Neurosymbolic Imitation Learning with Human Guidance: A Privileged Information Approach

Nikhilesh Prabhakar[1⋆], Varun Balaji[1], Athresh Karanam[1], Kristian Kersting[2], and Sriraam Natarajan[1]

[1] Department of Computer Science, The University of Texas at Dallas, Richardson TX 75080, USA
{nikhilesh.prabhakar,varun.balaji2,athresh.karanam,sriraam.natarajan}@utdallas.edu
[2] Computer Science Department and Centre for Cognitive Science, TU Darmstadt, Germany kristian.kersting@tu-darmstadt.de

Abstract. Imitation learning is widely used for learning to act in complex environments. While pure neural-based methods handle high dimensional data effectively, they suffer from the requirement of large number of samples and are prone to overfitting. Pure symbolic approaches, while generalize well, do not handle high-dimensional data effectively. We propose a neurosymbolic approach that achieves the best of both worlds, i.e, handling high-dimensional data while achieving generalization. The key advantage of our approach is that it can effectively exploit additional privileged information that is available only during training (in our case, gaze data). Our empirical evaluations demonstrate the effectiveness, efficiency and the generalization capability of our proposed approach.

Keywords: Neurosymbolic Learning · Imitation Learning · Privileged Information ·

1 Introduction

Skill acquisition by observation has long been addressed in the AI community as Imitation Learning [29]. The key idea is to learn a policy (i.e., a mapping from states to actions) from observations of states and corresponding policy from an expert. This problem has been extensively studied under different conditions – learning by observation [37], learning from demonstrations [2], programming by demonstrations [6], programming by example [22], apprenticeship learning [1], behavioral cloning [34], learning to act [16], and some others.

Recent advances in deep learning has accelerated imitation learning from games by learning from raw image data such as Atari games [4, 26]. While effective, these methods learn complex policies that are opaque, complex and uninterpretable. While certain local explanation methods exist, the explanations are typically post-hoc. Symbolic methods [27] on the other hand, learn interpretable, explainable, most importantly generalizable and potentially interactive models.

⋆ Preprint under review

However, these models assume that the input representation is symbolic, an assumption that restrict the adaptation of these methods to large domains.

Neurosymbolic approaches [11] bridge this gap between deep learning and symbolic methods and have recently re-attracted significant attention due to the emergence of the test-time scaling of autoregressive models. Inspired by the success of these methods, we develop a neurosymbolic method for imitation learning. The key idea in our approach is to use neural methods to create symbolic inputs from raw image data, and learn interpretable and generalizable probabilistic logic rules to model the policy.

A key aspect of our method is its ability to use additional information. Advice-taking methods that use additional information from the domain expert have a long and cherished history in AI [43]. Building on the successes of these methods, we consider additional information during training as privileged information (one that is not available during testing/deployment) to learn probabilistic logic rules. We specifically focus on Atari games, a traditional setting hard for symbolic models, and use gaze data as the privileged information. Our proposed framework, Gaze Guided Relational Approach to Imitation Learning (GRAIL) demonstrates efficacy, effectiveness, and generalization on empirical evaluations against purely neural and purely symbolic methods. Our key observation in this work is that combining ideas from classical AI literature such as neural, symbolic, advice-based and imitation learning can result in a robust, generalizable imitation learning model. We make the following key contributions: (1) We develop a novel neurosymbolic framework that learns interpretable probabilistic logic rules for imitation learning. (2)We consider additional information in the form of human gaze as privileged information and develop a robust learning procedure. (3) Our paper combines ideas from multiple areas of AI – neural learning, symbolic learning, imitation learning and advice-based learning. (4) We empirically validate that the proposed GRAIL framework is both sample efficient, and learns robust policies while generalizing to an unknown number of objects at test-time.

The rest of the paper is organized as follows: we next provide the necessary background and related work on the key topics of the paper – imitation learning, neurosymbolic learning and advice-based learning. We next present our key contribution, the GRAIL framework with necessary examples and notations before presenting our empirical evaluations on two ATARI domains, namely, *Asterix* and *Seaquest*. Finally, we conclude the paper by outlining areas of future research.

2 Background and Related Work

Notation: Let $o_t \in \mathcal{O} \subseteq \mathbb{R}^{C \times H \times W}$ denote the observation at time step t, and $a_t \in \mathcal{A}$ represent the corresponding action. We define a trajectory of length T as the sequence $\tau = (o_1, a_1, \ldots, o_T, a_T) \in (\mathcal{O} \times \mathcal{A})^T$. We define the policy $\pi_W : \mathcal{O} \times \mathcal{A} \rightarrow [0, 1]$, parameterized by a vector of weights $W \in \mathbb{R}^M$ and defined

over a set of M first-order definite clauses $\{c_i\}_{i=1}^{M}$, as a function mapping the joint observation space and action space to a probability distribution.

We introduce the rest of the notation as necessary.

2.1 Markov Decision Processes (MDPs)

A *Markov Decision Process* (MDP) provides the formal framework for modeling sequential decision-making problems. Formally, an MDP is defined as a tuple $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{T}, \mathcal{R}, \gamma \rangle$, where $\mathcal{S}$ is a set of states, $\mathcal{A}$ is a set of actions, $\mathcal{T}(s, a, s') = P(s_{t+1} = s' \mid s_t = s, a_t = a)$ is the transition function, $\mathcal{R}(s, a, s')$ is the reward function, and $\gamma \in [0, 1)$ is a discount factor [40].

