THE COLOR OF CONSTRAINTS: A GEOMETRIC-SEMANTIC INTERFACE FOR NEUROSYMBOLIC INVERSE CONSTRAINED REINFORCEMENT LEARNING

Jingxuan Hou *
University of Pennsylvania
jingxhou@seas.upenn.edu

ABSTRACT

We address the fundamental challenge of learning interpretable constraints from expert demonstrations in inverse constrained reinforcement learning (ICRL). Current neurosymbolic approaches suffer from semantic entanglement in their predicate representations and lack temporal coherence. We propose a novel geometric-semantic interface that projects high-dimensional states into a structured HSV color space, where the cylindrical topology naturally encodes semantic hierarchies and temporal evolution of constraints. Our three-stage learning algorithm progressively builds constraint understanding from basic concept separation to geometric manifold learning and complex temporal pattern discovery. Preliminary results on cart-pole dynamics demonstrate the framework's ability to learn visually interpretable constraint representations that separate expert and violator behaviors. We outline a comprehensive research roadmap for theoretical analysis and empirical validation on complex domains.

1 Introduction

Inverse Constrained Reinforcement Learning (ICRL) aims to infer the unknown constraints that guide expert behavior, a crucial capability for building safe and aligned AI systems. While neurosymbolic methods have emerged as the state-of-the-art approach, they face fundamental limitations in their interface design – the critical bridge between continuous, high-dimensional perception and discrete, symbolic reasoning. This broken interface manifests as a **semantic gap**, where neurally-learned predicate representations often fail to capture the coherent, hierarchical, and temporally smooth nature of real-world constraints.

The standard neurosymbolic ICRL pipeline maps raw states to a flat vector of predicate probabilities, which are then composed via differentiable temporal logic [Deane and Ray, 2025, Chou et al., 2019, Li et al., 2021]. However, this approach suffers from three interconnected pathologies: (1) **Semantic Entanglement**, where neural networks encode multiple unrelated concepts within single predicate dimensions without guarantees of separability [Szegedy et al., 2014, Locatello et al., 2019]; (2) **Temporal Fragility**, where predicate activations at each timestep depend only on the instantaneous state, ignoring the temporal context essential for constraint reasoning [Gilpin et al., 2020, Sutton et al., 1999]; and (3) **Compositional Rigidity**, where the expressive power is limited to predefined temporal logic templates, requiring expensive combinatorial search over formula structures [Mascle et al., 2023].

We posit that the core issue lies not in the individual components – the neural perception or the symbolic reasoning – but in the impoverished interface connecting them. An effective interface must provide a structured substrate that respects the geometry of semantic concepts while enabling robust temporal composition. Drawing inspiration from cognitive science, where conceptual spaces often exhibit natural geometric and topological properties [Gärdenfors, 2000, 2004,

*This research preview presents work conducted in preparation for doctoral studies. The framework introduces geometric-semantic interfaces for neurosymbolic ICRL, with preliminary validation on constrained cart-pole dynamics. This work is not currently submitted for publication and represents independent research vision and execution. Code and visualizations available at: https://github.com/Paisley77/colorful-constraint-learning.

2014, Lieto et al., 2017], we propose moving beyond flat predicate vectors toward **geometrically structured semantic interfaces**.

In this work, we introduce the HSV color space as a particularly elegant instantiation of such an interface. The cylindrical topology of HSV – with its circular hue, radial saturation, and axial value dimensions – naturally embodies the properties we seek: hue provides a continuous, circular coordinate system for semantic categories; saturation represents confidence in semantic assignment; and value captures the prominence or intensity of concepts. This is not merely an analogy; we provide a principled mathematical mapping from neural concept activations to this structured color space using Principal Component Analysis (PCA), transforming the abstract problem of "learning constraints" into the concrete geometric problem of "separating colored trajectories."

Our **Colorful Constraint Learning** framework implements this vision through a three-stage algorithm that progressively builds constraint understanding: First, a contrastive learning stage separates expert and violator trajectories in the concept space. Second, an alternating optimization learns a cylindrical manifold in the HSV space that captures spatial constraints. Third, Temporal Convolutional Networks (TCNs) discover complex temporal patterns that complement the geometric manifold. This multi-stage approach provides a computationally tractable path from raw demonstrations to interpretable constraint representations.

The contributions of this work are fourfold:

1. **A Formal Framework**: We introduce the HSV color space as a geometrically structured semantic interface for neurosymbolic ICRL, with a principled mapping from neural concepts to this interpretable latent space.

2. **A Multi-Stage Algorithm**: We develop a three-stage learning procedure that progressively builds constraint understanding from basic separation to geometric manifolds and complex temporal patterns.

3. **Preliminary Validation**: We provide proof-of-concept results on a cart-pole system, demonstrating the framework's ability to learn visually interpretable constraint representations that separate expert and violator behaviors.

4. **Research Roadmap:** We outline a comprehensive path for theoretical analysis and empirical extension to complex domains, positioning this work as a foundation for future research.

2 Related Works

2.1 Inverse Constrained Reinforcement Learning

Inverse Constrained Reinforcement Learning (ICRL) extends the classical Inverse Reinforcement Learning (IRL) problem by focusing on inferring constraints, or "avoidance behaviors," rather than reward functions. Given expert demonstrations that implicitly respect unknown safety constraints, the goal is to learn a constraint function that explains the expert's behavior, typically within a Constrained Markov Decision Process (CMDP) framework. Existing ICRL methods can be broadly categorized into three paradigms: maximum margin, probabilistic, and game-theoretic approaches.

Maximum Margin Approaches, inspired by their IRL counterparts [Abbeel and Ng, 2004], aim to find a constraint that maximally separates the expert's policy from other feasible policies. These methods often rely on linear programming or quadratic programming formulations. For instance, Scobee and Sastry [2020] proposed a maximum-margin framework that extracts constraints as logical formulas, providing a degree of interpretability. However, these methods often struggle with scalability and make strong linearity assumptions about the constraint function, limiting their application to high-dimensional or complex domains.

