From Traits to Roles: Consensus-Guided Composition of Orthogonal Experts for Cooperative MARL

Yewei Zhou[1,2,3], Bin Zhang[4], Ying Zhou[1], Xuri Ge[1], Dapeng Li[5,6], Hangyu Mao[7], Pengjie Ren[1,2,3], Zhiwei Xu[1*]

[1]Shandong University
[2]Shandong Provincial Key Laboratory of Computing-Network Integration
[3]Qingdao Key Laboratory of Trustworthy Artificial Intelligence
[4]The Key Laboratory of Cognition and Decision Intelligence for Complex Systems, Institute of Automation, Chinese Academy of Sciences
[5]Li Auto Inc.
[6]Tsinghua University
[7]Institute of Microelectronics, Chinese Academy of Sciences
yeweizhou@mail.sdu.edu.cn, zhiwei_xu@sdu.edu.cn

Abstract

Parameter sharing is a central design choice in cooperative multi-agent reinforcement learning, yet it fundamentally conflicts with the need for role specialization in heterogeneous cooperative environments. Existing role-based methods typically learn monolithic role representations, which often suffer from gradient interference and fail to capture the compositional structure of complex behaviors. Inspired by Trait Theory, we propose **DE**compose and **CO**nstruct **R**oles (**DECOR**), a framework that models agent roles as dynamic compositions of orthogonal behavioral traits. DECOR introduces an orthogonal Mixture-of-Experts architecture to decompose behaviors into independent traits, mitigating destructive gradient interference under parameter sharing, and a group-consensus guided mechanism to extract team-level tactical intents that guide role composition. Experiments on multiple benchmarks demonstrate that DECOR consistently improves sample efficiency and overall performance over existing related methods.

1 Introduction

Cooperative multi-agent reinforcement learning (MARL) has been successfully applied to a wide range of complex decision-making scenarios that require coordinated behaviors, such as traffic signal control [Chu *et al.*, 2019], recommendation systems [Zhao *et al.*, 2024], and multi-robot coordination [Feng *et al.*, 2024]. Most existing cooperative MARL methods adopt the centralized training with decentralized execution (CTDE) paradigm [Lowe *et al.*, 2017], in which agents are trained with access to global information while executing policies based solely on local observations. Within this paradigm, parameter sharing is commonly employed, where all agents are instantiated with the same network, in order to improve training efficiency and scalability through parameter reuse and shared experience learning. However, this design implicitly assumes agent homogeneity, an assumption that is frequently violated in real-world cooperative tasks involving heterogeneous agents with diverse roles, capabilities, and functional responsibilities. As a consequence, enforcing shared parameters on heterogeneous agents induces destructive gradient interference, where distinct functional requirements drive the shared parameters in competing directions, leading to suboptimal convergence and behavioral homogenization [Li *et al.*, 2024b].

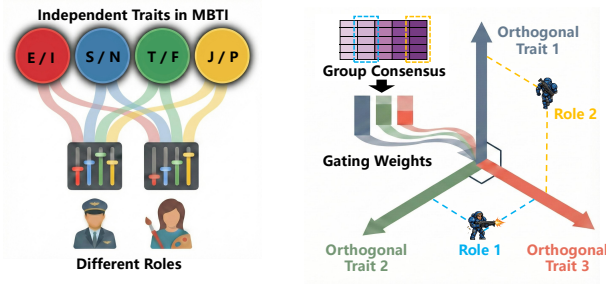

Figure 1: From Trait Theory to the proposed DECOR method. **Left:** Trait Theory views roles as compositions of independent traits. **Right:** DECOR implements this principle via orthogonal trait decomposition and group-consensus guided role composition.

In response to the behavioral homogenization induced by parameter sharing, recent studies have explored two representative directions to enhance agent specialization in cooperative MARL. One line of work focuses on modifying the parameter-sharing structure itself, for example by adopting fully independent networks or selective-sharing schemes [Christianos *et al.*, 2021]. While these approaches alleviate behavioral homogenization to some extent, they often incur reduced sample efficiency or rely on rigid, manually designed grouping heuristics, resulting in static forms of di-

versity that fail to adapt to the emergent and dynamic role requirements of complex cooperative tasks. Alternatively, a growing body of work introduces explicit role modeling to induce specialization among agents [Wang *et al.*, 2020b; Wang *et al.*, 2020c]. Despite their algorithmic differences, these role-based methods share several common design choices. In these approaches, role representations are typically learned as compact and holistic embeddings that encode multiple functional aspects within a single latent vector. Although effective in practice, this design can introduce optimization challenges, as improving one capability may affect others through the shared latent space. Moreover, role inference is often driven primarily by individual observation histories, which may not fully capture the global coordination cues necessary for team-level cooperation.

In psychology, the *Trait Theory* of personality suggests that complex human behaviors and personalities are not monolithic constructs, but rather emerge from the composition of a set of relatively independent and stable traits [Jung and Beebe, 2016]. This perspective underlies widely used personality frameworks, such as the Myers-Briggs Type Indicator (MBTI) [Briggs, 1976], where individuals are characterized through combinations of fundamental behavioral dimensions rather than a single unified type, as shown in Figure 1. A key implication of this view is that diversity in behavior arises naturally from different compositions of shared traits, enabling both commonality and specialization within a population.