In the reinforcement learning setting, the transition dynamics and reward function are typically unknown, and the agent must learn an optimal policy (and in model-based methods, the transition and reward dynamics) through interaction with the environment.

2.2 Imitation Learning (IL)

Imitation Learning (IL) is a learning paradigm in which an agent acquires a policy by observing expert demonstrations. Given a set of expert trajectories, the goal of IL is to learn a policy π that replicates the expert's behavior. This is particularly useful in settings where the reward function is difficult to specify manually, as the expert demonstrations implicitly encode the desired behavior. The two main approaches to IL are *Behavioral Cloning* (BC) [34], which treats policy learning as a (sequential) supervised learning problem, and *Inverse Reinforcement Learning* (IRL) [28], which infers the latent reward function that best explains the observed demonstrations [29]. Imitation learning has seen BC-based approaches applied to increasingly high-dimensional tasks. However, such approaches often suffer from covariate-shift. This occurs when small errors lead the agent to states not covered by the training distribution. Interactive frameworks such as DAgger [33], SEARN [9] and SMILe [32] address this by querying experts during training, they require "human-in-the-loop" simulators and incur high operational costs. Beyond covariate shift, purely neural BC architectures also suffer from two additional structural limitations: (1) they do not generalize to scenes with a varying number of objects, and (2) the learned policy remains opaque, offering no symbolic interpretation of the agent's decision-making. Our work addresses all three limitations by replacing the neural policy with a differentiable symbolic reasoner.

2.3 Learning using privileged information

The Learning Using Privileged Information (LUPI) paradigm [43, 42] aims to utilize auxiliary features, $\mathbf{Z}$, - in addition to regular features, $\mathbf{X}$ - that are available during training and not during testing with the goal of improving modeling accuracy, enabling faster model convergence, and improving robustness to noise,

among others. LUPI methods can be broadly categorized into two complementary categories: distillation-based transfer [14] and PI-inference-based.

Distillation-based LUPI methods use teacher-student pairs to train a teacher model with access to PI and transfer its information to a student that does not have access to PI. This line of work hinges upon the idea that soft targets from such a teacher can mimic the benefits of PI at test time, as shown analytically by [24]. In effect, this provides a model-agnostic framework to incorporate PI, where the teacher captures the PI which then implicitly guides the student during training through the distillation process [15]. This generic formulation is adapted by several LUPI algorithms [25, 12, 46, 45].

PI-inference-based LUPI methods, on the other hand, treat $\mathbf{Z}$ as latent variables to be inferred at test-time. These methods are particularly effective when the PI exerts complex and instance-specific influence on the target variable, allowing fine-grained calibrated predictions and explicit probabilistic semantics. In particular, they are effective at dealing with aleatoric uncertainty, especially in the presence of heteroscedastic noise [8, 20, 13], making them ideal candidates for dealing with heteroscedasticity inherent to human demonstrations [5].

Our work belongs to the latter category. However, instead of training a predictive model on the joint input space of $(\mathbf{X}, \mathbf{Z})$ and marginalizing $\mathbf{Z}$ out at test-time, we propose using the PI to augment the input space directly.

2.4 Treating gaze as privileged information

Treating gaze as an auxiliary signal can allow the agent to learn a model that focuses on causally relevant features without requiring additional expert interaction. Although not explicitly categorized under the LUPI paradigm, much prior work has used gaze as a privileged attention signal available at train time. Gaze-based regularization methods penalize the policy for attending to regions inconsistent with human gaze, either by modifying the loss [35, 3] or by modulating dropout within convolutional layers [7]. A second family of methods directly filters or crops the input observations using a predicted gaze saliency map, instantiated variously as masking [47, 21], observation cropping [17, 18], and gaze concatenation in drone racing and autonomous driving contexts [30, 44]. A third family adopts a multi-objective architecture, jointly predicting gaze and control commands so that gaze acts as an inductive bias during training [41]. Critically, all of these methods apply gaze guidance to a neural policy, preserving the opacity inherent to such architectures. In contrast, we apply gaze as an attention prior over the symbolic grounding stage, guiding which relational atoms the perception module attends to, rather than regularizing pixel-level feature maps.

2.5 Relational and Neurosymbolic Imitation Learning

Relational representations offer a natural solution to the generalization problem, as policies defined over object relations rather than fixed-size feature vectors can operate on a varying number of entities [10]. Inductive Logic Programming

(ILP) provides a framework for learning relational policies by inducing first-order rules from demonstrations. Formally, given background knowledge B and a language bias $\mathcal{L}$, ILP seeks a hypothesis $H \subseteq \mathcal{L}$ of definite clauses of the form $\alpha \leftarrow \alpha_1, \ldots, \alpha_m$, where each atom is an n-ary predicate over variables or constants [23]. While expressive and interpretable, classical ILP requires clean, pre-specified symbolic inputs and cannot operate directly on high-dimensional visual observations. differentiable ILP (∂ILP) addresses this by replacing hard symbolic inference with gradient-based optimization, enabling end-to-end rule learning via backpropagation [11]. More recent neurosymbolic approaches couple a neural perception front-end with a differentiable symbolic reasoner, enabling end-to-end training from raw observations The Neural Symbolic Forward Reasoner (NSFR) [39] performs differentiable forward-chaining inference over a library of first-order logic (FOL) clauses, using soft t-norm and t-conorm operators to approximate logical conjunction and disjunction in a fully differentiable manner. While NSFR has been applied in supervised classification settings, its application to imitation learning from visual demonstrations with gaze-guided symbolic grounding has not been explored. Our work bridges this gap by training an NSFR-based policy end-to-end on the Atari-HEAD dataset, using human gaze to supervise the neural grounding module that maps raw frames to probabilistic relational atoms.