Probabilistic Methods frame ICRL as a problem of probabilistic inference, where the expert is assumed to be acting optimally under a maximum entropy principle [Ziebart et al., 2008]. Building on the Maximum Entropy IRL framework, Maximum Entropy ICRL [Kalweit et al., 2020] models the expert as following a Boltzmann distribution over trajectories, with the likelihood of a trajectory decreasing exponentially with its constraint cost. While probabilistic methods are more robust to noise and suboptimality in the demonstrations, they are often computationally expensive, as they typically require repeated policy optimization in an inner loop, making them impractical for large-scale problems.

Game-Theoretic and Adversarial Formulations cast ICRL as a two-player game between a constraint learner and a policy optimizer. In this minimax framework, the constraint learner aims to find a function that assigns low cost to expert trajectories and high cost to others, while the policy optimizer finds a policy that minimizes the expected cost [Chen et al., 2025]. Adversarial ICRL methods Peng et al. [2020], Wang et al. [2023] leverage generative adversarial networks (GANs) to implicitly learn constraints by distinguishing between expert and generated trajectories. While these methods can be highly expressive, they often suffer from training instability and can be difficult to balance, with the potential for the constraint learner and policy to enter degenerate equilibria.

Neurosymbolic Integration has recently emerged as a promising direction that seeks to combine the representational power of neural networks with the interpretability and structure of symbolic reasoning. Unlike purely symbolic methods that require hand-engineered state abstractions, neurosymbolic frameworks can learn relevant features directly from raw sensory input while maintaining the data efficiency and generalization benefits of symbolic priors. Popular work in this area focused on integrating differentiable logic layers, such as Signal Temporal Logic (STL), with neural perception modules Xu et al. [2022]. Alternatively, Chou et al. [2018] developed a grid-based neural network that maps raw states to predicate satisfaction probabilities, which are then composed via a differentiable STL robustness calculator. This enables end-to-end learning of spatio-temporal constraints from high-dimensional inputs. Similarly, Li et al. [2021] proposed a differentiable logic layer that allows for the integration of symbolic rule templates into deep RL policies. While representing the state-of-the-art, these neurosymbolic methods often rely on a flat predicate vector interface, which can suffer from semantic entanglement and temporal fragility – a core limitation our work aims to address.

2.2 Semantic Representations in Reinforcement Learning

The quest for meaningful, structured state representations is a central challenge in reinforcement learning, with implications for generalization, interpretability, and sample efficiency. This line of research spans several interconnected strands, including concept-based RL, successor representations, and geometric approaches, all of which inform our work.

Concept-based Reinforcement Learning aims to bridge the gap between high-dimensional sensory inputs and abstract, human-interpretable concepts. A primary goal is to learn a bottleneck of semantically meaningful features that can be used for policy learning and explanation. Shindo et al. [2025] proposed a framework for blending neural and symbolic representations, while Paleja et al. [2023] focused on learning a tree-based concept bottleneck for policy guidance. These methods demonstrate the value of abstraction but often rely on pre-defined or separately learned concepts, lacking an end-to-end mechanism for discovering the semantic structure most relevant for constraint learning.

Successor Representations (SR) [Dayan, 1993] and their modern reinforcement learning counterparts [Momennejad et al., 2017] offer a different perspective by decoupling the dynamics of the environment from the immediate rewards. An agent learns a representation that captures "what happens where you go," rather than "what is where." This allows for fast adaptation to new reward functions but is primarily concerned with state transition dynamics rather than the semantic or geometric properties of the state itself. Our work shares the motivation of creating more generalizable representations but focuses on the semantic structure of individual states as it pertains to constraint satisfaction.

Geometric and Topological Approaches leverage the inherent structure of data manifolds for representation learning. This includes methods that learn equivariant representations respecting environmental symmetries [Grinsztajn et al., 2020], as well as those using topological data analysis to understand the shape of the state space [Shahidullah, 2022]. Tiwari et al. [2025] recently proved that policy training dynamics naturally induce low-dimensional manifolds in state space, with dimensionality linked to the action space. This provides a theoretical basis for our geometric approach, where we explicitly construct a structured manifold in semantic space rather than discovering it implicitly in state space. While their work focuses on state space geometry for policy improvement, we extend this geometric perspective to semantic representations for constraint learning.

Our work sits at the confluence of these fields and draws explicit inspiration from the **Conceptual Spaces** framework in cognitive science [Gärdenfors, 2000], which posits that human knowledge is organized in geometric spaces where concepts are regions with natural metric properties. We instantiate this theory by constructing a geometric-semantic interface, connecting it to modern **Metric Learning** techniques [Musgrave et al., 2020] that aim to learn embeddings where distances meaningfully reflect semantic similarity. By doing so, we aim to create a representation space where the structure itself – the distances and regions within the HSV cylinder – directly encodes the semantics of safety and constraint satisfaction.

3 The Colorful Constraint Learning Framework

3.1 Problem Formulation

We consider the Inverse Constrained Reinforcement Learning (ICRL) problem within the Constrained Markov Decision Process (CMDP) framework. A CMDP is defined by the tuple $(\mathcal{S}, \mathcal{A}, T, R, C, \gamma, \rho_0, d)$, where:

- $\mathcal{S}$ is the state space
- $\mathcal{A}$ is the action space
- $T : \mathcal{S} \times \mathcal{A} \to \Delta(\mathcal{S})$ is the transition dynamics

- $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the reward function
- $C : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is the constraint function
- $\gamma \in [0, 1)$ is the discount factor
- ρ_0 is the initial state distribution
- $d \in \mathbb{R}$ is the constraint threshold

In standard RL, we seek an optimal policy that maximizes expected cumulative reward:

$$\pi^* = \arg\max_{\pi} \mathbb{E}_{\tau\sim\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right] \tag{1}$$

In constrained RL, we additionally require policy feasibility:

$$\mathbb{E}_{\tau\sim\pi}\left[\sum_{t=0}^{\infty} \gamma^t C(s_t, a_t)\right] \leq d \tag{2}$$

The ICRL problem inverts this formulation: given expert demonstrations $\mathcal{D}_{\text{expert}} = \{\tau_1, \tau_2, \ldots, \tau_N\}$ where each trajectory $\tau_i = \{(s_0, a_0), (s_1, a_1), \ldots, (s_T, a_T)\}$, and assuming the reward function R is known, we aim to recover the unknown constraint function C such that the expert policy appears optimal and feasible:

$$\pi_E = \arg\max_{\pi} \mathbb{E}_{\tau\sim\pi}\left[\sum_{t=0}^{\infty} \gamma^t R(s_t, a_t)\right] \quad \text{subject to} \quad \mathbb{E}_{\tau\sim\pi}\left[\sum_{t=0}^{\infty} \gamma^t C(s_t, a_t)\right] \leq d \tag{3}$$

