Symmetries-enhanced Multi-Agent Reinforcement Learning

Nikolaos Bousias[1] NBOUSIAS@SEAS.UPENN.EDU
Stefanos Pertigkiozoglou[1] PSTEFANO@SEAS.UPENN.EDU
Kostas Daniilidis[1,2] KOSTAS@CIS.UPENN.EDU
George Pappas[1] PAPPASG@SEAS.UPENN.EDU
[1]GRASP Lab, University of Pennsylvania, Philadelphia, PA, [2]Archimedes, Athena RC

Editors: N. Ozay, L. Balzano, D. Panagou, A. Abate

Abstract

Multi-agent reinforcement learning has emerged as a powerful framework for enabling agents to learn complex, coordinated behaviors but faces persistent challenges regarding its generalization, scalability and sample efficiency. Recent advancements have sought to alleviate those issues by embedding intrinsic symmetries of the systems in the policy. Yet, most dynamical systems exhibit little to no symmetries to exploit. This paper presents a novel framework for embedding extrinsic symmetries in multi-agent system dynamics that enables the use of symmetry-enhanced methods to address systems with insufficient intrinsic symmetries, expanding the scope of equivariant learning to a wide variety of MARL problems. Central to our framework is the Group Equivariant Graphormer, a group-modular architecture specifically designed for distributed swarming tasks. Extensive experiments on a swarm of symmetry-breaking quadrotors validate the effectiveness of our approach, showcasing its potential for improved generalization and zero-shot scalability. Our method achieves significant reductions in collision rates and enhances task success rates across a diverse range of scenarios and varying swarm sizes.

Keywords: Representation Learning, Multi-agent RL, Swarm Robotics, Distributed Control

1. Introduction

The study of multi-agent dynamical systems, has garnered significant attention due to its wide-ranging applications in areas like autonomous navigation, environmental monitoring, target tracking, collaborative manipulation etc. Since obtaining large-scale datasets of expert demonstrations is often impractical, multi-agent reinforcement learning (MARL) has emerged as a powerful framework for enabling agents to learn complex, coordinated behaviors. However, the inherently high-dimensional and decentralized nature of multi-agent systems poses significant challenges, particularly in terms of scalability, generalization, and data efficiency. To enhance generalization, synthetic samples can be generated by applying symmetry transformations to the original data, increasing diversity without additional sampling (e.g. Laskin et al. (2020); Kostrikov et al. (2021)). Data augmentation, though, increases computation time and offers no generalization assurances. Outside of data augmentation, one promising avenue to address these challenges lies in leveraging the symmetries of dynamical systems as a policy inductive bias. This paper focuses on extrinsic symmetry exploitation in cooperative-competitive MARL problems.

From an RL perspective, the existence of symmetries morphs into equivalence of state-action pairs under a group transformation, meaning that the policy only needs to learn one mapping for each equivalence class, rather than separately learning the same behavior for all symmetrically related pairs Ravindran and Barto (2001); Zinkevich and Balch (2001). The presence of symmetries, therefore, reduces the hypothesis space and improves sample efficiency and generalization, as the policy learned for one pair generalizes to all equivalent pairs. Symmetry-aware methods can effectively operate on a reduced state-action space, corresponding to the equivalence classes rather than individual pairs (Yu and Lee (2022); Sonmez et al. (2024); Huang et al. (2024); Smith et al. (2024)). Though working with representations of the equivalence classes benefits from off-the-shelf neural architectures to parametrize the policy, it may restrict its expressivity, thus motivating the design of neural architectures that inherently respect the symmetries as an inductive bias (Mondal et al. (2020); van der Pol et al. (2020a,b); Zhu et al. (2022); Simm et al. (2020); Wang et al. (2022, 2020); Nguyen et al. (2023)). These approaches are only suitable for single agent or centralized multi-agent controllers and the neural networks used are tailored to the specificities of the system.

Omnipresent in homogeneous multi-agent systems is permutation equivariance, i.e. dynamics and reward functions are identical across robots, so indexing is interchangeable. This discrete S_n symmetry, leading to weight sharing among agents, is the basis of the representation power of GNNs (Jiang et al. (2020); Tzes et al. (2023)) and directly limits the required sample complexity for training the distributed policy. Permutation symmetries have been explored in MARL via policy sharing (HAO et al. (2023); Chen et al. (2022); Nguyen et al. (2017); Van der Pol and Oliehoek (2016); Lin and Lee (2024)), permutation equivariant critics (Almasan et al. (2022); Liu et al. (2019); Rashid et al. (2020); Liu et al. (2019)) or mean field approximation (Yang et al. (2018)). These approaches exploit the homogeneity structure but ignore the geometry properties of the problem. From a geometric standpoint, few symmetries in MARL have been explored, namely discrete C_n rotation symmetries van der Pol et al. (2022) and $E(3)$ Chen et al. (2023); Chen and Zhang (2023); McClellan et al. (2024). The existing works commonly employ equivariance-by-construction policies based on equivariant operations on message-passing networks (Satorras et al. (2021)) that assume $E(n)$ symmetry of the problem and, thus, are not transferable to MARL problems with different symmetries. Inspired by Wang et al. (2023); Hampsey et al. (2023), we propose a simple, yet modular, Group Equivariant Graphormer that can be adapted for a variety of symmetries in differently structured MARL problems via canonicalizing group actions on tensorial graph features. Existing literature focuses on exploiting existing symmetries in the systems, whereas the current paper extends to systems with partial or broken symmetries by embedding them in associated symmetry-enhanced systems to include desirable equivariant properties

1.1. Contributions

The aforementioned literature[1] utilizing symmetries as inductive biases in policy learning, which, however, presumes that the system exhibits said symmetries. In practice, most dynamical systems exhibit little to no symmetries. To the best of our knowledge, this is the first paper that attempts to construct a MARL framework that leverages symmetries in policies, even if the system is not endowed with them. Our contributions are outlined below:

- A formalization of the symmetry properties of multi-robot dynamical systems and the conditions under which optimal policies are equivariant functions to be approximated by equivariant networks. These conditions showcase the restrictive nature of current equivariant RL methods to systems with explicit symmetries shared by the task specification.
- A methodology for embedding extrinsic symmetries in systems where intrinsic symmetries are insufficient or broken, thus broadening the applicability of equivariant learning frameworks to any multi-agent dynamical system.

[1] Supplemental material can be found in Appendices I-IV of https://arxiv.org/pdf/2501.01136

- The introduction of a Group Equivariant Graphormer architecture tailored to distributed swarming tasks. This network is by design modular in terms of the symmetry considered (any group compatible with the manifold of the dynamics, discrete or continuous), in contrast to the vast majority of networks in the literature.

By embedding extrinsic symmetries, our approach demonstrates enhanced learning efficiency and generalization capabilities, paving the way for broader adoption of symmetry-aware methods in MARL and control. We subsequently validate its efficacy through extensive experimentation on a swarm of $SE(3)$ symmetry-breaking quadrotors that showcases the effectiveness of our scheme regarding scalability and generalization.