3 Neurosymbolic Imitation Learning using Privileged Information

We formally define the learning problem as follows

Given: A dataset $\mathcal{D} = \{\tau_i\}_{i=1}^{N}$ of expert trajectories, its corresponding privileged gaze heatmap $G_t \in \Delta^{H \times W}$ [43], and a rule base $\mathcal{B} = \{c_j\}_{j=1}^{M}$ of first-order definite clauses encoding expert domain knowledge over grounded atoms $\mathcal{P}$.

To Do: Learn a policy π_W such that the rollout regret $J(\pi^*) - J(\pi_W)$ to the expert policy π^* is minimal.

To this end, we propose Gaze-guided Relational Imitation Learning (GRAIL), which integrates three tightly coupled modules: a neural perception module that grounds high-dimensional visual observations into (probabilistic) atoms; a gaze-conditioned attention mechanism that modulates atom valuations using privileged human eye-tracking data; and a differentiable forward-chaining reasoner that learns an interpretable, rule-weighted policy via behavior cloning. Figure 1 illustrates the full pipeline. As far as we are aware, this is the first neurosymbolic learning framework that can exploit (propositional) additional knowledge, in this case gaze information, to learn symbolic policies.

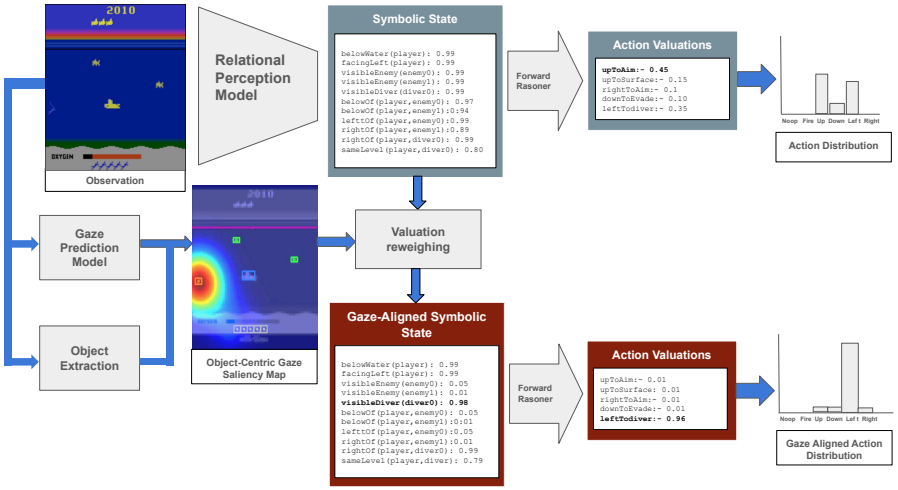

Fig. 1: **Neurosymbolic Imitation Learning (NESY-IL) Architecture.** The framework integrates high-dimensional visual perception with structured symbolic reasoning. A perception model extracts a set of grounded atoms from raw observations. Concurrently, a gaze prediction model and an object extractor provide object-level saliency, which is used to reweigh the states and align them with human attention. This gaze-aligned symbolic representation is then processed by a differentiable forward reasoner to generate action valuations and a final action distribution that reflects expert attentional biases.

3.1 Neurosymbolic Grounding via Probabilistic Atom Valuation

Note that unlike other typical symbolic learning methods, we do not assume an input representation that is also symbolic. Instead, we assume that the input is a set of raw images. Consequently, the first key challenge in our imitation learning is *grounding* of symbols: mapping raw pixel observations to a structured representation suitable for symbolic inference. The grounded state of a domain is defined by a finite set of grounded atoms $\mathcal{P} = \{p_1, \ldots, p_{|\mathcal{P}|}\}$, where each atom p_i is an instantiated predicate over domain entities (e.g., `close(player, enemy)`).

We train a neural perception network $f_\theta : \mathbb{R}^{C\times H\times W} \to [0,1]^{|\mathcal{P}|}$ to produce a probabilistic valuation vector:

$$\mathbf{v}_t^{(0)} = f_\theta(o_t) \in [0,1]^{|\mathcal{P}|} \tag{1}$$

where $v_{t,i}^{(0)}$ denotes the inferred truth probability of atom p_i at timestep t. This is inspired from several probabilistic logic frameworks such as Probabilistic Horn Abduction [31], PRISM [36] and Problog [19] to name a few. The probabilities are assigned to ground atoms and first-order clauses are learned as unweighted clauses.

Algorithm 1 GRAIL: Gaze-guided Relational Imitation Learning

Require: Expert trajectories $\mathcal{D} = \{\tau_i\}_{i=1}^{N}$, gaze heatmaps $\{G_t\}$, atom set $\mathcal{P}$, rule base $\mathcal{B}$, max steps $T_{\max}$