3.1.1 Key Assumptions

To ensure identifiability of constraints, we make the following assumptions:

1. **Known Reward**: The task reward function R is provided, eliminating reward-constraint ambiguity.
2. **Constraint-Activation Diversity**: Expert demonstrations contain sufficient visits to constraint boundaries to make constraints identifiable:
$$\exists \tau \in \mathcal{D}_{\text{expert}} \text{ such that } \mathbb{E}[C(\tau)] \approx d \tag{4}$$
3. **Complementary Violation Data**: Access to some violation trajectories $\mathcal{D}_{\text{viol}}$ that explicitly break constraints, providing negative examples for contrastive learning.
4. **Smoothness Prior**: The constraint function C and its representation in semantic space vary smoothly with respect to state transitions.

3.2 The HSV Geometric-Semantic Interface

The core contribution of our framework is the introduction of a structured geometric-semantic interface that bridges the gap between high-dimensional sensory inputs and symbolic constraint reasoning. We propose using the HSV (Hue-Saturation-Value) color space as a universal semantic substrate, leveraging its inherent cylindrical topology to encode hierarchical semantic relationships and temporal evolution of concepts.

3.2.1 Geometric Motivation: Why HSV?

The HSV color space provides three key geometric properties that make it ideal as a semantic interface:

1. **Circular Topology**: The hue dimension forms a circle S^1, naturally representing cyclic semantic categories and preventing artificial boundary effects.
2. **Radial Hierarchy**: Saturation provides a natural radial coordinate where distance from center represents semantic confidence or specificity.
3. **Axial Prominence**: The value axis encodes semantic significance or activation strength, completing a cylindrical coordinate system.

Formally, the HSV space $\mathbb{H} \subset \mathbb{R}^3$ can be represented as a cylindrical manifold:

$$\mathbb{H} = \{(h, s, v) \in [0, 1) \times [0, 1] \times [0, 1]\} \tag{5}$$

with the identification $h \sim h + 1$ for the hue dimension, giving it the topology of $S^1 \times [0, 1]^2$.

The choice of HSV space is justified by both geometric principles and cognitive foundations:

- **Representational Efficiency**: The cylindrical topology of $\mathbb{H}$ provides a compact representation for hierarchical semantics without artificial discontinuities.
- **Cognitive Plausibility**: The hue-saturation-value structure mirrors psychological models of conceptual categorization, where basic-level categories (hue) are organized within broader taxonomic hierarchies.
- **Mathematical Tractability**: The well-defined metric structure of $\mathbb{H}$ enables rigorous analysis of semantic relationships and constraint satisfaction.

3.2.2 Semantic Embedding Formulation

Let $\mathbf{a}_t \in \mathbb{R}^k$ be a concept activation vector at time t, obtained from a neural perception module $g_\theta : \mathcal{S} \to \mathbb{R}^k$. Our semantic embedding $\phi : \mathbb{R}^k \to \mathbb{H}$ maps concept vectors to the HSV space through a principled statistical transformation.

Step 1: Statistical Whitening via PCA Given a dataset of concept vectors $\mathbf{A} = [\mathbf{a}_1, \mathbf{a}_2, \ldots, \mathbf{a}_T]^\top \in \mathbb{R}^{T \times k}$, we perform Principal Component Analysis (PCA) to capture the dominant semantic variations:

$$\mathbf{A} - \boldsymbol{\mu} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^\top \tag{6}$$

where $\boldsymbol{\mu} = \frac{1}{T}\sum_{t=1}^{T} \mathbf{a}_t$ is the mean concept vector, $\mathbf{V} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_k]$ are the principal components, and $\boldsymbol{\Sigma} = \text{diag}(\lambda_1, \lambda_2, \ldots, \lambda_k)$ contains the eigenvalues.

The projection onto principal components is:

$$\mathbf{p}_t = \mathbf{V}^\top(\mathbf{a}_t - \boldsymbol{\mu}) \tag{7}$$

where $\mathbf{p}_t = [p_{t,1}, p_{t,2}, \ldots, p_{t,k}]^\top$ represents the coordinates in the semantically aligned PCA space.

Step 2: Geometric Mapping to HSV Coordinates We define the HSV embedding $\phi(\mathbf{a}_t) = (H_t, S_t, V_t)$ as follows:

- **Hue (Semantic Category)**: The circular coordinate encoding primary semantic theme:

$$H_t = \frac{\text{atan2}(p_{t,2}, p_{t,1}) + 2\pi}{2\pi} \mod 1 \tag{8}$$

  This uses the first two principal components to define a circular coordinate system, where angular position represents semantic category. The atan2 function ensures correct quadrant determination, and the modulo operation enforces circular continuity.

- **Saturation (Semantic Confidence)**: The radial coordinate measuring confidence or specificity of semantic assignment:

$$S_t = \sqrt{\sum_{i=1}^{k} \frac{p_{t,i}^2}{\lambda_i}} \tag{9}$$

  This Mahalanobis-like distance normalizes by eigenvalue magnitudes, giving equal importance to each semantic dimension regardless of variance scaling.

- **Value (Semantic Prominence)**: The axial coordinate representing overall activation strength:

$$V_t = \tanh\left(\frac{\|\mathbf{a}_t\|_2}{\sigma}\right) \tag{10}$$

  where σ is a scaling parameter, typically set to $\sigma = \text{std}(\{\|\mathbf{a}_t\|_2\}_{t=1}^{T})$. The hyperbolic tangent ensures bounded output while preserving monotonicity.

3.2.3 Geometric Properties and Semantic Interpretation

The resulting embedding possesses several desirable geometric properties:

Proposition 1 (Semantic Distance Preservation)**.** *For concept vectors $\mathbf{a}, \mathbf{b} \in \mathbb{R}^k$, the chordal distance in HSV space approximates their semantic dissimilarity:*

$$d_{\mathbb{H}}(\phi(\mathbf{a}), \phi(\mathbf{b})) \approx \sqrt{d_{circ}^2(H_a, H_b) + (S_a - S_b)^2 + (V_a - V_b)^2} \tag{11}$$

where $d_{circ}(H_a, H_b) = \min(|H_a - H_b|, 1 - |H_a - H_b|)$ is the circular distance on hue.