2. Preliminaries

2.1. Group Theory & Equivariant Functions

A group $(G, \cdot)$ is a set G equipped with an operator $\cdot : G \times G \to G$ that satisfies the properties of *Identity:* $\exists e \in G$ such that $e \cdot g = g \cdot e = e$; *Associativity:* $\forall g, h, f \in G$, $g \cdot (h \cdot f) = (g \cdot h) \cdot f$; *Inverse:* $\forall g \in G$, $\exists g^{-1}$ such that $g^{-1} \cdot g = g \cdot g^{-1} = e$. Additional to its structure we can define the way that the group elements act on a space X via a group action:

Definition 1 *A map $\phi_g : X \to X$ is called an action of group element $g \in G$ on X if for e is the identity element $\phi_e(x) = x$ for all $x \in X$ and $\phi_g \circ \phi_h = \phi_{g \cdot h}$ for all $g, h \in G$.*

Note here that a group action on a given space X allow us to group different elements of X in sets of orbits. More precisely given a group action ϕ_* an orbit of a element $x \in X$ is the set $\mathcal{O}_x^{\phi_*} = \{\phi_g(x) | g \in G\}$. In many application we require functions that respect the structure of a group acting on their domain and codomain. We refer to these functions as equivariant and we formally define them as follow:

Definition 2 *Given a group G and corresponding group actions $\phi_g : X \to X$, $\psi_g : X \to X$ for $g \in G$ a function $f : X \to Y$ is said to be (G, ϕ_*, ψ_*) equivariant if and only if:*

$$\psi_g \left[f(x) \right] = f \left(\phi_g[x] \right) \quad \forall x \in X, g \in G \tag{1}$$

2.2. Notation

Let $\mathcal{X}$ be a smooth manifold and $T_x\mathcal{X}$ the tangent space at an arbitrary $x \in \mathcal{X}$. A smooth vector field is a smooth map $f : \mathcal{X} \to T\mathcal{X}$ with $f(x) \in T_x\mathcal{X}$. The set of smooth vector fields on a manifold $\mathcal{X}$, denoted $\mathfrak{X}(\mathcal{X})$, is a linear infinite dimensional vector space. Let G be a d-dimension real Lie group, with identity element e. For a smooth manifold $\mathcal{X}$ and Lie group G, the left natural action $\phi : G \times \mathcal{X} \to \mathcal{X}$ satisfies $\phi(e, x) = x, \forall x \in \mathcal{X}$ and $\phi(\hat{g}, \phi(g, x)) = \phi(\hat{g} \cdot g, x), \forall g.\hat{g} \in G, x \in \mathcal{X}$, thus inducing families of smooth diffeomorphisms $\phi_g(x) := \phi(g, x)$. A group action is *free* if $\forall x \in \mathcal{X}, \phi(g, x) = x \Leftrightarrow g = e$ and *transitive* if $\forall x, y \in \mathcal{X}$, $\exists g \in G$ such that $\phi(g, x) = y$ (i.e. the nonlinear smooth projections $\phi_x(g) := \phi(g, x)$ are surjective). A *homogeneous* space is a smooth manifold $\mathcal{X}$ that admits a transitive group action $\phi : G \times \mathcal{X} \to \mathcal{X}$ and the Lie group G is, then, called the *symmetry* of $\mathcal{X}$. The group torsor $\mathfrak{G}$ of a Lie group G is defined as the underlying manifold of G without the group structure, allowing for identification of the torsor elements by the group elements, denoted $\chi \in \mathfrak{G} \simeq g \in G$, and inheriting the free and transitive group action ϕ induced by the group operator, i.e. for $\hat{g} \in G$ and $\chi \in \mathfrak{G} \simeq g \in G$ it stands that $\phi(\hat{g}, \chi) \simeq \hat{g} \cdot g$. Crucially, a manifold that serves as a torsor for multiple Lie groups may admit multiple symmetries.

2.3. Problem Statement

Consider a homogeneous multi-robot dynamical system, comprising of N autonomous robots indexed $i \in \{1, \ldots, N\} \equiv I_N$. Let $\mathcal{X}$ be a smooth manifold and $\mathcal{U}$ a finite dimensional input space. The agents are described by dynamics:

$$\dot{x}_i(t) = f(x_i(t), u_i(t)), \quad x_i(0) = x_i^0, \forall i \in I_N \tag{2}$$

where $u_i \in \mathcal{U}, x_i \in \mathcal{X}$ and $f : \mathcal{X} \times \mathcal{U} \to \mathfrak{X}(\mathcal{X})$ a linear morphism. Consider a Lie group G and a smooth transitive group action $\phi : G \times \mathcal{X} \to \mathcal{X}$.

Graph Representation of Multi-robot Systems: We assume that the robots are equipped with sensing/communication capabilities with a range $\hat{\epsilon}$. Let $\mathcal{N}_i := \{j \in I_N \setminus \{i\} \; : \; d(x_i, x_j) \leq \hat{\epsilon}\}$, $\forall i \in I_N$ denote the neighbourhoods, thus giving rise to a graph representation of the system $\mathcal{G}_t = \{V_{\mathcal{G}_t}, \mathcal{E}_{\mathcal{G}_t}\}$, with nodes $V_{\mathcal{G}_t} = \{x_i(t)\,,\, i \in I_N\}$ and edges $\mathcal{E}_{\mathcal{G}_t} = \{(i,j)\,,: \forall i \in I_N, j \in \mathcal{N}_i\}$. Information is propagated over the graph structure, with each robot receiving local observations of the system, denoted $o_i(t) = \{x_j(t)\,,\, j \in \mathcal{N}_i\}\,,\, i \in I_N$. We denote $x(t) = [\oplus_{\forall i \in I_N} x_i(t)] \in \cup_{i \in I_N} \mathcal{X}$ and $u(t) = [\oplus_{\forall i \in I_N} u_i(t)] \in \cup_{i \in I_N} \mathcal{U}$ the centralized state and action of the multi-robot system respectively. Assuming that a submanifold $\bar{X} \subseteq \mathcal{X}$ forms the torsor $\mathfrak{G}$ of the Lie group G and that $\mathcal{X} \setminus \bar{X}$ is compatible with G, then every robot inherits a group representation element $g_i\,, \forall i \in I_N$ and the node attributes of the graph become $V_{\mathcal{G}_t} = \{(g_i(t), x_i(t))\,,\, i \in I_N\}$.

Problem statement: Given a swarm of N robots of known dynamics, the trajectories $\{x_i(t)\,,\, i \in I_N\}, t \in \mathbb{R}^+$ evolve in a complex environment, known or estimated continuously through onboard sensors, with obstacles represented as a point cloud set $\mathcal{O} \in \mathbb{R}^{3 \times d}$. Assuming a metric function denoted $\mathcal{L} : \cup_{i \in I_N} \mathcal{X} \times \cup_{i \in I_N} \mathcal{U} \to \mathbb{R}$ that codifies the specific task (e.g. distance for navigation, alignment for flocking), we formally define the multi-agent geometric optimization problem

$$\begin{aligned} u^*_{0:T, I_N} = & \operatorname*{arg\,min}_{u_{0:T}} \int_0^T \mathcal{L}(\oplus_{I_N} x_i(\tau), \oplus_{I_N} u_i(\tau)) d\tau \\ s.t. \quad & \dot{x}_i(t) = f(x_i(t), u_i(t)) \in \mathfrak{X}(\mathcal{X})\,,\, i \in I_N \\ & u_i(t) = \pi(o_i(t) \cup x_i(t); \mathcal{O}) \in \mathcal{U} \end{aligned} \tag{3}$$

Problem (3) amounts to N robots learning the distributed control policy $\pi_\theta : \cup_{\mathcal{N}_i \cup \{i\}} \mathcal{X} \times \mathbb{R}^{3 \times d} \to \mathcal{U}$ using only local observations and knowledge of the environment. This problem can be recast as a distributed MARL problem by maximizing a reward function instead of minimizing a loss function, with robot dynamics described by a transition function, forming an MDP. For clarity reasons we opt to use the optimal control jargon.