function GRAIL($\mathcal{D}, \{G_t\}, \mathcal{P}, \mathcal{B}$)
  for each $(o_t, \mathbf{y}_t)$ from $\mathcal{D}$ **do** ▷ Train perception module with ground-truth atom labels
    $\mathcal{L}_{\text{NLL}} \leftarrow -\sum_i y_{t,i} \log v_{t,i}^{(0)}$
    Update θ via $\nabla_\theta \mathcal{L}_{\text{NLL}}$
  end for
  for each (o_t, G_t) from $\mathcal{D}$ **do** ▷ Train gaze prediction module
    $\mathcal{L}_{\text{gaze}} \leftarrow D_{\text{KL}}(G_t \parallel g_\phi(o_t))$
    Update ϕ via $\nabla_\phi \mathcal{L}_{\text{gaze}}$
  end for
  Freeze θ, ϕ; initialize W ▷ Begin policy learning
  for each (o_t, a_t) from $\mathcal{D}$ **do**
    $\mathbf{v}_t^{(g)} \leftarrow \textsc{GazeModulate}(o_t, \mathcal{P})$ ▷ Gaze-filtered grounding
    $\mathbf{v}^{(T_{\max})} \leftarrow \textsc{ForwardChain}(\mathbf{v}_t^{(g)}, \mathcal{B}, W)$ ▷ Differentiable inference
    $s_c \leftarrow \max_{r \in \mathcal{R}_c} v^{(T_{\max})}(r) \quad \forall c \in \mathcal{A}$
    $\mathcal{L}_{\text{policy}} \leftarrow -\log \pi_W(a_t \mid o_t)$
    Update W via $\nabla_W \mathcal{L}_{\text{policy}}$
  end for
  return W
end function

To supervise this module, we first apply a non-differentiable object extraction oracle (OCAtari) to parse expert trajectories into ground-truth binary valuations $\mathbf{y}_t \in \{0,1\}^{|\mathcal{P}|}$. The perception network is then trained independently by minimizing the negative log-likelihood (NLL) loss:

$$\mathcal{L}_{\text{NLL}}(\theta) = -\frac{1}{|\mathcal{P}|} \sum_{i=1}^{|\mathcal{P}|} y_{t,i} \log v_{t,i}^{(0)} \tag{2}$$

Once trained, f_θ serves as a fixed symbolic feature extractor, decoupling visual perception from downstream logical reasoning.

3.2 Privileged Gaze Modulation of Atom Valuations

Even after grounding, the valuation vector $\mathbf{v}_t^{(0)}$ may assign non-trivial truth probabilities to atoms involving task-irrelevant background entities, introducing spurious correlations into downstream rule learning. We resolve this credit assignment problem via *privileged information* in the form of human eye-tracking data, available only during training under the LUPI paradigm [43].

Gaze Predictor. Prior to policy training, we train a dedicated gaze prediction network $g_\phi : \mathbb{R}^{1\times H\times W} \to \mathbb{R}^{H\times W}$ that maps a single grayscale frame $o_t \in$

$\mathbb{R}^{1\times H\times W}$ to a spatial probability distribution over fixation locations. Ground-truth gaze coordinates are converted into 2D Gaussian heatmaps G_t, and g_ϕ is optimized by minimizing the Kullback-Leibler divergence:

$$\mathcal{L}_{\text{gaze}}(\phi) = D_{\text{KL}}(G_t \parallel g_\phi(o_t)) \tag{3}$$

Atom-Level Gaze Scoring. During policy training, g_ϕ is frozen and produces a normalized gaze heatmap $\hat{G}_t$, where $\sum_{x,y} \hat{G}_t(x,y) = 1$. An atom p_i in $\mathcal{P}$ may reference one or more entities $e_i^{(1)} \cdots e_i^{(k)}$, each with an associated bounding box β_i. We compute a per-entity saliency score by calculating a gaze mass as follows

$$s_i = \sum_{(x,y)\in\beta_i} \hat{G}_t(x,y) \in [0,1] \tag{4}$$

For atoms referencing a single entity, the scalar gaze score is $s_i = s_i^{(1)}$. For atoms referencing multiple entities, we aggregate using the product t-conorm

$$s_i = 1 - \prod_{j=1}^{n} (1 - s_i^{(j)}) \tag{5}$$

Gaze-Modulated Valuation. The gaze score is applied as a soft multiplicative scaling over the initial valuation vector, yielding the gaze-modulated symbolic state:

$$v_{t,i}^{(g)} = v_{t,i}^{(0)} \cdot s_i \tag{6}$$

The intuition is that we essentially reweigh the probabilities of the ground atoms based on the gaze information. If the gaze heatmap covers the grounding object, the corresponding probability is increased else it is decreased. This is akin to the idea of relevance information used in earlier ILP learning systems [38] where the importance of a grounded predicate can be decreased or increased using domain knowledge thus changing the score the clause induced by the ground atom. We instead change the probabilities of these atoms.

3.3 Differentiable Forward-Chaining Inference for Policy Learning

The gaze-filtered valuation vector $\mathbf{v}^{(g)_t}$ is passed to a differentiable logic reasoning engine, implemented via the Neurosymbolic Forward Reasoner (NSFR). NSFR performs forward-chaining inference over a predefined rule base $\mathcal{B}$, a set of first-order definite clauses of the form:

$$\alpha_0 \leftarrow \alpha_1, \alpha_2, \ldots, \alpha_m \tag{7}$$

where each α_j is a grounded FOL atom. Discrete Boolean inference is relaxed into continuous arithmetic via fuzzy logic semantics: logical conjunction ($\wedge$) is approximated by the product T-norm, and disjunction ($\vee$) by the corresponding T-conorm, enabling end-to-end differentiation through the inference graph.

Forward Chaining. Starting from the initial facts $\mathbf{v}_t^{(g)}$, the engine performs $T_{\max}$ deductive steps, iteratively updating atom valuations by applying the weighted rule base. This yields a final valuation tensor $\mathbf{v}_t^{(T_{\max})}$ encoding the inferred truth probabilities of all action predicates.