Proof Sketch. The PCA transformation aligns the coordinate system with directions of maximal semantic variation. The hue coordinate captures angular relationships in the most significant semantic plane, while saturation and value encode magnitude information in a normalized fashion. □

3.2.4 Temporal Coherence via Recurrent Smoothing

To ensure temporal coherence in the semantic evolution, we introduce a recurrent smoothing mechanism on the HSV trajectories. Given a sequence of HSV points $\{\mathbf{c}_t\}_{t=1}^T$ where $\mathbf{c}_t = (H_t, S_t, V_t)$, we compute smoothed coordinates $\{\tilde{\mathbf{c}}_t\}_{t=1}^T$ as:

$$\tilde{\mathbf{c}}_t = \alpha \cdot \tilde{\mathbf{c}}_{t-1} + (1 - \alpha) \cdot \mathbf{c}_t^{\text{target}} \tag{12}$$

with special handling for the circular hue dimension:

$$\tilde{H}_t = \alpha \cdot \tilde{H}_{t-1} + (1 - \alpha) \cdot H_t^{\text{adjusted}} \tag{13}$$

$$H_t^{\text{adjusted}} = \begin{cases} H_t - 1 & \text{if } H_t - \tilde{H}_{t-1} > 0.5 \\ H_t + 1 & \text{if } \tilde{H}_{t-1} - H_t > 0.5 \\ H_t & \text{otherwise} \end{cases} \tag{14}$$

The smoothing parameter $\alpha \in (0, 1)$ controls the temporal coherence strength, with higher values enforcing smoother semantic transitions. This recurrent smoothing transforms discrete symbolic sequences into continuously evolving semantic trajectories, enabling robust temporal reasoning.

3.3 Three-Stage Learning Algorithm

As shown in Figure 1, our framework employs a progressive three-stage learning procedure that builds constraint understanding from basic semantic separation to complex temporal patterns. Each stage addresses a distinct aspect of constraint learning while maintaining the interpretability of the geometric-semantic interface.

3.3.1 Stage 1: Concept Space Foundation Learning

The first stage learns the concept embedding $g_\theta : \mathcal{S} \to \mathbb{R}^k$ that maps raw states to semantically meaningful concept vectors. We employ contrastive learning to separate expert and violator behaviors in the concept space.

Definition 1 (Contrastive Loss)**.** *Given batches of expert trajectories $\mathcal{B}_E$ and violator trajectories $\mathcal{B}_V$, the contrastive loss is defined as:*

$$\mathcal{L}_{contrast} = \alpha \mathcal{L}_{sep} + \beta \mathcal{L}_{exp\text{-}cluster} + \gamma \mathcal{L}_{viol\text{-}cluster} \tag{15}$$

where:

$$\mathcal{L}_{sep} = \frac{1}{|\mathcal{B}_E||\mathcal{B}_V|} \sum_{\tau_E \in \mathcal{B}_E} \sum_{\tau_V \in \mathcal{B}_V} \max(0, m - \|\boldsymbol{\mu}_E - \boldsymbol{\mu}_V\|_2)^2 \tag{16}$$

$$\mathcal{L}_{exp\text{-}cluster} = \frac{1}{|\mathcal{B}_E|(|\mathcal{B}_E| - 1)} \sum_{i \neq j} \|\boldsymbol{\mu}_{E_i} - \boldsymbol{\mu}_{E_j}\|_2^2 \tag{17}$$

$$\mathcal{L}_{viol\text{-}cluster} = \frac{1}{|\mathcal{B}_V|(|\mathcal{B}_V| - 1)} \sum_{i \neq j} \|\boldsymbol{\mu}_{V_i} - \boldsymbol{\mu}_{V_j}\|_2^2 \tag{18}$$

Here $\boldsymbol{\mu}_E = \frac{1}{T} \sum_{t=1}^T \mathbf{a}_t^E$ and $\boldsymbol{\mu}_V = \frac{1}{T} \sum_{t=1}^T \mathbf{a}_t^V$ are trajectory centroids, and $m > 0$ is a margin parameter.

This stage ensures that the concept space captures fundamental semantic distinctions between constraint-satisfying and constraint-violating behaviors.

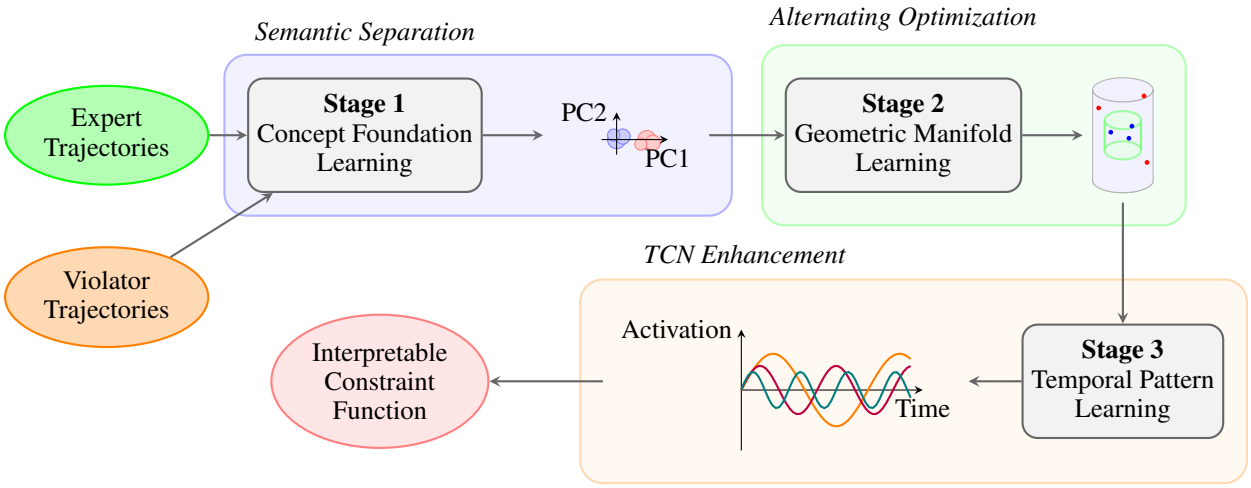

Figure 1: Complete pipeline of our colorful constraint learning framework. The three stages progressively build constraint understanding: semantic clustering in concept space (violet), geometric manifold learning in HSV space (green), and temporal pattern discovery (orange). Each stage enhances interpretability while maintaining separation between expert and violator behaviors.