3. Exploiting Symmetries of Dynamical Systems

In this section we show that if Problem 3 exhibits some structural symmetries then the optimal control policy needs to exhibit them as well, leading to sample efficient learning.

Definition 3 *Consider a real Lie Group G, a smooth manifold with group properties satisfying differentiability of group operations. Problem 3 is G-equivariant if, for transitive actions induced by elements of the group G on a vector field $\mathcal{X}$, $\phi : \mathcal{X} \times G \to \mathcal{X}$ and $\psi : \mathcal{U} \times G \to \mathcal{U}$, it satisfies the following properties:*

1. *The objective function is type-0 equivariant (invariant function), i.e.*
$$\mathcal{L}(\oplus_{I_N} x_i(t), \oplus_{I_N} u_i(t)) = \mathcal{L}(\oplus_{I_N} \phi_g(x_i(t)), \oplus_{I_N} \psi_g(u_i(t))),\ \forall g \in G$$
2. *The robot dynamics are equivariant w.r.t. elements of G, i.e.*
$$d\phi_g f(x_i(t), u_i(t)) = f(\phi_g(x_i(t)), \psi_g(u_i(t)),\ \forall g \in G$$

where $d\phi : \mathfrak{X}(\mathcal{X}) \times G \to \mathfrak{X}(\mathcal{X})$ the differential of the diffeomorphism defining the symmetry. In RL terminology, this is equivalent to requiring the transition, reward, and observations functions to be invariant under group actions, giving rise to the G-invariant Markov Games.

Assumption 1 *The neighborhood structure $\mathcal{N}_i, \forall i \in I_N$ is G-invariant.*

Theorem 1 *The optimal control policy $\pi^*(x_i(t) \cup o_i(t))$ for the G-equivariant multi-robot problem, i.e. equation (3) with definition (3), is equivariant under group actions from elements of G. (Proof is appended in Appendix I of the supplemental material[1].)*

Theorem 1 offers a way to shrink the hypothesis class of the learned controller to that of all G-equivariant functions. This inductive bias usually translates to fewer parameters required and greater sampling efficiency as multiple states $x(t)$ of the G-equivariant problem identified via the actions of the group G, are essentially equivalent, and therefore any G-equivariant policy trained on a dataset $D = \{x^1, x^2, \dots\}$ would generalize to the dataset $\hat{D} = \cup_{\forall g \in G}\{\phi_g(x^1), \phi_g(x^2), \dots\}$ where $D \subseteq \hat{D} \subseteq \cup_{I_N}\mathcal{X}$. Leveraging symmetries is particularly important in multi-agent RL as it is a sampling inefficient learning process.

4. When symmetries are broken

Section 3 demonstrated why symmetries of dynamical systems are useful. Even though most tasks exhibit geometric symmetries (condition 1 of Definition 3), most dynamical systems have little to no symmetries to exploit (condition 2), rendering Theorem 1 void. To circumvent this issue, we propose embedding the non-equivariant system into an extended dynamical system that is G-equivariant, so that Theorem 1 stands. The extended learned policy is G-equivariant, and optimal action for the original non-equivariant dynamical system is recovered via the symmetry-breaking projective function that guarantees equivalence between the two associated dynamical systems. This way we can provide structure in the learned policy, even if the original dynamical system does not support it. A consequence of this framework is that the structure of the policy is solely constrained by the symmetries that the reward function admits, meaning that any problem can be recast as a G-equivariant problem as long as condition 1 holds and the Lie group G is compatible with the manifold $\mathcal{X}$. Consider the push-forward smooth map $d_*\phi : \mathfrak{X}(\mathcal{X}) \times G \to \mathfrak{X}(\mathcal{X})$ defined as $d_*\phi_g f(x,u) := d\phi_g f(\phi_{g^{-1}}(x), u)$, naturally induced by the diffeomorphism $\phi : G \times \mathcal{X} \to \mathcal{X}$.

Lemma 1 *The push-forward function $d_*\phi$ is a well defined linear group action (automorphism) on the infinite dimensional vector field $\mathfrak{X}(\mathcal{X})$. (Proof in Appendix I of supplemental material[1].)*

Theorem 2 *Let $f : \mathcal{X} \times \mathcal{U} \to \mathfrak{X}(\mathcal{X})$ denote a control affine dynamical system on a smooth manifold $\mathcal{X}$ that admits a transitive group action from a compatible Lie group G, over an input vector space $\mathcal{U}$. For the extended input vector space $\hat{\mathcal{U}} := span\{d_*\phi_g f(x,u) \,|\, u \in \mathcal{U}, g \in G\}$, the associated system dynamics $F : \hat{\mathcal{U}} \times \mathcal{X} \to \mathfrak{X}(\mathcal{X})$ are equivariant with respect to actions induced by elements of the Lie group G. (Proof in Appendix I of supplemental material[1].)*

Leveraging theorem 2, any control affine dynamical system can be embedded in an associated equivariant dynamical system, by extending the input to include the closure of the image of the dynamics under the pushforward operator. The associated dynamics then are $F(x, \hat{u}) := \sum_{j=1}^{K} d_*\phi_{g_j} f(x, u_j)$, for $g_i \in G, u_i \in \mathcal{U}, K \in \mathbb{N}, x \in \mathfrak{X}(\mathcal{X}), \hat{u} \in \hat{\mathcal{U}}$. A limiting factor is that, even if the original input space $\mathcal{U}$ is finite dimensional, the extended input vector space $\hat{\mathcal{U}} \subset \mathfrak{X}(\mathcal{X})$ may be infinite dimensional. In practice, its dimension depends on the complexity of the dynamical system and the selected symmetry group.

Remark 1 *(Mahony et al. (2020) The trajectory of a control affine dynamical system $f : \mathcal{X} \times \mathcal{U} \to \mathfrak{X}(\mathcal{X})$ on a smooth manifold $\mathcal{X}$ is identical to the trajectory of its associated system dynamics $F : \hat{\mathcal{U}} \times \mathcal{X} \to \mathfrak{X}(\mathcal{X})$, defined in theorem (2), if the extended control input is constrained in $\mathcal{U}$, i.e. for some $u(t) \in \mathcal{U}$ $F(x(t), u(t)) = f(x(t), u(t))$.*

We can, then, reformulate problem (3) to incorporate artificial symmetries even if the dynamical system does not satisfy the condition 2 of definition 3 (only condition 1) as

$$\begin{aligned} u^*_{0:T,I_N} =& h_{\mathcal{U}}(\arg\min_{\hat{u}_{0:T}\in\hat{\mathcal{U}}} \int_0^T \mathcal{L}(\oplus_{I_N} x_i(\tau), \oplus_{I_N} h_{\mathcal{U}}(\hat{u}_i(\tau)))d\tau) \\ s.t. \quad & \dot{x}_i(t) = F(x_i(t), h_{\mathcal{U}}(\hat{u}_i(t))) \in \mathfrak{X}(\mathcal{X})\,,\ i \in I_N \\ & \hat{u}_i(t) = \hat{\pi}(o_i(t) \cup x_i(t); \mathcal{O}) \in \hat{\mathcal{U}} \end{aligned} \tag{4}$$

where $h_{\mathcal{U}} : \hat{\mathcal{U}} \to \mathcal{U}$ a smooth idempotent surjective morphism constraining the extended input to the original input vector space $\mathcal{U} \subset \hat{\mathcal{U}}$.

