TU Delft

Delft University of Technology

Document Version
Final published version

Citation (APA)
Diwan, A. A., Urain, J., Kober, J., & Peters, J. (2025). *Noise-conditioned Energy-based Annealed Rewards (NEAR): A generative framework for imitation learning from observation*. Poster session presented at 13th International Conference on Learning Representations, ICLR 2025, Singapore, Singapore.

Important note
To cite this publication, please use the final published version (if applicable).
Please check the document version above.

Copyright
In case the licence states “Dutch Copyright Act (Article 25fa)”, this publication was made available Green Open Access via the TU Delft Institutional Repository pursuant to Dutch Copyright Act (Article 25fa, the Taverne amendment). This provision does not affect copyright ownership.
Unless copyright is transferred by contract or statute, it remains with the copyright holder.

Sharing and reuse
Other than for strictly personal use, it is not permitted to download, forward or distribute the text or part of it, without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license such as Creative Commons.

Takedown policy
Please contact us and provide details if you believe this document breaches copyrights.
We will remove access to the work immediately and investigate your claim.

This work is downloaded from Delft University of Technology.

Noise-conditioned Energy-based Annealed Rewards (NEAR): A Generative Framework for Imitation Learning from Observation

TU Delft | TECHNISCHE UNIVERSITÄT DARMSTADT | IAS

* anish.diwan@tu-darmstadt.de

Anish Diwan * , Julen Urain , Jens Kober † , Jan Peters †

Cognitive Robotics, TU Delft | Intelligent Autonomous Systems, TU Darmstadt

ICLR

TL;DR

Non-adversarial IRL using **state-only expert data**!

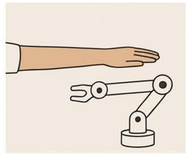
NEAR uses learnt **energy functions (score-matching) as reward functions.**

Problem Setting

Difficult Action Data Collection



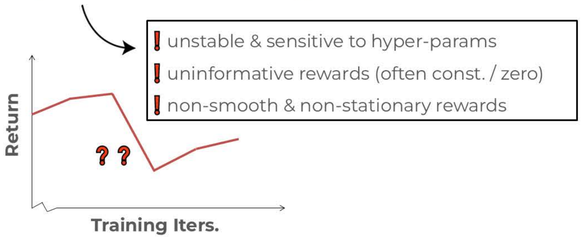
Actuator Embodiment Mismatch



Our paper focuses on **inverse RL** only using trajectories of the expert's states.

Challenges!

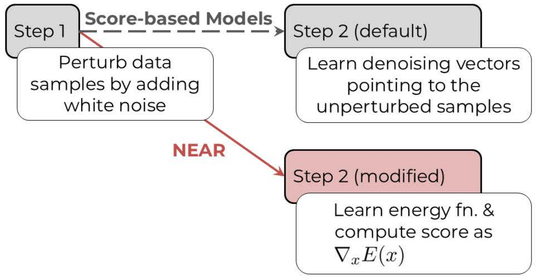
Existing methods (Adversarial IL) are prone to **optimization & RL challenges**!



Contributions

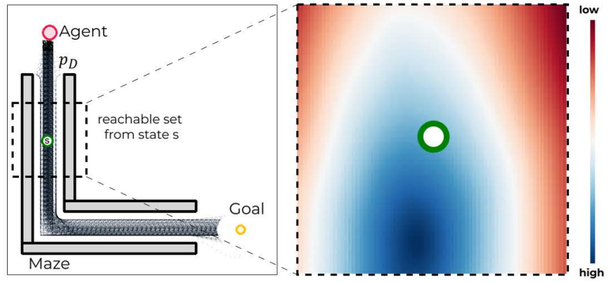
1 Energy Fn. As Reward Fn.

Can we use score-based generative models in inverse RL?



Perturbed distribution $\mathcal{N}(x, \sigma) = \exp E(x \mid \sigma)/Z$

$E(x \mid \sigma) \Rightarrow$ **scalar indicating closeness to expert** $\Rightarrow$ **reward function!**



- ✔ Smooth and easy to optimize w. RL (stationary during RL)
- ✔ Easily combines with other objectives
- ✔ Not prone to min-max optimization issues of adversarial IL

Contributions

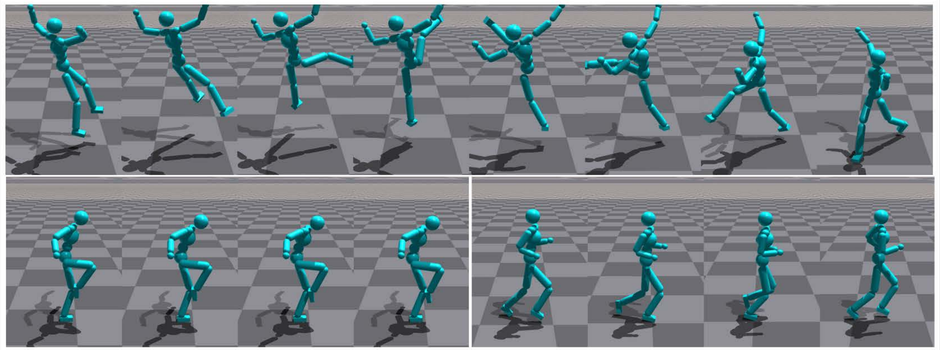
2 Reward Fn. Annealing: How Much Noise To Add ?

Low $\sigma \Rightarrow E(\cdot \mid \sigma)$ **is closer to** p_D **but less noisy** (and vice versa)

Noisiness ensures that the reward function is well-defined and informative

Solution: learn multiple $E(\cdot \mid \sigma_i)$ and **change** σ_i **during RL**

Results



Algorithm	Walking (74 clips)		Running (26 clips)		Left Punch (19 clips)	
NEAR	**0.51 ± 0.15**	**-7.52 ± 1.32**	**0.62 ± 0.17**	**-7.24 ± 1.59**	0.37 ± 0.05	**-6.87 ± 1.47**
AMP	**0.51 ± 0.07**	-8.78 ± 1.04	0.65 ± 0.01	-9.71 ± 1.54	**0.32 ± 0.01**	-9.93 ± 3.28
Expert	-	-5.4	-	-3.79	-	-1.73

Algorithm	Crane Pose (3 clips)		Mummy Walk (1 clip)		Spin Kick (1 clip)	
NEAR	0.94 ± 0.15	-6.6 ± 1.97	0.66 ± 0.39	**-4.72 ± 1.2**	0.78 ± 0.05	-5.59 ± 2.26
AMP	**0.82 ± 0.09**	**-8.1 ± 1.18**	**0.41 ± 0.01**	-13.84 ± 1.12	**0.58 ± 0.1**	**-3.16 ± 0.73**
Expert	-	-12.28	-	-4.71	-	-3.39

Avg. pose error (lower is better) | Spectral arc length (closer to expert is better)