3.3.2 Stage 2: Geometric Manifold Learning

The second stage learns a constraint manifold $\mathcal{M} \subset \mathbb{H}$ that defines the safe region in HSV space through alternating optimization.

Definition 2 (Cylindrical Constraint Manifold)**.** *We parameterize the constraint manifold as a cylindrical safe region:*

$$\mathcal{M} = \{\mathbf{c} = (h, s, v) \in \mathbb{H} : d(\mathbf{c}, \mathcal{M}) \leq 0\} \tag{19}$$

where the signed distance function is:

$$d(\mathbf{c}, \mathcal{M}) = \max\left(|h - h_c|_{circ} - w_h, |s - s_c| - w_s, |v - v_c| - w_v\right) \tag{20}$$

with parameters $\Theta_{\mathcal{M}} = (h_c, w_h, s_c, w_s, v_c, w_v)$ learned from data.

Definition 3 (Smooth Temporal Logic Operator)**.** *We define a differentiable Always operator for temporal constraint satisfaction:*

$$\rho_{\square}^{smooth}(\mathbf{C}, \mathcal{M}) = \xi - \frac{1}{\beta} \log\left(\frac{1}{T}\sum_{t=0}^{T-1}\int_0^1 \exp\left(\beta \cdot d(\tilde{\mathbf{c}}_t(\tau), \mathcal{M})\right) d\tau\right) \tag{21}$$

where $\tilde{\mathbf{c}}_t(\tau)$ is the linear interpolation between $\mathbf{c}_t$ and $\mathbf{c}_{t+1}$, $\xi > 0$ is a safety margin, and $\beta > 0$ controls approximation smoothness.

The alternating optimization proceeds in two phases:

1. **Phase A (Manifold Update)**: Fix concept network g_θ and update manifold parameters $\Theta_{\mathcal{M}}$ to maximize separation between expert and violator satisfaction scores.
2. **Phase B (Concept Refinement)**: Fix manifold $\mathcal{M}$ and refine g_θ using combined contrastive and temporal logic losses.

3.3.3 Stage 3: Temporal Pattern Enhancement

The final stage learns complex temporal patterns using Temporal Convolutional Networks (TCNs) that complement the geometric manifold.

Definition 4 (Temporal Pattern Bank)**.** *We employ a bank of N TCNs $\{TCN_j\}_{j=1}^N$ with sparse combination:*

$$C_{temporal}(\tau) = \sum_{j=1}^{N} w_j^{TCN} \cdot TCN_j(\mathbf{C}(\tau)) \tag{22}$$

where $\mathbf{C}(\tau) = [\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_T]$ *is the HSV trajectory, and weights* $\mathbf{w}^{TCN} \in \Delta^N$ *are learned with* ℓ_1 *regularization for sparsity.*

Definition 5 (TCN Architecture)**.** *Each TCN employs dilated causal convolutions:*

$$TCN_j(\mathbf{C}) = MLP\left(DilatedConv_L \circ \cdots \circ DilatedConv_1(\mathbf{C})\right) \tag{23}$$

with exponential dilation factors $[1, 2, 4, \ldots, 2^{L-1}]$ *to capture multi-scale temporal dependencies.*

3.3.4 Constraint Function

The learned color manifold naturally induces a differentiable cost function for policy optimization:

$$C(\tau) = \underbrace{\lambda_M \cdot C_{\text{manifold}}(\tau)}_{\text{spatial attraction}} + \underbrace{\lambda_T \cdot C_{\text{temporal}}(\tau)}_{\text{temporal patterns}} \tag{24}$$

where

$$C_{\text{manifold}}(\tau) = \frac{1}{T}\sum_{t=0}^{T} \text{Distance}(\mathbf{c}_t, \mathcal{M}) \tag{25}$$

$$\text{Distance}(\mathbf{c}_t, \mathcal{M}) = \xi - \rho_{\square}^{\text{smooth}}(\mathbf{C}(\tau), \mathcal{M}) \tag{26}$$

$$C_{\text{temporal}}(\tau) = \sum_{j=1}^{N} w_j^{\text{TCN}} \cdot TCN_j(\mathbf{C}(\tau); \Omega_j) \tag{27}$$

3.3.5 Complete Algorithm

The complete algorithm (Algorithm 1) progressively builds constraint understanding: Stage 1 establishes semantic foundations, Stage 2 learns geometric constraints in HSV space, and Stage 3 captures complex temporal patterns. This hierarchical approach ensures both interpretability and expressivity in the learned constraint representation.

4 Preliminary Experiments

We present preliminary experimental validation of our colorful constraint learning framework on a modified inverted pendulum environment. Our experiments demonstrate the framework's ability to learn interpretable constraint representations and separate expert from violator behaviors.

4.1 Experimental Setup

4.1.1 Environment and Constraints

We use a physics-accurate inverted pendulum environment defined by continuous state and action spaces:

- **State**: $s_t = [x, \dot{x}, \theta, \dot{\theta}]^\top \in \mathcal{S} \subset \mathbb{R}^4$
  - $x, \dot{x}$: Cart position (m) and velocity (m/s)
  - $\theta, \dot{\theta}$: Pole angle (rad) from upright and angular velocity (rad/s)
- **Action**: $a_t = F \in \mathcal{A} = [-F_{\text{max}}, F_{\text{max}}] \subset \mathbb{R}$
  - F: Force applied to cart (N)

The system dynamics are governed by the standard equations of motion for an inverted pendulum in a cart-pole system, derived using Lagrangian's equation of motion:

$$\ddot{x} = \frac{F - \mu_c \operatorname{sgn}(\dot{x}) - m_p l \dot{\theta}^2 \sin\theta + m_p g \sin\theta\cos\theta - \dfrac{\mu_p \dot{\theta}\cos\theta}{l}}{m_c + m_p \sin^2\theta} \tag{28}$$

$$\ddot{\theta} = \frac{g(m_c + m_p)\sin\theta - \dfrac{(m_c + m_p)\mu_p \dot{\theta}}{m_p l} + \cos\theta\left[F - \mu_c \operatorname{sgn}(\dot{x})\right] - m_p l \dot{\theta}^2 \sin\theta\cos\theta}{l(m_c + m_p \sin^2\theta)} \tag{29}$$

where:

Algorithm 1 Three-Stage Colorful Constraint Learning

Require: Expert trajectories $\mathcal{D}_{\text{expert}}$, violator trajectories $\mathcal{D}_{\text{viol}}$, margin $\epsilon > 0$
Ensure: Learned parameters $\Theta = \{\theta, \Theta_{\mathcal{M}}, \mathbf{w}^{\text{TCN}}, \Omega\}$

Stage 1: Foundation Learning
for $iteration = 1$ to N_1 **do**
  Sample batch $(\tau_E, \tau_V) \sim (\mathcal{D}_{\text{expert}}, \mathcal{D}_{\text{viol}})$
  Compute concept trajectories: $\mathbf{A}_E = g_\theta(\tau_E)$, $\mathbf{A}_V = g_\theta(\tau_V)$
  Compute HSV trajectories: $\mathbf{C}_E = \phi(\mathbf{A}_E)$, $\mathbf{C}_V = \phi(\mathbf{A}_V)$
  $\mathcal{L}_1 = \text{ColorContrastiveLoss}(\mathbf{C}_E, \mathbf{C}_V)$
  Update $\theta \leftarrow \theta - \eta_1 \nabla_\theta \mathcal{L}_1$
end for
Stage 2: Geometric Manifold Learning
for $alternation = 1$ to N_2 **do**
  Phase A: Update $\Theta_{\mathcal{M}}$ with fixed θ
  for $iteration = 1$ to K_A **do**
    Sample batch (τ_E, τ_V)
    $\mathbf{C}_E, \mathbf{C}_V \leftarrow \text{ComputeHSVTrajectories}(\tau_E, \tau_V)$
    $s_E = \rho_\square(\mathbf{C}_E, \mathcal{M})$, $s_V = \rho_\square(\mathbf{C}_V, \mathcal{M})$
    $\mathcal{L}_A = \max(0, \epsilon - (s_E - s_V))$
    Update $\Theta_{\mathcal{M}} \leftarrow \Theta_{\mathcal{M}} - \eta_2 \nabla_{\Theta_{\mathcal{M}}} \mathcal{L}_A$
  end for
  Phase B: Update θ with fixed $\Theta_{\mathcal{M}}$
  for $iteration = 1$ to K_B **do**
    Sample batch (τ_E, τ_V)
    $\mathcal{L}_B = \mathcal{L}_{\text{contrast}} + \lambda \mathcal{L}_A$
    Update $\theta \leftarrow \theta - \eta_3 \nabla_\theta \mathcal{L}_B$
  end for
end for
Stage 3: Temporal Pattern Learning
Freeze θ, $\Theta_{\mathcal{M}}$ {Foundation is fixed}
for $iteration = 1$ to N_3 **do**
  Sample batch (τ_E, τ_V)
  $\mathbf{C}_E, \mathbf{C}_V \leftarrow \text{ComputeHSVTrajectories}(\tau_E, \tau_V)$
  $s_E^{\text{TCN}} = \sum_j w_j^{\text{TCN}} \cdot \text{TCN}_j(\mathbf{C}_E; \Omega_j)$
  $s_V^{\text{TCN}} = \sum_j w_j^{\text{TCN}} \cdot \text{TCN}_j(\mathbf{C}_V; \Omega_j)$
  $\mathcal{L}_3 = \max(0, \epsilon - (s_E^{\text{TCN}} - s_V^{\text{TCN}})) + \lambda_{\text{reg}} \|\mathbf{w}^{\text{TCN}}\|_1$
  Update $\Omega, \mathbf{w}^{\text{TCN}} \leftarrow \Omega, \mathbf{w}^{\text{TCN}} - \eta_4 \nabla \mathcal{L}_3$
end for

return $\Theta = \{\theta, \Theta_{\mathcal{M}}, \mathbf{w}^{\text{TCN}}, \Omega\}$

- $x, \dot{x}, \ddot{x}$: cart position, velocity, and acceleration
- $\theta, \dot{\theta}, \ddot{\theta}$: pole angle, angular velocity, and angular acceleration
- m_c: cart mass, m_p: pole tip mass
- l: pole length, g: gravitational acceleration
- F: applied force, μ_c: cart friction, μ_p: pole hinge friction

These equations are integrated using a numerical method (e.g., Euler or Runge-Kutta) with timestep Δt.

The environment configuration is summarized in Table 1.

Figure 2 presents the schematics of the environment setup.

The unknown constraint to be learned is pole angle stability: expert policies maintain the pole within $\pm 20^\circ$ of vertical, while violator policies allow larger deviations that risk failure.

Parameter	Value	Description
Cart mass	1.0 kg	
Pole mass	0.1 kg	
Pole length	0.5 m	
Time step	0.02 s	Simulation resolution
Max force	10.0 N	Control limit
Max episode steps	500	10 seconds of simulation

Table 1: Inverted pendulum physical parameters.

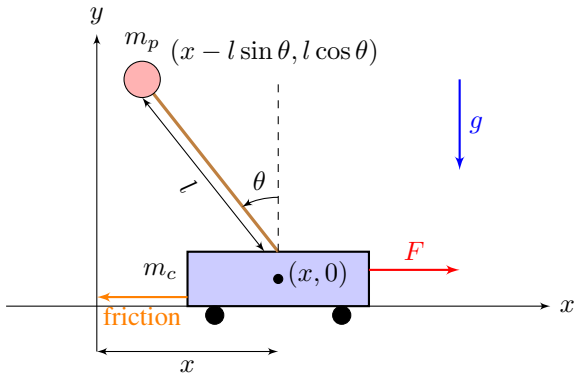

Figure 2: Schematic of the cart-pole system. The system consists of a cart of mass m_c moving horizontally along the x-axis, mounted with a massless pole of length l, to which is attached a tip mass m_p. The angle θ is measured from the vertical axis, positive counterclockwise. A force F is applied to the cart in the horizontal direction, with friction opposing cart motion. Gravity g acts downward.