Behavior Cloning Objective. Let $\mathcal{R}_c \subseteq \mathcal{B}$ denote the subset of clauses with action c as their head. The aggregated action score is:

$$s_c = \max_{r \in \mathcal{R}_c} v_t^{(T_{\max})}(r) \tag{8}$$

producing an unnormalized score vector $\mathbf{s} \in [0,1]^{|\mathcal{A}|}$. The learnable rule weight matrix W is optimized end-to-end by minimizing the negative log-likelihood of the expert action under the induced policy $\pi_W(a \mid o_t)$:

$$\mathcal{L}_{\text{policy}}(W) = -\sum_{t=1}^{T} \log \pi_W(a_t \mid o_t) \tag{9}$$

Because $\mathbf{v}_t^{(g)}$ has been filtered by the privileged gaze mechanism, backpropagation updates W exclusively on task-critical relational features, preventing the formation of spurious correlations with background distractors and improving both sample efficiency and out-of-distribution generalization.

3.4 GRAIL Framework

Putting all of these together, we get the GRAIL framework that is outlined in Figure 1. To reiterate, the framework obtains raw images as inputs, converts them to probabilistic grounded atoms using the perception module. Then given the gaze information for the corresponding raw image, the probabilities of the ground atoms are reweighted accordingly. Next, these grounded atoms are used to learn relational policies using a forward reasoner that employs a behavioral cloning objective. The final policies obtained in this fashion are inrepretable, explainable and most importantly, as we demonstrate, generalizable. As far as we are aware, this is the first end-to-end neurosymbolic framework capable of seamlessly integrating gaze information as privileged information to learn effective, efficient and generalizable policies.

4 Experimental Evaluation

Our experiments explicitly aim at answering the following questions:

- **Q1 — Effectiveness:** Does GRAIL learn a competitive game-playing policy compared to the baseline methods while effectively leveraging gaze data?
- **Q2 — Sample Efficiency:** Does access to human gaze data during training improve performance in low-data regimes?

Table 1: Mean Game Score across Games

Game	BC	AGIL	BC+Mask	NSFR	GRAIL
Asterix	306.00 ± 223.75	228.00 ± 175.26	251.00 ± 153.46	2519 ± 1488.65	$\mathbf{6308.00 \pm 3765.63}$
Seaquest	139.20 ± 39.18	192.80 ± 60.89	182.00 ± 60.82	454.20 ± 272.71	$\mathbf{501.60 \pm 216.84}$

– **Q3 — Generalization:** Does the symbolic state representation generalize better than pixel-based approaches when the test environment contains unseen objects?

Our code is publicly available at `https://anonymous.4open.science/status/Gaze-Based-Neurosymbolic-Imitation-Learner-665E`.

4.1 Domain

We evaluate GRAIL on two Atari 2600 environments, Seaquest, and, Asterix, drawn from the Atari Human Eye-Tracking and Demonstration (Atari-HEAD) dataset [48]. Atari-HEAD creates gameplay recordings with high-precision eye-tracking data and human action demonstrations across 20 games and multiple human subjects, comprising over 100 hours of recorded play. To ensure consistency, we restrict our experiments to the trajectories of a single human player and use only primitive actions, excluding key combinations for a controlled demonstration dataset.

4.2 Baselines

We compare our algorithm, **GRAIL**, against four baselines — (i) a purely neural behavioral cloning baseline with no gaze information (**BC**); (ii) a gaze-augmented neural baseline that masks pixel observations with a predicted gaze saliency map, requiring the gaze predictor at test time (**AGIL**); (iii) a gaze-augmented neural baseline that multiplies input observations by a gaze-derived attention mask to suppress task-irrelevant pixel regions (**BC+Mask**); and (iv) an ablation of our own framework that removes the gaze modulation module entirely, substituting unmodulated atom valuations $\mathbf{v}_t^{(0)}$ for the gaze-filtered $\mathbf{v}_t^{(g)}$ (**NSFR-IL**), directly isolating the contribution of privileged gaze information within the neurosymbolic setting.

4.3 Experimental Results

Q1. Effectiveness To assess whether the neurosymbolic architecture learns a competitive game-playing policy, we compare mean game scores across all methods on Seaquest and Asterix, reported across 50 evaluation seeds in Table 1. On both domains, both symbolic methods considerably outperform the neural baselines, with GRAIL achieving the highest mean score overall. Comparing GRAIL

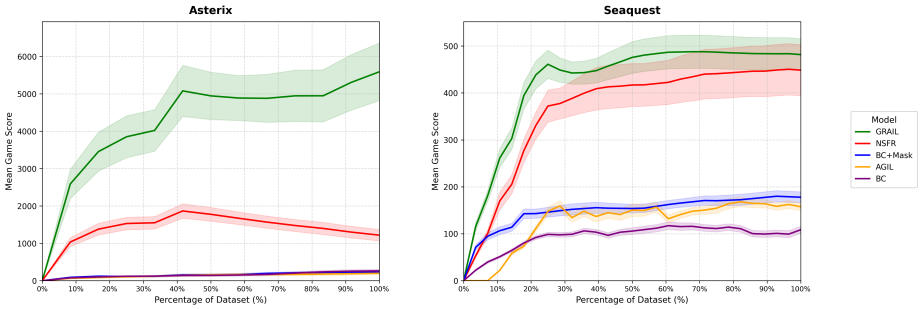

Fig. 2: **Comparison of sample efficiency in Atari games.** Mean Game Score is plotted against the percentage of training dataset used for two domains, Asterix (left) and Seaquest (right). In both the domains, GRAIL achieves higher performance earlier and converges to a higher asymptote.

directly to NSFR-IL demonstrates the significant marginal gain of gaze modulation in our framework, as the two methods share identical architectures and rule bases, differing only in whether gaze modulation is applied. This confirms the effectiveness of GRAIL's neurosymbolic architecture as well as its ability to effectively leverage gaze data.

