as an example. Let $s_i^t = (x_i^t, y_i^t)$ denote the position of agent $i$ at time $t$ and $s_t = \{s_1^t, \ldots, s_M^t\}$ denote the joint state of $M$ agents in the scene. Given a sequence of history observations $\{s_1, \ldots, s_t\}$, the task is to predict future trajectories of all agents $\{s_{t+1}, \ldots, s_T\}$ 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 $h_i^t$ with respect to agent $i$. 

![Figure 3: Social contrastive learning in the multi-agent context. Given a scenario that contains a primary agent of interest (blue) and multiple neighboring agents in the vicinity (gray), our Social-NCE loss encourages the extracted motion representation, in an embedding space, to be close to a positive future event and apart from some synthetic negative events that could have caused collisions or discomforts.](image-url)
and the decoder $g(\cdot)$ subsequently rolls out its future trajectory $\hat{s}^{i}_{t+1:T}$:

$$
\begin{align*}
    h^i_t &= f(s^{i}_{1:t}, i), \\
    \hat{s}^{i}_{t+1:T} &= g(h^i_t).
\end{align*}
$$

(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.,

$$
\begin{align*}
    z^i_t &= f_S(h^i_{t-1}, s^i_t), \\
    h^i_t &= f_I(z^i_t, i).
\end{align*}
$$

(3)

where $z^i_t$ is the latent representation of agent $i$ given the observation of its own state at time $t$ and $z^i_t = \{z^i_1, \ldots, z^i_M\}$.

A variety of architectures have been explored for each modules and validated on accuracy measures [3, 57, 58]. Nevertheless, their robustness remains an open issue. Several recent works [10, 49] 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 adapted to socially-aware motion representations. 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 $q = \psi(h^i_t)$, where $\psi(\cdot)$ is an MLP projection head.
- **key**: embedding of a future event $k = \phi(s^i_{t+\delta t}, \delta t)$, where $\phi(\cdot)$ is an event encoder modeled by an MLP; $s^i_{t+\delta t}$ is a sampled spatial location and $\delta t > 0$ is the sampling horizon.

By tuning $\delta t \in \Lambda$ in a range, e.g., $\Lambda = \{1, \ldots, 4\}$, 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., $\phi(s^i_{t+\delta t})$.

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 [16, 36, 98], 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:

$$
L_{SocialNCE} = - \log \frac{\exp(\psi(h^i_t) \cdot \phi(s^i_{t+\delta t}, \delta t)/\tau)}{\sum_{\delta t' \in \Lambda} \sum_{n=0}^{N} \exp(\psi(h^i_t) \cdot \phi(s^{\cdot,n}_{t+\delta t'}, \delta t)/\tau)}.
$$

(4)

The full training objective is a weighted combination of the conventional task loss, e.g., mean squared error (MSE) or negative log-likelihood (NLL) for trajectory forecasting, and the proposed social contrastive loss:

$$
L(f, g, \psi, \phi) = L_{task}(f, g) + \lambda L_{SocialNCE}(f, \psi, \phi),
$$

(5)

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

### 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 [16, 32, 36, 65, 73, 98]. 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 to the multimodal nature 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 social 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 $t + \delta t$.

$$
\hat{s}^{i,n}_{t+\delta t} = s^j_{t+\delta t} + \Delta s^i_p + \epsilon,
$$

(6)

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

negative samples are synthesized. We also add random perturbations to each sampled location \( \epsilon \sim \mathcal{N}(0; \sigma, I) \), where \( \sigma \) 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 \( t + \delta t \),

\[ s_{i, t+\delta t}^{i+} = s_{i, t+\delta t}^i + \epsilon. \]  

(7)

The key intuition behind our method is that the conventional learning approaches only focus on replicating “positive” behaviors in normal states, without the need to understand the consequence of “negative” ones. In contrast, by using Social-NCE in tandem with our proposed sampling strategy, we actively enforce the extracted motion representation \( h_i^t \) to contain necessary information for identifying the scenarios that could have led to catastrophic outcomes. This subtle but essential difference enables the model to learn a significantly more robust representation 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 the social 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 to 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 [46].

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 [55, 72] and the Trajnet++ interaction-centric benchmark [49]. 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 [33, 49, 100], 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, 33, 66, 85]. We consider the following two state-of-the-art multi-modal forecasting models as baselines:

- Social-STGCNN [66]: a spatio-temporal convolutional network with graph representations of pedestrian trajectories.
- Trajectron++ [85]: 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++ [49], 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 [66] | Trajectron++ [85] |
|-----------|--------------------|-------------------|
|           | Vanilla | Ours | Gain | Vanilla | Ours | Gain |
|           | FDE ↓ | COL ↓ | FDE ↓ | COL ↓ | FDE ↓ | COL ↓ | FDE ↓ | COL ↓ | FDE ↓ | COL ↑ |
| ETH       | 1.223 | 1.33  | 1.224 | 0.61  | 54.1% | 0.810 | 1.16  | 0.791 | 0.00  | 100.0% |
| Hotel     | 0.687 | 3.82  | 0.678 | 3.35  | 12.3% | 0.184 | 0.84  | 0.177 | 0.38  | 54.6%  |
| Univ      | 0.912 | 9.11  | 0.879 | 6.44  | 29.3% | 0.450 | 3.38  | 0.435 | 3.08  | 8.9%   |
| Zara1     | 0.525 | 2.27  | 0.515 | 1.02  | 55.1% | 0.320 | 0.46  | 0.330 | 0.18  | 61.5%  |
| Zara2     | 0.480 | 6.86  | 0.482 | 3.37  | 50.9% | 0.250 | 1.03  | 0.255 | 0.99  | 3.3%   |
| Average   | 0.765 | 4.70  | 0.756 | 2.96  | 37.0% | 0.403 | 1.37  | 0.398 | 0.93  | 45.7%  |

Table 1: Comparison between Social-NCE and the vanilla predictive learning for two recent multi-modal trajectory forecasting models, Social-STGCNN [66] and Trajectron++ [85], on the ETH [72] / UCY [55] 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 [49] |
|----------|----------------|-----------------------|
|          | Vanilla | Ours | Gain | Vanilla | Ours | Gain |
|          | FDE ↓ | COL ↓ | FDE ↓ | COL ↓ | FDE ↓ | COL ↓ | FDE ↓ | COL ↓ | FDE ↓ | COL ↑ |
| Avoidance| 1.23   | 13.12 | 1.23  | 10.83  | 17.5% | 1.33  | 12.92 | 1.33  | 10.42 | 19.4% |
| Group    | 0.97   | 7.44  | 0.97  | 4.65   | 37.5% | 1.05  | 6.51  | 1.05  | 5.89  | 9.5%  |
| Overall  | 1.14   | 6.44  | 1.14  | 5.31   | 17.6% | 1.22  | 5.49  | 1.22  | 4.59  | 16.4% |

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 [49]. 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 [49]: a LSTM-based model with an interaction module sharing 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 [15, 18, 19]. We use an open-sourced simulator of crowd navigation [15], 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 [15], which quantifies the performance of a policy using three metrics: navigation time, collision rate, and the accumulated reward as follows:

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

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 ↑ | Time (s) ↓ | Collision (%) ↓ |
|--------|----------|------------|-----------------|
| Vanilla | 0.28 ± 0.01 | 10.31 ± 0.07 | 11.11 ± 1.45 |
| Random | 0.24 ± 0.02 | 10.32 ± 0.12 | 18.60 ± 4.69 |
| Ours   | **0.32 ± 0.01** | **10.33 ± 0.07** | **3.40 ± 1.36** |

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 [15] 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 policy. 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 [15].

### 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 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 [39], a state-of-the-art model-free RL method, as baseline and follow the architecture of value-based SARL policy [15] 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 [39]. In order to isolate the impact of Social-NCE, we use the dense reward function proposed in [86], which

![Learning curves of Rainbow DQN [39] 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.](image-url)

Figure 6: Learning curves of Rainbow DQN [39] 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    | **91.6%** | **84.6%** | **79.2%** | **69.0%** |

Table 4: Offline RL normalized scores attained by the vanilla rainbow and Social-NCE agents (higher is better). Normalized score is calculated as: $100 \times (\text{agent score} - \text{random play score}) / (\text{optimal agent score} - \text{random play score})$. Our Social-NCE consistently facilitates the recovery of a near-optimal policy and is particularly advantageous in the low-data regime.
We use temperature sampled from the dataset, where $f$ line Rainbow policy on $N$ made up of 30k episodes of trials in total. We train an off-policy exploration contains 5k episodes, and the full dataset is carried out using online RL checkpoint models trained from scratch, (ii) multiple free explorations are in the learning process: (i) 10k episodes are collected during online RL training from scratch, (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 $\lambda = 0.1$.

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

Table 5: Social-NCE for imitation learning with different sampling horizons. 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.

### 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 [56]. 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, 31, 42, 51, 52].

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 method for learning socially-aware motion representations. The proposed Social-NCE loss, combined with our sampling strategy, significantly boosts the performance of recent human trajectory forecasting and crowd navigation algorithms. Our result suggests that learning from the opposite by means of negative data augmentations can be a promising alternative to the traditional interactive data collection for robust sequential predictions.