4.1.2 Data Collection

We collect two types of demonstration data:

Dataset	Count	Length	Characteristics
Expert	100 trajectories	200 steps	Stable pole behavior
Violator	100 trajectories	200 steps	Unstable behavior

Table 2: Demonstration dataset composition.

Expert trajectories are generated using a Linear Quatratic Regulator (LQR) controller, while violator trajectories use destabilizing or random control policies. Each trajectory represents approximately 4 seconds of system dynamics. The data collection is summarized in Table 2.

4.1.3 Implementation Details

Our implementation uses the configuration summarized in Table 3:

4.2 Quantitative Result Analysis

4.2.1 Cluster Separation Metrics

We evaluate the separation between expert and violator behaviors using standard clustering metrics summarized in Table 4.

The progressive improvement across stages demonstrates the effectiveness of our hierarchical learning approach.

4.2.2 Constraint Satisfaction Accuracy

We evaluate the learned constraint function's accuracy in classifying expert vs. violator trajectories in Table 5.

Component	Parameter	Value
Concept Network	Hidden layers	[64, 32]
	Concept dimension	8
	Activation	ReLU
Color Embedding	PCA components	8
	Smoothing factor α	0.8
Training	Batch size	32
	Learning rate	1×10^{-3}
	Contrastive margin m	2.0
Manifold	Initial hue center	0.5
	Initial hue width	0.2
	Initial saturation center	0.6
	Initial saturation width	0.3
	Initial value center	0.5
	Initial value width	0.4

Table 3: Model architecture and training parameters.

Stage	Separation Ratio	Silhouette Score	DB Index
Initial	1.12	0.18	2.34
Stage 1	1.87	0.42	1.56
Stage 2	2.45	0.61	1.12
Stage 3	2.83	0.68	0.89

Table 4: Cluster separation metrics across learning stages. Higher separation ratio and silhouette score indicate better separation, while lower Davies-Bouldin (DB) index indicates tighter clustering.

Method	Precision	Recall	F1 Score
Stage 1 (Concepts only)	0.78	0.82	0.80
Stage 2 (+ Geometric)	0.88	0.91	0.89
Stage 3 (+ Temporal)	0.93	0.94	0.93

Table 5: Constraint classification performance. Each stage adds complementary information that improves classification accuracy.

4.2.3 Ablation Study

We conduct an ablation study to understand each component's contribution (Table 6).

The ablation study confirms that all components of our framework contribute to its performance, with recurrent smoothing and the cylindrical manifold providing particularly important geometric structure.

5 Limitations and Future Work

While our preliminary results demonstrate the feasibility of geometric-semantic interfaces for constraint learning, several limitations highlight directions for future research.

5.1 Theoretical Foundations

Our framework, while empirically motivated, lacks formal theoretical guarantees. The choice of HSV space, while geometrically appealing, is not derived from first principles. Future work should establish:

- **Theoretical justification** for the HSV topology based on information-theoretic principles. Long-term goal will be develop a unified semantic framework that generalizes beyond HSV space to other structured latent spaces.
- **Formal connections** to metric learning and optimal transport theory

Component Removed	Separation Ratio	F1 Score
Full framework	2.83	0.93
No recurrent smoothing	2.21	0.85
No contrastive loss	1.45	0.72
No temporal patterns	2.45	0.89
Linear manifold (vs. cylindrical)	2.12	0.83

Table 6: Ablation study results. Each component contributes significantly to overall performance.

- **Convergence guarantees** for the three-stage alternating optimization procedure

5.2 Empirical Validation

Our experiments are limited to simple dynamical systems with synthetic constraints:

- **Environment complexity**: The inverted pendulum represents a minimal test case; validation on high-dimensional systems like MuJoCo benchmarks or autonomous driving simulators is needed
- **Constraint diversity**: We focus on simple stability constraints; complex temporal logic specifications and multi-objective constraints require evaluation
- **Comparison baselines**: Comprehensive comparison against state-of-the-art ICRL methods [Yue et al., 2025, Cao and Xie, 2023, Fang et al., 2024] is absent

6 Conclusion

While our colorful constraint learning framework shows promise for interpretable neurosymbolic AI, significant work remains to establish its theoretical foundations, empirical robustness, and practical applicability. The limitations identified here provide a clear roadmap for future research, with the ultimate goal of developing AI systems that can learn, reason about, and explain their constraints in human-aligned ways. The geometric-semantic interface perspective offers a fertile ground for bridging the gap between neural representations and symbolic reasoning across multiple domains.

References

Pieter Abbeel and Andrew Ng. Apprenticeship learning via inverse reinforcement learning. *Proceedings, Twenty-First International Conference on Machine Learning, ICML 2004*, 09 2004. doi:10.1007/978-0-387-30164-8_417.

Kun Cao and Lihua Xie. Game-theoretic inverse reinforcement learning: A differential pontryagin's maximum principle approach. *IEEE Transactions on Neural Networks and Learning Systems*, 34(11):9506–9513, 2023. doi:10.1109/TNNLS.2022.3148376.

Yang Chen, Yitan Zhang, Gael Gendron, Mitchell Rogers, David Arturo Soriano Valdez, Mihailo Azhar, Shahrokh Heidari, Dr Padriac Amato Tahua O'Leary, Kobe Knowles, Jiamou Liu, Patrice Jean Delmas, and Michael Witbrock. Adversarial inverse reward-constraint learning with reward-feasibility contrast prior inspired by animal behaviour, 2025. URL `https://openreview.net/forum?id=eszQcR5F1e`.

Glen Chou, Dmitry Berenson, and Necmiye Ozay. Learning constraints from demonstrations with grid and parametric representations. *The International Journal of Robotics Research*, 40:1255 – 1283, 2018. URL `https://api.semanticscholar.org/CorpusID:53621830`.

Glen Chou, Dmitry Berenson, and Necmiye Ozay. Learning constraints from demonstrations, 2019. URL `https://arxiv.org/abs/1812.07084`.

Peter Dayan. Improving generalization for temporal difference learning: The successor representation. *Neural Computation*, 5(4):613–624, 1993. doi:10.1162/neco.1993.5.4.613.