Proposition 2 *Problems (3) and (4) are equivalent and, if $\mathcal{L} : \mathcal{X} \to \mathbb{R}$ is G-invariant, there exists an optimal lifted $\hat{\pi}^* : \mathcal{X}^N \to \hat{\mathcal{U}}$ that is G-equivariant. (Proof in supplemental material[1].)*

From Proposition 2 there exists a G-equivariant extended optimal policy $\hat{\pi}^* : \mathcal{X}^{|\mathcal{N}_i|+1} \to \hat{\mathcal{U}}$ such that the optimal policy can be decomposed as $\pi^*(x_i(t), o_i(t)) = h_{\mathcal{U}} \circ \hat{\pi}^*(x_i(t), o_i(t))$. Assuming that the extended input space is finite-dimensional, the universal approximation theorem holds, meaning that the equivariant policy for the associated system can be ϵ-approximated by any G-equivariant neural network $\hat{\pi}_\theta$. Unfortunately, finding the associated equivariant dynamical system is a difficult process, even for systems with simple dynamics, so $h_{\mathcal{U}}$ is unknown and must also be learned by a neural network h_θ. Without the equivariant analogue system F, the policy $\pi_\theta(x_i, o_i) = h_\theta \circ \hat{\pi}_\theta(x_i, o_i) \in \mathcal{U}$ is fitted using data drawn from the original non-equivariant system f, whose trajectories identify with F, thus there is no guarantee that $\hat{\pi}_\theta \to \hat{\pi}^*$. Still, the existence of F alone guarantees that the compositional policy π_θ is valid, i.e. is a universal approximator of π^*. We demonstrate experimentally in Section 6 the benefits in generalization and scalability of parametrizing the policy as a composition of a G-equivariant and a non-equivariant neural network.

5. Equivariant Graphormer

To solve the Geometric Swarming problem, we must learn the distributed equivariant policy $\hat{u}_i(t) = \hat{\pi}(o_i \cup x_i; \mathcal{O})\,,\ i \in I_N$, as described in Section 4. Generally deep learning models require specialized architectures to ensure the satisfaction of the equivariant constraints Fuchs et al. (2020). Such architectures can result in more challenging optimization (Pertigkiozoglou et al., 2024) that can complicate their integration with standard RL techniques. To address these challenges we proposed to leverage the structure of our input graph representation and achieve equivariance through group canonicalization, described in Section 5.1. This technique doesn't impose any additional constraints to the model architecture and thus it allows us to easily extend pre-existing arcitectures used by previous work to learn distributed swarming policies. To that end, in Section 5.2 we describe an equivariant extension of the Graph Transformer Müller et al. (2023).

5.1. Group Canonicalization

Given a group G acting on space X through action ϕ_g we can define an extended group action ϕ'_g on space $G \times X$ as $\phi'_g[(p,x)] = (g \cdot p, \phi_g[x])$, where $\forall g, p \in G, x \in X$. Since ϕ_g is an action of group G, ϕ'_g satisfies the properties of definition 1 and, thus, it is also an action. Assuming an additional space Y with corresponding group action $\psi_g : G \times Y \to Y$, we can show the following:

Lemma 2 *A function $f : G \times X \to Y$ satisfies the equivariant constraint $f(\phi'_g[p]) = \psi_g[f(p)]$, $\forall g \in G$ and $p \in G \times X$, if and only if, for $h : X \to Y$, it can be written as a composition:*

$$f(g,x) = \psi_g[h(\phi_{g^{-1}}[x])], \quad \forall g \in G, x \in X$$

with $h(x) = f(e,x)$ for all $x \in X$ and e being the identity element of G. (Proof in Appendix I of the supplementary material[1].)

Notice that in Lemma 2, h is a function from X to Y without any additional constraints. This implies that we can use any general function approximator (e.g. MLP, Transformer) to approximate an equivariant function $f : G \times X \to Y$, by only applying the appropriate input-output transformations. We will use this observation to extend the baseline non-equivariant models to be equivariant with minimal changes to the underlying architecture.

5.2. Equivariant Graph Transformer

Given a feature augmented graph representation (V, E, F) with a finite set of nodes V, a finite set of edges $E(G) \subset \{(u,v)|u,v \in V\}$ and a set of per-node features $F = \{f_v \in X|v \in V\}$, a graph transformer sequentially updates the nodes features using a local attention layer to aggregate information from neighboring nodes. Specifically the l^{th} update layer for node $v \in V$ is a function $M_v : X^{(l)} \to X^{(l+1)}$ defined as $f_v^{(l+1)} \leftarrow M_v(F^{(l)}) = \mu(f_v^{(l)} + \mathrm{attn}(F^{(l)})_v)$, where μ corresponds to a fully-connected feedforward network, and attn corresponds to the local attention layer:

$$\mathrm{attn}(F^{(l)})_v = \sum_{p \in \mathcal{N}_v} \frac{\exp\left(f_v^{(l)^T} W_Q^T W_K f_p^{(l)}\right)}{\sum_{p \in \mathcal{N}_v} \exp\left(f_v^{(l)^T} W_Q^T W_K f_p^{(l)}\right)} \left(W_V f_p^{(l)}\right)$$

with $\mathcal{N}_v$ being the set of neighbors for nodes v. As discussed in Section 2.3, the input graph representation is endowed an additional structure that allows for an simple extention of the graph transformer to be equivariant. Specifically, each node $v \in V$ additional to a feature $f_v \in X$ describing each state is also equipped with a local frame $g_v \in G$. This means that the input graph is represented as (V, E, F_{tens}), with $F_{\text{tens}} = \{(g_v, f_v) \in G \times X|v \in V\}$ being a set of "tensorial" features that are described by their local frame $g_v \in G$ along with their feature value f_v. For such a feature (g_v, f_v) expressed in local frame $g_v \in G$ we can compute the equivalent feature

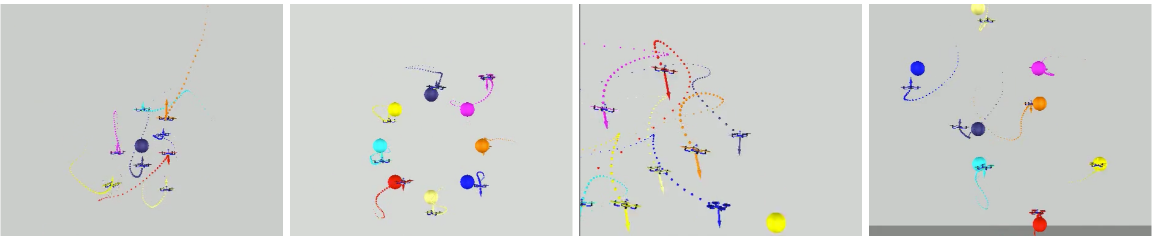

Figure 1: Instances of the swarm in various scenarios.

expressed in a new frame $g_n \in G$ by applying the action $\phi'_{g_n g_v^{-1}}[(g_v, f_v)] = (g_n, \phi_{g_n g_v^{-1}}[f_v])$. This structure allows us to leverage the results of Lemma 2 and define an equivariant update rule $M_v^{\text{eq}} : G \times X^{(l)} \to X^{(l+1)}$ for the features of node $v \in V$ as follows:

$$f_v^{(l+1)} \leftarrow M_v^{\text{eq}}(g_v, F^{(l)}) = \phi_{g_v}^{(l+1)}\left[M_v\left(\phi_{g_v^{-1}}^{(l)}\left[F^{(l)}\right]\right)\right]$$

with $\phi^{(l)}, \phi^{(l+1)}$ being actions of group G on the corresponding input/ouput feature space $X^{(l)}, X^{(l+1)}$. By sequentially composing the equivariant update layers we define an edge-preserving isomorphism, end-to-end Group Equivariant Graphomer, which is used to learn the equivariant policy $\hat{\pi}$. It is easy to see that the Group Equivariant Graphormer is permutation equivariant[1].