### Acknowledgments

This work is supported by the Swiss National Science Foundation under the Grant 200021-L92326. We thank Parth Kothari, Yifan Sun, Taylor Mordan, Mohammadhossein Bahari, Lorenzo Bertoni and Sven Kreiss 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 Kodak, 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] Samaneh Azadi, Catherine Olsson, Trevor Darrell, Ian Goodfellow, and Augustus Odena. Discriminator Rejection Sampling. In International Conference on Learning Representations, Sept. 2018. 3

[9] Mohammadhossein Bahari, Nejjar Ismail, and Alexandre Alahi. Injecting knowledge in data-driven vehicle trajectory predictors. Transportation Research Part C, 2021. 2

[10] 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

[11] 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

[12] Kianne Brantley, Wen Sun, and Mikael Henaff. Disagreement-Regularized Imitation Learning. In International Conference on Learning Representations, Sept. 2019. 1, 2

[13] Yuning Chai, Benjamin Sapp, Mayank Bansal, and Dragomir Anguelov. MultiPath: Multiple Probabilistic Anchor Trajectory Hypotheses for Behavior Prediction. In Conference on Robot Learning, pages 86–99. PMLR, May 2020. 2

[14] 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 Unsupervised Pedestrian Detection. In 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 5030–5039, June 2018. ISSN: 2575-7075. 13

[15] 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

[16] 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. 2, 3, 4

[17] Xinlei Chen and Kaiming He. Exploring Simple Siamese Representation Learning. In Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, pages 15750–15758, 2021. 2

[18] 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

[19] 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

[20] Chiho Choi, Srikanth Malla, Abhishek Patil, and Joon Hee Choi. DROGON: A Trajectory Prediction Model based on Intention-Conditioned Behavior Reasoning. arXiv:1908.00024 [cs], Nov. 2020. arXiv: 1908.00024. 2

[21] Ching-Yao Chuang, Joshua Robinson, Lin Yen-Chen, Antonio Torralba, and Stefanie Jegelka. Debias Contrastive Learning. arXiv:2007.00224 [cs, stat], July 2020. arXiv: 2007.00224. 3

[22] 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

[23] Pasquale Coscia, Francesco Castaldo, Francesco AN Palmieri, Alexandre Alahi, Silvio Savarese, and Lamberto Ballan. Long-term path prediction in urban scenarios using circular distributions. Journal on Image and Vision Computing (IJCV), 2018. 2

[24] 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

[12] Kiante Brantley, Wen Sun, and Mikael Henaff. Disagreement-Regularized Imitation Learning. In International Conference on Learning Representations, Sept. 2019. 1, 2
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

Joshua Robinson, Ching-Yao Chuang, Suvrit Sra, and Stephanie Jegelka. Contrastive Learning with Hard Negative Samples. arXiv:2010.04592 [cs, stat], Oct. 2020. arXiv: 2010.04592. 3

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

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 Conference Proceedings, June 2011. ISSN: 1938-7228. 1, 2

Andrey Rudenko, Luigi Palmieri, Michael Herman, Kris M Kitani, Dariu M Gavrila, 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

Amir Sadeghian, Vineet Kosaraju, Ali Sadeghian, Noriaki Hirose, Hamid Rezatofighi, and Silvio Savarese. SoPhie: An Attention 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, Oct. 2019. 2

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

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

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

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

Jiaming Song and Stefano Ermon. Multi-label Contrastive Predictive Coding. arXiv:2007.09852 [cs, stat], July 2020. arXiv: 2007.09852. 3

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. 1

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

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

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks. arXiv:1312.6199 [cs], Dec. 2013. arXiv: 1312.6199. 3

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

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

Anirudh Vemula, Katharina Muellling, and Jean Oh. Social Attention: Modeling Attention in Human Crowds. arXiv:1710.04689 [cs], Oct. 2017. arXiv: 1710.04689. 1, 2

Ruohan Wang, Carlo Ciliberto, Pierlugi 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

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. 2, 4

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

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

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 [14], L-CAS [91] and CFF [4]. This benchmark has been used in a series of recent competitions.

Our evaluation protocol follows the previous work [49, 66, 85]. 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 \( t_0 \) but not on any steps from \( t_0 \) to \( t_p \) that have already been predicted, i.e., \( s^i_{t_p+1} = f(s^i_{1:t_p}, s^M_{1:t_0}) \). 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, \( s^i_{t_p+1} = f(s^i_{1:t_p}, s^i_{1:t_0}, s_{t_p}^M, ...) \), we measure the collision rate over the entire prediction horizon.

A.2. Reinforcement Learning Details

The original SARL policy [15] 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 [86] for the model-free Rainbow algorithm,

\[
\begin{align*}
    r(s_t, a_t) &= \alpha (d^{-1}_g - d^0_g) \\
    &+ \begin{cases} 
    -1 & \text{if } d^{t}_m < 0 \\
    10d^{t}_m - 1 & \text{else if } d^{t}_m < 0.1 \\
    1 & \text{else if goal is reached} \\
    0 & \text{otherwise}
    \end{cases} 
\end{align*}
\]

(9)

where \( d_g \) is the Euclidean distance between the robot and its goal, \( \alpha = 0.08 \) is a control parameter. Other settings are 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 [49] trained with the vanilla predictive learning outputs colliding trajectories between the
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.