Oliver Deane and Oliver Ray. Neuro-symbolic inverse constrained reinforcement learning. In Leilani H. Gilpin, Eleonora Giunchiglia, Pascal Hitzler, and Emile van Krieken, editors, *Proceedings of The 19th International Conference on Neurosymbolic Learning and Reasoning*, volume 284 of *Proceedings of Machine Learning Research*, pages 913–925. PMLR, 08–10 Sep 2025. URL `https://proceedings.mlr.press/v284/deane25a.html`.

Nan Fang, Guiliang Liu, and Wei Gong. Offline inverse constrained reinforcement learning for safe-critical decision making in healthcare, 2024. URL https://arxiv.org/abs/2410.07525.

P. Gärdenfors. *Conceptual Spaces: The Geometry of Thought*. MIT Press, 2000.

P. Gärdenfors. Conceptual spaces as a framework for knowledge representation. *Mind and Matter*, 2(2):9–27, 2004.

P. Gärdenfors. The geometry of meaning. In *The Geometry of Meaning*. MIT Press, 2014.

Yann Gilpin, Vince Kurtz, and Hai Lin. A smooth robustness measure of signal temporal logic for symbolic control, 2020. URL https://arxiv.org/abs/2006.05239.

Nathan Grinsztajn, Olivier Beaumont, Emmanuel Jeannot, and Philippe Preux. Geometric deep reinforcement learning for dynamic dag scheduling, 2020. URL https://arxiv.org/abs/2011.04333.

Gabriel Kalweit, Maria Huegle, Moritz Werling, and Joschka Boedecker. Deep inverse q-learning with constraints, 2020. URL https://arxiv.org/abs/2008.01712.

Xiao Li, Guy Rosman, Igor Gilitschenski, Jonathan DeCastro, Cristian-Ioan Vasile, Sertac Karaman, and Daniela Rus. Differentiable logic layer for rule guided trajectory prediction. In Jens Kober, Fabio Ramos, and Claire Tomlin, editors, *Proceedings of the 2020 Conference on Robot Learning*, volume 155 of *Proceedings of Machine Learning Research*, pages 2178–2194. PMLR, 16–18 Nov 2021. URL https://proceedings.mlr.press/v155/li21b.html.

Antonio Lieto, Antonio Chella, and Marcello Frixione. Conceptual spaces for cognitive architectures: A lingua franca for different levels of representation. *Biologically Inspired Cognitive Architectures*, 19:1–9, January 2017. ISSN 2212-683X. doi:10.1016/j.bica.2016.10.005. URL http://dx.doi.org/10.1016/j.bica.2016.10.005.

Francesco Locatello, Stefan Bauer, Mario Lucic, Gunnar Rätsch, Sylvain Gelly, Bernhard Schölkopf, and Olivier Bachem. Challenging common assumptions in the unsupervised learning of disentangled representations, 2019. URL https://arxiv.org/abs/1811.12359.

Corto Mascle, Nathanaël Fijalkow, and Guillaume Lagarde. Learning temporal formulas from examples is hard, 2023. URL https://arxiv.org/abs/2312.16336.

Ida Momennejad, Evan M Russek, Jin H Cheong, Matthew M Botvinick, Nathaniel Douglass Daw, and Samuel J Gershman. The successor representation in human reinforcement learning. *Nature human behaviour*, 1(9):680–692, 2017.

Kevin Musgrave, Serge Belongie, and Ser-Nam Lim. A metric learning reality check, 2020. URL https://arxiv.org/abs/2003.08505.

Rohan Paleja, Yaru Niu, Andrew Silva, Chace Ritchie, Sugju Choi, and Matthew Gombolay. Learning interpretable, high-performing policies for autonomous driving, 2023. URL https://arxiv.org/abs/2202.02352.

Xue Bin Peng, Angjoo Kanazawa, Sam Toyer, Pieter Abbeel, and Sergey Levine. Variational discriminator bottleneck: Improving imitation learning, inverse rl, and gans by constraining information flow, 2020. URL https://arxiv.org/abs/1810.00821.

Dexter R. R. Scobee and S. Shankar Sastry. Maximum likelihood constraint inference for inverse reinforcement learning, 2020. URL https://arxiv.org/abs/1909.05477.

Archie Shahidullah. Topological data analysis of neural network layer representations, 2022. URL https://arxiv.org/abs/2208.06438.

Hikaru Shindo, Quentin Delfosse, Devendra Singh Dhami, and Kristian Kersting. Blendrl: A framework for merging symbolic and neural policy learning, 2025. URL https://arxiv.org/abs/2410.11689.

Richard S. Sutton, Doina Precup, and Satinder Singh. Between mdps and semi-mdps: A framework for temporal abstraction in reinforcement learning. *Artificial Intelligence*, 112(1):181–211, 1999. ISSN 0004-3702. doi:https://doi.org/10.1016/S0004-3702(99)00052-1. URL https://www.sciencedirect.com/science/article/pii/S0004370299000521.

Christian Szegedy, Wojciech Zaremba, Ilya Sutskever, Joan Bruna, Dumitru Erhan, Ian Goodfellow, and Rob Fergus. Intriguing properties of neural networks, 2014. URL https://arxiv.org/abs/1312.6199.

Saket Tiwari, Omer Gottesman, and George Konidaris. Geometry of neural reinforcement learning in continuous state and action spaces, 2025. URL https://arxiv.org/abs/2507.20853.

Qi Wang, Yongsheng Hao, and Jiawei Zhang. Generative inverse reinforcement learning for learning 2-opt heuristics without extrinsic rewards in routing problems. *Journal of King Saud University - Computer and Information Sciences*, 35(9):101787, 2023. ISSN 1319-1578. doi:https://doi.org/10.1016/j.jksuci.2023.101787. URL https://www.sciencedirect.com/science/article/pii/S1319157823003415.

Ziwei Xu, Yogesh Rawat, Yongkang Wong, Mohan S Kankanhalli, and Mubarak Shah. Don't pour cereal into coffee: Differentiable temporal logic for temporal action segmentation. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 14890–14903. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/5f96a21345c138da929e99871fda138e-Paper-Conference.pdf.

Bo Yue, Jian Li, and Guiliang Liu. Provably efficient exploration in inverse constrained reinforcement learning, 2025. URL https://arxiv.org/abs/2409.15963.

B. D. Ziebart et al. Maximum entropy inverse reinforcement learning. In *AAAI Conference on Artificial Intelligence*, 2008.