6. Experiments

To evaluate the effectiveness of incorporating artificial symmetries in MARL, we offer extensive experimentation on formation flight of swarms of N Crazyflie quadrotors. Quadrotors have few inherent geometric symmetries to exploit (Yu and Lee (2023a); Smith et al. (2024); Huang et al. (2024); Hampsey et al. (2023)), so equivariant MARL approaches like Chen and Zhang (2023); Yu and Lee (2023b) are not applicable. Our scheme, presented in sections 4-5, allows for the embedding of artificial $SE(3)$ symmetry in the distributed controllers, even though the quadrotor model is not $SE(3)$-equivariant, as gravity is $O(3)$-invariant. The multi-agent environment implementation is based on Huang et al. (2023), extended to include aerodynamic effects like drag and downwash, similarly to Panerati et al. (2021). For $x_i^d(t) \in \mathbb{R}^3$ desired target position, the state of each robot in the world-frame is $s_i(t) = (x_i(t), \dot{x}_i(t), R_i(t), \omega_i(t), x_i^d(t))$, where $R_i(t) \in SO(3)$ is the rotation matrix from the body frame to the world frame, $x_i(t) \in \mathbb{R}^3$ and $\omega_i(t) \in \mathbb{R}^3$ the position and angular acceleration in the world frame. Each robot receives neighborhood observations $o_i(t) = \cup_{j \in \mathcal{N}_i} s_j(t)$. The task is to learn a distributed, collision-free control policy that guides the swarm to a desired formation $x_i^d \in \mathbb{R}^{3 \times N}, \forall i \in I_N$, i.e. learn $\pi_\theta(u_i(t)|s_i(t), o_i(t))$ mapping state-observations to Gaussian distribution parametrized continuous actions $u_i(t)$. For enhanced generalization, a pool of geometric formations are used to construct diverse scenarios (static and dynamic formations, evasion pursuit etc.), e.g. Figure 1. The reward function for every quadrotor is $\bar{R}(s_i(t), o_i(t), u_i(t)) = \bar{R}^x + \bar{R}^c + \bar{R}^s$, where $\bar{R}^x = -c_1||x_i(t) - x_i^d(t)||$ a penalty motivating the quadrotor to approach the target, $\bar{R}^c = -c_2 \mathbf{1}_{||x_i(t) - x_j(t)||_2 < d_m \, \forall j \in \mathcal{N}_i} - c_3 \sum_{j \in \mathcal{N}_i} [1 - ||x_i(t) - x_j(t)||_2 / d_p]_+$ a collision penalty and $\bar{R}^s = -c_4||\omega_i(t)||_2 - c_5||u_i(t)||_2$ an auxiliary reward facilitating learning relatively stable controllers at the early stages of training.

Architecture & Training: If the neighborhood $\mathcal{N}_i$ is constructed via a Euclidean relative distance, for $SE(3)$ group actions Assumption 1 stands. From $(s_i(t), o_i(t))$ every agent constructs a local approximation $\mathcal{G}_{i,t}$ of the graph representation $\mathcal{G}_t$ from Section 2.3. As the rewards function is invariant to actions from $SE(3)$ and the S_N permutation group, via Section 4 the policy becomes

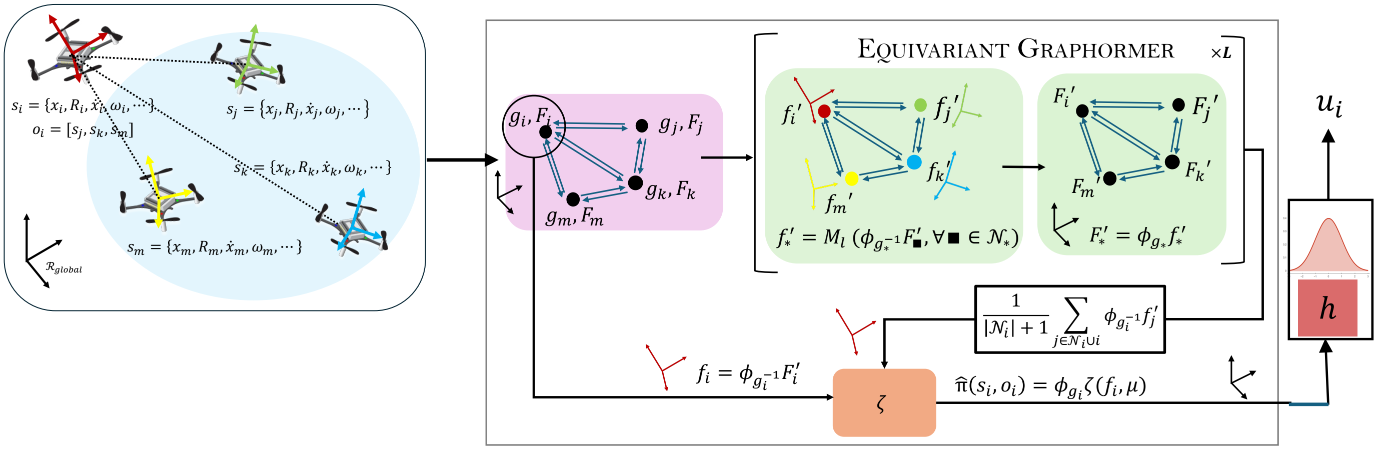
$s_i = \{x_i, R_i, \dot{x}_i, \omega_i, \cdots\}$
$o_i = [s_j, s_k, s_m]$
$s_j = \{x_j, R_j, \dot{x}_j, \omega_j, \cdots\}$
$s_k = \{x_k, R_k, \dot{x}_k, \omega_k, \cdots\}$
$s_m = \{x_m, R_m, \dot{x}_m, \omega_m, \cdots\}$
$\mathcal{R}_{global}$
g_i, F_i
g_j, F_j
g_k, F_k
g_m, F_m
EQUIVARIANT GRAPHORMER
×L
f_i'
f_j'
f_k'
f_m'
$f_*' = M_l\,(\phi_{g_*^{-1}} F_\blacksquare', \forall \blacksquare \in \mathcal{N}_*)$
F_i'
F_j'
F_k'
F_m'
$F_*' = \phi_{g_*} f_*'$
u_i
h
$\frac{1}{|\mathcal{N}_i| + 1} \sum_{j \in \mathcal{N}_i \cup i} \phi_{g_i^{-1}} f_j'$
$f_i = \phi_{g_i^{-1}} F_i'$
ζ
$\hat{\pi}(s_i, o_i) = \phi_{g_i} \zeta(f_i, \mu)$

Figure 2: Architecture schematic.

$\pi_\theta(u_i(t)|s_i(t), o_i(t)) = h_{\theta_p}(u_i(t)|\hat{\pi}_{\theta_e}(s_i(t), o_i(t)))$, where h_{θ_p} is parametrized by a simple MLP and $\hat{\pi}_{\theta_e}(s_i(t), o_i(t)) = \phi_{g_i}\zeta_{\theta_\zeta}(\phi_{g_i^{-1}} s_i(t) \oplus \frac{1}{|\mathcal{N}_i|+1} \sum_{j \in \mathcal{N}_i \cup \{i\}} \phi_{g_i^{-1}} f_j^L)$ the equivariant part of the policy with ζ_{θ_ζ} an MLP and $\{f_j^L, \forall j \in \mathcal{N}_i \cup \{i\}\} = M_{\theta_M}(\mathcal{G}_{i,t} \leftarrow s_i(t), o_i(t))$ the updated node features from the Group Equivariant Graphormer of Section 5. A schematic of our architecture is depicted in Figure 2. To train the distributed policy we use anasynchronous adaptation of the PPO algorithm from Petrenko et al. (2020). We use as baselines K-Attention and DeepSets Batra et al. (2021), rMAPPO Yu et al. (2022) and InfoMARL Nayak et al. (2023). Further details regarding architecture, hyperparameters, the reward function and training algorithm are offered in the supplemental material[1]. We, also, offer an adaptation of K-Attention with $SE(3)$ symmetry, following Lemma 2. It should be noted that if the manifold or part of it forms a torsor shared by various groups and the reward function is invariant to actions induced by them, the problem admits multiple extrinsic symmetries. The selection process of the group if multiple choices are valid is outside the scope of this paper and is left as future work.