Q2. Sample Efficiency To assess whether access to human gaze data improves performance in low-data regimes, we compare the learning curves of all methods as a function of the number of training demonstrations, shown in Figure 2. The learning curves show that GRAIL converges to a higher asymptote than all baselines in both domains and does so with substantially fewer samples Most strikingly, GRAIL trained on 10% of the available data already outperforms NSFR-IL trained on the full dataset. Neural baselines plateau at lower performance levels and do not recover with additional data, suggesting that the bottleneck is representational rather than statistical. A notable phenomenon is the performance degradation of NSFR-IL on Asterix as the number of samples grows. We believe that this reflects a gap between the expert rules crafted and the behavior of the human gameplay data. As more samples are observed, the model increasingly fits spurious correlations in the training data that do not generalize to evaluation. GRAIL does not exhibit this degradation, suggesting that gaze modulation acts as a regularizer. By down-weighting atom valuations for entities the human expert does not attend to, it prevents the rule weights from latching onto relational features that co-occur with correct actions in the training data but are not relevant to the policy. In effect, human attention provides a supervisory signal that steers the learner away from spurious correlations that additional data would otherwise support.

Q3. Generalization To assess whether the symbolic state representation generalizes better than pixel-based approaches when the test environment contains

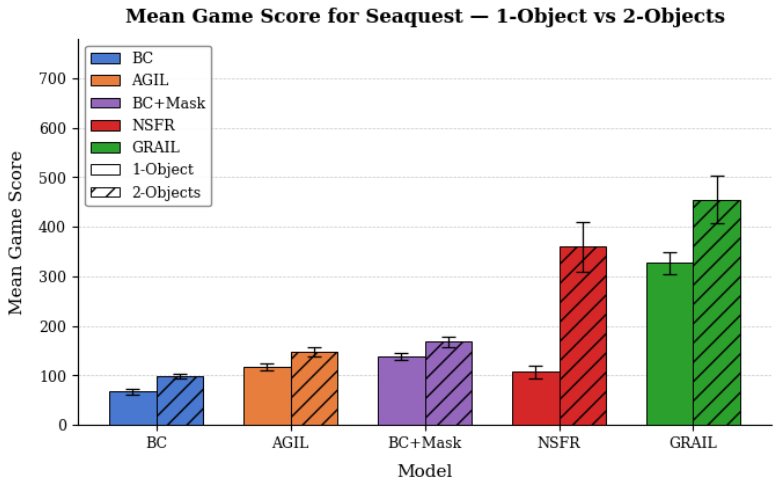

Fig. 3: **Generalization across variable object counts in Seaquest.** Mean game scores are reported for configurations trained on a dataset containing a maximum of one and two objects per object type.

unseen objects, we train all methods on datasets containing a maximum of one and two objects per object type in Seaquest. The results are presented in Figure 3. Neural baselines — BC, AGIL, and BC+Mask — show only marginal improvement when moving from the 1-object to the 2-objects training regime, with mean game scores remaining below 200 across both conditions. Both neurosymbolic methods, on the other hand, benefit considerably from the richer training distribution. NSFR-IL improves substantially in the 2-objects setting, and GRAIL improves further still, achieving the highest mean reward overall. This is consistent with the compositionality of first-order relational representations. The policy is defined over object-level predicates rather than fixed-size feature vectors, so it naturally extends to scenes containing a greater number of entities. The additional advantage of GRAIL over NSFR-IL in the 2-objects setting further suggests that gaze modulation helps the reasoner maintain focus on task-critical atoms as scene complexity increases.

In summary, firstly, our experimental results across the two domains show that our framework, GRAIL, outperforms both the neurosymbolic as well as the purely neural baseline methods. Secondly, as evidenced by NSFR-IL - which GRAIL subsumes by removing the gaze modulation module - surpassing neural baselines in effectiveness (Q1), sample efficiency (Q2) and generalization (Q3), neurosymbolic methods for imitation learning outperform purely neural methods regardless of incorporation of PI. Finally, our results show that the proposed gaze modulation effectively and consistently leverages PI to achieve higher overall performance (Q1), faster convergence in low-data regimes (Q2), and generalizes

to unknown number of objects (Q3) at test-time through a combination of the neurosymbolic architecture and PI.

5 Conclusion

We presented GRAIL, a gaze-guided neurosymbolic imitation learning framework that is capable of learning from raw image data and additional privileged information in the form of gaze maps. This is the first-of-its-kind framework that effectively employs privileged information that is available during training but not during deployment. Note that because we considered raw game playing data, gaze information served as a natural privileged information. Our framework is certainly not limited to the use of only gaze information. Additional privileged information, if available can be easily integrated to the framework as they are mainly used to reweigh the probabilities. Another key aspect of the system is that it is a plug and play framework and one can employ any differentiable learner for learning the policies, any perception module capable of extracting probabilistic facts and as mentioned earlier, any form of privileged information for reweighing the probabilities. Our end-to-end learning system is capable of learning effective, efficient, explainable and generalizable policies as we demonstrated empirically in multiple image-based domains, traditionally hard ones for symbolic learning.

Scaling the algorithms to larger problems including clinical decision-support is an interesting future directions. Integrating other forms of domain knowledge in the form of qualitative constraints, preferences and shaping functions is another potential direction. Similarly, incorporating gaze information as other forms including regularization as well as implementing rule learners are immediate next steps. Allowing for richer human interaction requires integration with a LLM-based interaction system, another interesting direction. Finally, developing imitation learning systems that knows what it knows and solicits information (whether it is gaze or preferences or shaping functions etc) when it does not know, is one of the most exciting questions in building true human-allied learning systems that learn to act from observations.

6 Generative AI Use

We declare that Generative AI was not used to write or edit the paper. All the content of the paper are original and the authors are solely responsible for the content of the paper.

References

1. Abbeel, P., Ng, A.Y.: Apprenticeship learning via inverse reinforcement learning. In: Proceedings of the Twenty-First International Conference on Machine Learning (ICML). p. 1. ACM (2004). https://doi.org/10.1145/1015330.1015430

2. Argall, B.D., Chernova, S., Veloso, M., Browning, B.: A survey of robot learning from demonstration. Robotics and Autonomous Systems **57**(5), 469–483 (2009). https://doi.org/10.1016/j.robot.2008.10.024
3. Banayeeanzade, A., Bahrani, F., Zhou, Y., Bıyık, E.: Gabril: Gaze-based regularization for mitigating causal confusion in imitation learning. arXiv preprint arXiv:2507.19647 (2025)
4. Bellemare, M.G., Naddaf, Y., Veness, J., Bowling, M.: The arcade learning environment: An evaluation platform for general agents. Journal of Artificial Intelligence Research **47**, 253–279 (2013)
5. Caldarelli, E., Chatalic, A., Colomé, A., Rosasco, L., Torras, C.: Heteroscedastic gaussian processes and random features: Scalable motion primitives with guarantees. In: CoRL. Proceedings of Machine Learning Research, vol. 229, pp. 3010–3029. PMLR (2023)
6. Calinon, S.: Robot Programming by Demonstration: A Probabilistic Approach. EPFL/CRC Press (2009)
7. Chen, Y., Liu, C., Tai, L., Liu, M., Shi, B.E.: Gaze training by modulated dropout improves imitation learning. In: 2019 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 7756–7761. IEEE (2019)
8. Collier, M., Jenatton, R., Kokiopoulou, E., Berent, J.: Transfer and marginalize: Explaining away label noise with privileged information. In: Chaudhuri, K., Jegelka, S., Song, L., Szepesvari, C., Niu, G., Sabato, S. (eds.) Proceedings of the 39th International Conference on Machine Learning. Proceedings of Machine Learning Research, vol. 162, pp. 4219–4237. PMLR (17–23 Jul 2022)
9. Daumé, H., Langford, J., Marcu, D.: Search-based structured prediction. Machine learning **75**(3), 297–325 (2009)
10. Džeroski, S., De Raedt, L., Driessens, K.: Relational reinforcement learning. Machine learning **43**(1), 7–52 (2001)
11. Evans, R., Grefenstette, E.: Learning explanatory rules from noisy data. Journal of Artificial Intelligence Research **61**, 1–64 (2018)
12. Garcia, N.C., Morerio, P., Murino, V.: Learning with privileged information via adversarial discriminative modality distillation. IEEE Trans. Pattern Anal. Mach. Intell. **42**(10), 2581–2593 (2020)
13. Hernández-Lobato, D., Sharmanska, V., Kersting, K., Lampert, C.H., Quadrianto, N.: Mind the nuisance: Gaussian process classification using privileged noise. In: Advances in Neural Information Processing Systems. pp. 55–63 (2014)
14. Hinton, G., Vinyals, O., Dean, J.: Distilling the knowledge in a neural network. arXiv preprint arXiv:1503.02531 (2015)
15. Hoffman, J., Gupta, S., Darrell, T.: Learning with side information through modality hallucination. In: 2016 IEEE Conference on Computer Vision and Pattern Recognition (CVPR). pp. 826–834 (2016). https://doi.org/10.1109/CVPR.2016.96
16. Khardon, R.: Learning action strategies for planning domains. Artificial Intelligence **113**(1–2), 125–148 (1999). https://doi.org/10.1016/S0004-3702(99)00060-0
17. Kim, H., Ohmura, Y., Kuniyoshi, Y.: Using human gaze to improve robustness against irrelevant objects in robot manipulation tasks. IEEE Robotics and Automation Letters **5**(3), 4415–4422 (2020)
18. Kim, H., Ohmura, Y., Kuniyoshi, Y.: Gaze-based dual resolution deep imitation learning for high-precision dexterous robot manipulation. IEEE Robotics and Automation Letters **6**(2), 1630–1637 (2021)
19. Kimmig, A., Demoen, B., Raedt, L.D., Costa, V.S., Rocha, R.: On the implementation of the probabilistic logic programming language ProbLog. In: Theory and