Generalization: Table 1 summarizes the evaluation of the policies trained in all scenarios with a swarm comprising of 8 quadrotors. The metrics were averaged over 50 episodes per task and include the average collected reward per agent, the average distance to the targets, the average number of collisions with environment or other agents, the success rate (drones remaining within a small distance of their assigned target while avoiding collisions) and the inter-agent collision rate. The policies with embedded $SE(3)$ symmetry, notably the $SE(3)$-K-Attention, outperform the baselines across all scenarios, attaining increased rewards and leading to fewer collisions and increased success probability. The average collision rate of the $SE(3)$-Graphormer across tasks is 2.5% and 4.2% for $SE(3)$-K-Attention compared to $7.4 - 16.1\%$ for the baselines. DeepSets and InfoMARL are more aggresive, occasionally achieving better positional rewards, as observed by the average distance to target, at the expense of collision related rewards. The reduced collisions of the symmetry-enhanced policies can be explained by the fact that during training the collision instances are significantly fewer than instances without them where policies mainly optimize the position reward, meaning that even thought there are enough samples for the policy to learn target tracking, the same is not true for collision avoidance. The symmetry-enhanced policies exploit the structure of the collision avoidance specification to shrink the hypothesis class and, thus, effectively restricting the models and leading to sampling efficient and accurate learning of collision avoidance without loss of expressivity of the final policy. Overall, the $SE(3)$-enhanced policies successfully

SCENARIOS

	Static same goal					Dynamic same goal					Evasion Pursuit (Lissajous)					Swarm-vs-Swarm				
	Rew	Dist	Col	Suc %	IColR	Rew	Dist	Col	Suc %	IColR	Rew	Dist	Col	Suc %	IColR	Rew	Dist	Col	Suc %	IColR
K-Attention	-6.49	0.56	1.25	71.9	0.14	-7.28	0.77	1.5	65.6	0.17	-8.11	1.25	1.88	63.5	0.203	-5.16	0.83	1	78.1	0.218
rMAPPO	-7.86	0.70	2.13	51.6	0.21	-8.54	1.49	1.75	53.1	0.205	-8.65	1.59	1	71.9	0.125	-8.13	0.71	2	46.8	0.235
InfoMARL	-2.73	0.42	0.05	98.8	0.013	-4.38	0.65	0.5	89.4	0.106	-5.68	0.83	0.68	85.1	0.148	-4.15	**0.50**	0.71	83.3	0.167
K-DeepSets	-3.24	0.44	0.1	96.9	0.006	-4.43	0.69	0.187	96.1	0.039	-5.05	**0.69**	0.51	88.1	0.117	-4.18	0.60	0.46	89.2	0.107
SE(3)-K-Attention (Ours)	-2.44	0.33	**0**	**100**	**0**	**-3.91**	0.87	**0**	**100**	**0**	-4.87	0.86	**0.13**	95.3	**0.031**	-3.99	0.53	0.55	87.7	0.123
SE(3)-Graphormer (Ours)	**-2.41**	**0.32**	**0**	**100**	**0**	-4.21	**0.64**	0.125	96.8	0.031	**-4.53**	0.74	**0.13**	**96.8**	**0.031**	**-3.47**	0.52	**0.33**	**92.7**	**0.073**
	Static different goals					**Dynamic different goals**					**Evasion Pursuit (Bezier)**					**Swap goals**				
	Rew	Dist	Col	Suc %	IColR	Rew	Dist	Col	Suc %	IColR	Rew	Dist	Col	Suc %	IColR	Rew	Dist	Col	Suc %	IColR
K-Attention	-3.88	0.41	0.5	87.5	0.063	-5.37	0.74	0.92	80.2	0.099	-11.52	1.35	3	31.2	0.335	-3.91	0.57	0.50	88.6	0.113
rMAPPO	-3.92	0.55	0.33	80.2	0.04	-6.33	1.01	1.17	74.4	0.119	-11.8	2.23	1.08	45.8	0.225	-4.01	0.81	0.5	81.2	0.126
InfoMARL	-2.29	**0.18**	0.25	94.3	0.057	-3.97	**0.60**	0.42	90.5	0.104	-6.73	1.18	0.84	78.9	0.183	-3.09	0.41	0.47	88.7	0.113
K-DeepSets	-2.73	0.32	0.07	98.1	0.017	-4.43	0.72	0.38	90.9	0.091	-6.14	1.03	0.625	86.7	0.129	-3.46	0.42	0.375	91.4	0.086
SE(3)-K-Attention (Ours)	-1.91	0.24	0.03	99.2	0.007	-3.88	0.82	0.23	94.6	0.054	-6.71	1.46	0.37	80.2	0.082	**-2.71**	**0.40**	0.15	96.2	0.038
SE(3)-Graphormer (Ours)	**-1.87**	0.21	**0.02**	**99.3**	**0.003**	**-3.18**	0.81	**0**	**100**	**0**	**-5.63**	**0.97**	**0.125**	**89.0**	**0.031**	-2.82	0.55	**0.125**	**96.9**	**0.031**

Table 1: Evaluation for a swarm of 8 quadrotors in various scenarios.

manage to guide the quadrotors to the targets with significantly fewer collisions as indicated by the increased success rate.

Scalability Ablation Studies: We examine the impact of extrinsic symmetries on scalability, i.e. using pretrained policy to control swarms of increasing size with zero-shot learning. The policies, trained for swarms of 8 quadrotors, are evaluated without further training in swarms of 16-128 quadrotors. The metrics are averaged over 50 episodes of static and dynamic scenarios (Table 2). As the room size remains unchanged, larger swarms lead to cluttered space, significantly increasing collisions affecting multiple drones in a cascade effect. However, the embedded symmetries leveraged locally are not affected by the size of the swarm. The symmetry-enhanced policies exhibit significantly fewer collisions than the baselines (for 128 drones the $SE(3)$-enhanced policies' collision rate rises to $\simeq 16\%$ compared to $35-40\%$ for the baselines) as the swarm size progressively increases, while approaching reasonably the designated targets (97% and 84.64% success rate in swarms of 64,128 compared to $< 91\%$ and $< 65\%$ for the baselines).