Practice of Logic Programming. vol. 11, pp. 235–262. Cambridge University Press (2011)
20. Lambert, J., Sener, O., Savarese, S.: Deep learning under privileged information using heteroscedastic dropout. In: The IEEE Conference on Computer Vision and Pattern Recognition (CVPR) (June 2018)
21. Liang, A., Thomason, J., Bıyık, E.: Visarl: Visual reinforcement learning guided by human saliency. In: 2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS). pp. 2907–2912. IEEE (2024)
22. Lieberman, H.: Programming by example (introduction). Communications of the ACM **43**(3), 72–74 (2000). https://doi.org/10.1145/330534.330543
23. Lloyd, J.W.: Foundations of Logic Programming. Springer-Verlag, Berlin, 2nd edn. (1987)
24. Lopez-Paz, D., Bottou, L., Schölkopf, B., Vapnik, V.: Unifying distillation and privileged information. arXiv preprint arXiv:1511.03643 (2015)
25. Markov, K., Matsui, T.: Robust speech recognition using generalized distillation framework. In: Interspeech. pp. 2364–2368 (2016)
26. Mnih, V., Kavukcuoglu, K., Silver, D., Rusu, A.A., Veness, J., Bellemare, M.G., Graves, A., Riedmiller, M., Fidjeland, A.K., Ostrovski, G., et al.: Human-level control through deep reinforcement learning. Nature **518**(7540), 529–533 (2015)
27. Muggleton, S.: Inductive logic programming. New generation computing **8**(4), 295–318 (1991)
28. Ng, A.Y., Russell, S.: Algorithms for inverse reinforcement learning. In: Proceedings of the Seventeenth International Conference on Machine Learning (ICML). pp. 663–670 (2000)
29. Osa, T., Pajarinen, J., Neumann, G., Bagnell, J.A., Abbeel, P., Peters, J.: An algorithmic perspective on imitation learning. Foundations and Trends® in Robotics **7**(1-2), 1–179 (2018)
30. Pfeiffer, C., Wengeler, S., Loquercio, A., Scaramuzza, D.: Visual attention prediction improves performance of autonomous drone racing agents. PLoS one **17**(3), e0264471 (2022)
31. Poole, D.: Probabilistic horn abduction and bayesian networks. Artificial Intelligence **64**(1), 81–129 (1993)
32. Ross, S., Bagnell, D.: Efficient reductions for imitation learning. In: Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics. pp. 661–668. JMLR Workshop and Conference Proceedings (2010)
33. Ross, S., Gordon, G., Bagnell, D.: 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 (AISTATS). pp. 627–635. JMLR Workshop and Conference Proceedings (2011)
34. Sammut, C., et al.: Building symbolic representations of intuitive real-time skills from performance data (1992)
35. Saran, A., Zhang, R., Short, E.S., Niekum, S.: Efficiently guiding imitation learning agents with human gaze. arXiv preprint arXiv:2002.12500 (2020)
36. Sato, T.: A statistical learning method for logic programs with distribution semantics. In: Proceedings of the 12th International Conference on Logic Programming (ICLP). pp. 715–729. MIT Press (1995)
37. Segre, A.M., DeJong, G.: Explanation-based manipulator learning: Acquisition of planning ability through observation. In: Proceedings of the 1985 IEEE International Conference on Robotics and Automation. pp. 555–560 (1985)

38. Shavlik, J., Natarajan, S.: Speeding up inference in Markov Logic Networks by preprocessing to reduce the size of the resulting grounded network. In: Proceedings of the 21st International Joint Conference on Artificial Intelligence (IJCAI). pp. 1951–1956 (2009)
39. Shindo, H., Dhami, D.S., Kersting, K.: Neuro-symbolic forward reasoning. arXiv preprint arXiv:2110.09383 (2021)
40. Sutton, R.S., Barto, A.G.: Reinforcement Learning: An Introduction. The MIT Press, 2nd edn. (2018)
41. Thakur, R.K., Sunbeam, M.N.S., Goecks, V.G., Novoseller, E., Bera, R., Lawhern, V.J., Gremillion, G.M., Valasek, J., Waytowich, N.R.: Imitation learning with human eye gaze via multi-objective prediction. In: Interactive Learning with Implicit Human Feedback Workshop at ICML (2023)
42. Vapnik, V., Izmailov, R.: Learning using privileged information: Similarity control and knowledge transfer. Journal of Machine Learning Research **16**(61), 2023–2049 (2015), `http://jmlr.org/papers/v16/vapnik15b.html`
43. Vapnik, V., Vashist, A.: A new learning paradigm: Learning using privileged information. Neural Networks **22**(5-6), 544–557 (2009)
44. Xia, Y., Kim, J., Canny, J., Zipser, K., Canas-Bajo, T., Whitney, D.: Periphery-fovea multi-resolution driving model guided by human attention. In: Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision. pp. 1767–1775 (2020)
45. Yan, S., Odom, P., Pasunuri, R., Kersting, K., Natarajan, S.: Learning with privileged and sensitive information: a gradient-boosting approach. Frontiers Artif. Intell. **6** (2023)
46. Yang, S., Sanghavi, S., Rahmanian, H., Bakus, J., Vishwanathan, S.V.N.: Toward understanding privileged features distillation in learning-to-rank. In: NeurIPS (2022)
47. Zhang, R., Liu, Z., Zhang, L., Whritner, J.A., Muller, K.S., Hayhoe, M.M., Ballard, D.H.: Agil: Learning attention from human for visuomotor tasks. In: Proceedings of the European Conference on Computer Vision (ECCV). pp. 663–679 (2018)
48. Zhang, R., Walshe, C., Liu, Z., Guan, L., Muller, K.S., Whritner, J.A., Zhang, L., Hayhoe, M.M., Ballard, D.H.: Atari-HEAD: Atari human eye-tracking and demonstration dataset. In: Proceedings of the AAAI Conference on Artificial Intelligence. vol. 34, pp. 6811–6820 (2020)