SWARM SIZE

	16				32				64				128			
	Rew	Col	Suc %	IColR	Rew	Col	Suc %	IColR	Rew	Col	Suc %	IColR	Rew	Col	Suc %	IColR
K-Attention	-5.38	1.6	84.6	0.154	-5.40	2.8	86.6	0.134	-5.58	7.6	82.8	0.172	-7.70	69.2	58.9	0.406
rMAPPO	-6.08	1.083	79.2	0.119	-6.80	1.83	81.8	0.104	-7.79	16	57.2	0.352	-11.88	58.167	48.4	0.457
InfoMARL	-4.07	0.343	96.7	0.033	-4.72	2.333	88.02	0.119	-5.19	11.25	77.3	0.227	-7.03	40.29	64.6	0.351
K-DeepSets	-**3.38**	0.2	97.5	0.025	-3.92	1.875	90.2	0.098	-4.56	3.8	90.2	0.121	-10.51	163.17	64.9	0.348
SE(3)K-Attention (Ours)	-4.09	**0.067**	**99.2**	0.008	-4.41	0.583	96.6	0.034	-4.96	2.43	93.7	0.063	-5.66	**12.75**	83.77	0.16
SE(3)-Graphormer (Ours)	-3.45	0.133	98.8	**0.0125**	**-3.69**	**0.375**	**98.7**	**0.023**	**-4.29**	**0.82**	**97.5**	**0.025**	**-5.47**	14.86	**84.64**	**0.154**

Table 2: Evaluation of zero-shot transferability to growing swarms.

7. Conclusions & Future Work

This paper provides a novel methodology for leveraging extrinsic symmetries, indicated solely by the task, for systems without intrinsic ones and introduce the Group Equivariant Graphormer, an equivariant neural architecture adaptable to different symmetry groups. The experimental results of our work suggest that embedding extrinsic equivariance in MARL policies is beneficial for the generalization, scalability and sample efficiency. However, a task may be invariant to multiple symmetry groups that are compatible with the manifold but not exhibited by the dynamics. So how do we pick the symmetry group? We leave the symmetry selection strategy for future work.

Acknowledgments

This work was supported by the ARL grant DCIST CRA W911NF-17- 2-0181. The authors thank Evangelos Chatzipantazis for the helpful discussions.

References

Paul Almasan, José Suárez-Varela, Krzysztof Rusek, Pere Barlet-Ros, and Albert Cabellos-Aparicio. Deep reinforcement learning meets graph neural networks: Exploring a routing optimization use case. *Computer Communications*, 196:184–194, December 2022. ISSN 0140-3664. doi: 10.1016/j.comcom.2022.09.029. URL http://dx.doi.org/10.1016/j.comcom.2022.09.029.

Sumeet Batra, Zhehui Huang, Aleksei Petrenko, Tushar Kumar, Artem Molchanov, and Gaurav S. Sukhatme. Decentralized control of quadrotor swarms with end-to-end deep reinforcement learning. In *5th Conference on Robot Learning, CoRL 2021, 8-11 November 2021, London, England, UK*, Proceedings of Machine Learning Research. PMLR, 2021. URL https://arxiv.org/abs/2109.07735.

Dingyang Chen and Qi Zhang. E(3)-equivariant actor-critic methods for cooperative multi-agent reinforcement learning. *arXiv preprint arXiv:2308.11842*, 2023.

Dingyang Chen, Yile Li, and Qi Zhang. Communication-efficient actor-critic methods for homogeneous markov games. In *International Conference on Learning Representations*, 2022. URL https://openreview.net/forum?id=xy_2w3J3kH.

Runfa Chen, Jiaqi Han, Fuchun Sun, and Wenbing Huang. Subequivariant graph reinforcement learning in 3D environments. In Andreas Krause, Emma Brunskill, Kyunghyun Cho, Barbara Engelhardt, Sivan Sabato, and Jonathan Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 4545–4565. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/chen23i.html.

Fabian B. Fuchs, Daniel E. Worrall, Volker Fischer, and Max Welling. Se(3)-transformers: 3d roto-translation equivariant attention networks. In *Advances in Neural Information Processing Systems 34 (NeurIPS)*, 2020.

Matthew Hampsey, Pieter van Goor, Tarek Hamel, and Robert Mahony. Exploiting different symmetries for trajectory tracking control with application to quadrotors*. *IFAC-PapersOnLine*, 56(1):132–137, 2023. ISSN 2405-8963. doi: https://doi.org/10.1016/j.ifacol.2023.02.023. URL https://www.sciencedirect.com/science/article/pii/S2405896323002112. 12th IFAC Symposium on Nonlinear Control Systems NOLCOS 2022.

Jianye HAO, Xiaotian Hao, Hangyu Mao, Weixun Wang, Yaodong Yang, Dong Li, YAN ZHENG, and Zhen Wang. Boosting multiagent reinforcement learning via permutation invariant and permutation equivariant networks. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=OxNQXyZK-K8.

Junchang Huang, Weifeng Zeng, Hao Xiong, Bernd R. Noack, Gang Hu, Shugao Liu, Yuchen Xu, and Huanhui Cao. Symmetry-informed reinforcement learning and its application to low-level attitude control of quadrotors. *IEEE Transactions on Artificial Intelligence*, 5(3):1147–1161, 2024. doi: 10.1109/TAI.2023.3249683.

Zhehui Huang, Sumeet Batra, Tao Chen, Rahul Krupani, Tushar Kumar, Artem Molchanov, Aleksei Petrenko, James A Preiss, Zhaojing Yang, and Gaurav S Sukhatme. Quadswarm: A modular multi-quadrotor simulator for deep reinforcement learning with direct thrust control. *arXiv preprint arXiv:2306.09537*, 2023.

Jiechuan Jiang, Chen Dun, Tiejun Huang, and Zongqing Lu. Graph convolutional reinforcement learning, 2020. URL https://arxiv.org/abs/1810.09202.

Ilya Kostrikov, Denis Yarats, and Rob Fergus. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels, 2021. URL https://arxiv.org/abs/2004.13649.

Misha Laskin, Kimin Lee, Adam Stooke, Lerrel Pinto, Pieter Abbeel, and Aravind Srinivas. Reinforcement learning with augmented data. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 19884–19895. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/e615c82aba461681ade82da2da38004a-Paper.pdf.

Bor-Jiun Lin and Chun-Yi Lee. HGAP: Boosting permutation invariant and permutation equivariant in multi-agent reinforcement learning via graph attention network. In Ruslan Salakhutdinov, Zico Kolter, Katherine Heller, Adrian Weller, Nuria Oliver, Jonathan Scarlett, and Felix Berkenkamp, editors, *Proceedings of the 41st International Conference on Machine Learning*, volume 235 of *Proceedings of Machine Learning Research*, pages 30189–30210. PMLR, 21–27 Jul 2024. URL https://proceedings.mlr.press/v235/lin24m.html.

Iou-Jen Liu, Raymond A. Yeh, and Alexander G. Schwing. Pic: Permutation invariant critic for multi-agent deep reinforcement learning, 2019. URL https://arxiv.org/abs/1911.00025.

Robert E. Mahony, T. Hamel, and Jochen Trumpf. Equivariant systems theory and observer design. *ArXiv*, abs/2006.08276, 2020. URL https://api.semanticscholar.org/CorpusID:219687199.

Joshua McClellan, Naveed Haghani, John Winder, Furong Huang, and Pratap Tokekar. Boosting sample efficiency and generalization in multi-agent reinforcement learning via equivariance, 2024. URL https://arxiv.org/abs/2410.02581.

Arnab Kumar Mondal, Pratheeksha Nair, and Kaleem Siddiqi. Group equivariant deep reinforcement learning, 2020. URL https://arxiv.org/abs/2007.03437.

Luis Müller, Mikhail Galkin, Christopher Morris, and Ladislav Rampášek. Attending to graph transformers, 2023.

Siddharth Nayak, Kenneth Choi, Wenqi Ding, Sydney Dolan, Karthik Gopalakrishnan, and Hamsa Balakrishnan. Scalable multi-agent reinforcement learning through intelligent information aggregation. In *Proceedings of the 40th International Conference on Machine Learning*, ICML'23. JMLR.org, 2023.

Duc Thien Nguyen, Akshat Kumar, and Hoong Chuin Lau. Policy gradient with value function approximation for collective multiagent planning. In I. Guyon, U. Von Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/c2ba1bc54b239208cb37b901c0d3b363-Paper.pdf.

Hai Huu Nguyen, Andrea Baisero, David Klee, Dian Wang, Robert Platt, and Christopher Amato. Equivariant reinforcement learning under partial observability. In *7th Annual Conference on Robot Learning*, 2023. URL https://openreview.net/forum?id=AnDDMQgM7-.

Jacopo Panerati, Hehui Zheng, SiQi Zhou, James Xu, Amanda Prorok, and Angela P. Schoellig. Learning to fly—a gym environment with pybullet physics for reinforcement learning of multi-agent quadcopter control. In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 7512–7519, 2021. doi: 10.1109/IROS51168.2021.9635857.

Stefanos Pertigkiozoglou, Evangelos Chatzipantazis, Shubhendu Trivedi, and Kostas Daniilidis. Improving equivariant model training via constraint relaxation. In A. Globerson, L. Mackey, D. Belgrave, A. Fan, U. Paquet, J. Tomczak, and C. Zhang, editors, *Advances in Neural Information Processing Systems*, volume 37, pages 83497–83520. Curran Associates, Inc., 2024. URL https://proceedings.neurips.cc/paper_files/paper/2024/file/98082e6b4b97ab7d3af1134a5013304e-Paper-Conference.pdf.

Aleksei Petrenko, Zhehui Huang, Tushar Kumar, Gaurav S. Sukhatme, and Vladlen Koltun. Sample factory: Egocentric 3d control from pixels at 100000 FPS with asynchronous reinforcement learning. In *Proceedings of the 37th International Conference on Machine Learning, ICML 2020, 13-18 July 2020, Virtual Event*, volume 119 of *Proceedings of Machine Learning Research*, pages 7652–7662. PMLR, 2020. URL http://proceedings.mlr.press/v119/petrenko20a.html.

Tabish Rashid, Mikayel Samvelyan, Christian Schroeder de Witt, Gregory Farquhar, Jakob Foerster, and Shimon Whiteson. Monotonic value function factorisation for deep multi-agent reinforcement learning, 2020. URL https://arxiv.org/abs/2003.08839.

Balaraman Ravindran and Andrew G. Barto. Symmetries and model minimization in markov decision processes. 2001. URL https://api.semanticscholar.org/CorpusID:59092891.

Víctor Garcia Satorras, Emiel Hoogeboom, and Max Welling. E(n) equivariant graph neural networks. In Marina Meila and Tong Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 9323–9332. PMLR, 18–24 Jul 2021. URL https://proceedings.mlr.press/v139/satorras21a.html.

Gregor N. C. Simm, Robert Pinsler, Gábor Csányi, and José Miguel Hernández-Lobato. Symmetry-aware actor-critic for 3d molecular design, 2020. URL https://arxiv.org/abs/2011.12747.

Henry Smith, Ajay Shankar, Jennifer Gielis, Jan Blumenkamp, and Amanda Prorok. So(2)-equivariant downwash models for close proximity flight. *IEEE Robotics Autom. Lett.*, 9(2): 1174–1181, February 2024. URL https://doi.org/10.1109/LRA.2023.3337701.

Yasin Sonmez, Neelay Junnarkar, and Murat Arcak. Exploiting symmetry in dynamics for model-based reinforcement learning with asymmetric rewards, 2024. URL https://arxiv.org/abs/2403.19024.

Mariliza Tzes, Nikolaos Bousias, Evangelos Chatzipantazis, and George J Pappas. Graph neural networks for multi-robot active information acquisition. In *2023 IEEE International Conference on Robotics and Automation (ICRA)*, pages 3497–3503. IEEE, 2023.

Elise Van der Pol and Frans A Oliehoek. Coordinated deep reinforcement learners for traffic light control. *Proceedings of learning, inference and control of multi-agent systems (at NIPS 2016)*, 8: 21–38, 2016.

Elise van der Pol, Thomas Kipf, Frans A. Oliehoek, and Max Welling. Plannable approximations to mdp homomorphisms: Equivariance under actions, 2020a. URL https://arxiv.org/abs/2002.11963.

Elise van der Pol, Daniel Worrall, Herke van Hoof, Frans Oliehoek, and Max Welling. Mdp homomorphic networks: Group symmetries in reinforcement learning. In H. Larochelle, M. Ranzato, R. Hadsell, M.F. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 4199–4210. Curran Associates, Inc., 2020b. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/2be5f9c2e3620eb73c2972d7552b6cb5-Paper.pdf.

Elise van der Pol, Herke van Hoof, Frans A. Oliehoek, and Max Welling. Multi-agent mdp homomorphic networks, 2022. URL https://arxiv.org/abs/2110.04495.

Dian Wang, Colin Kohler, and Robert Platt. Policy learning in se(3) action spaces, 2020. URL https://arxiv.org/abs/2010.02798.

Dian Wang, Robin Walters, Xupeng Zhu, and Robert Platt. Equivariant q learning in spatial action spaces. In Aleksandra Faust, David Hsu, and Gerhard Neumann, editors, *Proceedings of the 5th Conference on Robot Learning*, volume 164 of *Proceedings of Machine Learning Research*, pages 1713–1723. PMLR, 08–11 Nov 2022. URL https://proceedings.mlr.press/v164/wang22j.html.

Dian Wang, Jung Yeon Park, Neel Sortur, Lawson L.S. Wong, Robin Walters, and Robert Platt. The surprising effectiveness of equivariant models in domains with latent symmetry. In *The Eleventh International Conference on Learning Representations*, 2023. URL https://openreview.net/forum?id=P4MUGRM4Acu.

Yaodong Yang, Rui Luo, Minne Li, Ming Zhou, Weinan Zhang, and Jun Wang. Mean field multi-agent reinforcement learning. In Jennifer Dy and Andreas Krause, editors, *Proceedings of the 35th International Conference on Machine Learning*, volume 80 of *Proceedings of Machine Learning Research*, pages 5571–5580. PMLR, 10–15 Jul 2018. URL https://proceedings.mlr.press/v80/yang18d.html.

Beomyeol Yu and Taeyoung Lee. Equivariant reinforcement learning for quadrotor uav. *2023 American Control Conference (ACC)*, pages 2842–2847, 2022. URL https://api.semanticscholar.org/CorpusID:249375239.

Beomyeol Yu and Taeyoung Lee. Equivariant reinforcement learning for quadrotor uav. In *2023 American Control Conference (ACC)*, pages 2842–2847, 2023a. doi: 10.23919/ACC55779.2023.10156379.

Beomyeol Yu and Taeyoung Lee. Equivariant reinforcement learning for quadrotor uav. In *2023 American Control Conference (ACC)*, pages 2842–2847. IEEE, 2023b.

Chao Yu, Akash Velu, Eugene Vinitsky, Jiaxuan Gao, Yu Wang, Alexandre Bayen, and Yi Wu. The surprising effectiveness of PPO in cooperative multi-agent games. In *Thirty-sixth Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2022.

Xupeng Zhu, Dian Wang, Ondrej Biza, Guanang Su, Robin Walters, and Robert Platt. Sample efficient grasp learning using equivariant models, 2022. URL https://arxiv.org/abs/2202.09468.

Martin A. Zinkevich and Tucker R. Balch. Symmetry in markov decision processes and its implications for single agent and multiagent learning. In *International Conference on Machine Learning*, 2001. URL https://api.semanticscholar.org/CorpusID:14856766.