Inspired by this compositional view of personality, we propose **DECOR** (**DE**compose and **CO**nstruct **R**oles), a framework that rethinks role representation in cooperative MARL from a trait-based perspective. Instead of modeling roles as static and monolithic embeddings, DECOR represents an agent's role as a dynamic composition of shared yet independent behavioral traits. To realize this idea, DECOR introduces an orthogonal role decomposition mechanism that parameterizes traits as independent expert subspaces within an orthogonal Mixture-of-Experts (MoE) architecture [Puigcerver *et al.*, 2023], explicitly decoupling different behavioral capabilities and mitigating gradient interference during parameter sharing. Based on the decomposed trait experts, a group-consensus guided module learns discrete consensus embeddings via a Vector Quantized-Variational Autoencoder (VQ-VAE) based discrete codebook [Van Den Oord *et al.*, 2017], implicitly grouping agents into coordination clusters that share the same consensus embedding. This shared consensus is then used to compute gating weights for agent-level trait composition. Through this gated composition process, agents can dynamically reconstruct their roles by combining orthogonal traits with role-specific weights, enabling both specialization and coordinated team behavior. Moreover, DECOR is designed as a plug-in framework that can be readily integrated into existing cooperative MARL algorithms, and extensive experiments across multiple domains demonstrate its consistent performance gains.

Our main contributions are summarized as follows:

- We propose the orthogonal role decomposition mechanism to parameterize traits as independent expert subspaces within an orthogonal MoE architecture, mitigating gradient interference under parameter sharing.
- A group-consensus guided role reconstruction module is proposed to compute trait composition weights, enabling agents to dynamically reconstruct their roles in response to evolving team objectives.
- Extensive experiments on multiple challenging benchmarks demonstrate that the proposed DECOR framework consistently outperforms existing baselines.

2 Preliminaries

2.1 Dec-POMDP and CTDE Methods

Cooperative MARL problems are commonly formulated as decentralized partially observable Markov decision processes (Dec-POMDPs) [Oliehoek *et al.*, 2016], defined by a tuple $\mathcal{G} = \langle S, U, P, r, \Omega, O, N, \gamma \rangle$, where N agents interact with the environment. At each timestep, the environment is in a global state $s \in S$. Each agent $i \in \{1, \dots, N\}$ receives a local observation $o_i \in \Omega$ from the observation function $O(s, i)$ and selects an action $u_i \in U$ based on its action-observation history τ_i. The joint action $\mathbf{u} = \langle u_1, \dots, u_N \rangle \in U^N$ induces a state transition according to $P(s' \mid s, \mathbf{u})$ and yields a shared reward $r(s, \mathbf{u})$. The objective is to learn a joint policy $\boldsymbol{\pi}$ that maximizes the expected discounted return $J(\boldsymbol{\pi}) = \mathbb{E}_{\boldsymbol{\pi}} \left[\sum_{t=0}^{\infty} \gamma^t r_t \right]$.

Centralized training with decentralized execution (CTDE) is a widely adopted learning paradigm for cooperative MARL [Lowe *et al.*, 2017]. During training, agents are allowed to leverage additional centralized information, such as the global state s or joint action $\mathbf{u}$, to facilitate learning and stabilize optimization. During execution, agents act in a fully decentralized manner using only local information, typically summarized by τ_i. This paradigm enables effective learning in partially observable cooperative environments while preserving decentralized deployability. Under the CTDE framework, two major classes of methods have been developed. Value-based approaches learn a centralized action-value function during training and enable decentralized decision-making through structured factorization or value decomposition [Sunehag *et al.*, 2017; Rashid *et al.*, 2020]. Policy-based approaches [Yu *et al.*, 2022], in contrast, directly parameterize decentralized policies while exploiting centralized information for training, such as centralized critics or joint value estimators.

2.2 Vector-Quantized Variational Autoencoder

The Vector-Quantized Variational Autoencoder (VQ-VAE) learns discrete latent representations, which enables capturing identifiable patterns from high-dimensional inputs [Van Den Oord *et al.*, 2017]. It consists of an encoder f_ϕ, a decoder f_ψ, and a learnable codebook $\mathcal{C} = \{\mathbf{e}_k\}_{k=1}^{K} \subset \mathbb{R}^D$, where D denotes the dimensionality of each codebook embedding. ϕ and ψ denote the learnable parameters of the encoder and decoder, respectively. Given an input $\mathbf{x}$, the encoder produces a continuous latent representation $z_e(x)$. This representation is then discretized by mapping it to the nearest codebook entry:

$$z_q(\mathbf{x}) = \mathbf{e}_k, \quad \text{where } k = \operatorname*{argmin}_j \|z_e(\mathbf{x}) - \mathbf{e}_j\|_2. \tag{1}$$

The training objective of VQ-VAE combines a reconstruction loss, a codebook loss, and a commitment loss:

$$\mathcal{L}_{\text{VQ}} = \|\mathbf{x} - \hat{\mathbf{x}}\|_2^2 + \|\text{sg}[z_e(\mathbf{x})] - \mathbf{e}_k\|_2^2 + \beta\|z_e(\mathbf{x}) - \text{sg}[\mathbf{e}_k]\|_2^2, \tag{2}$$

where $\hat{\mathbf{x}} = f_\psi(z_q(\mathbf{x}))$ denotes the reconstruction, and $\text{sg}[\cdot]$ represents the stop-gradient operator. β is a hyperparameter that balances the commitment loss, encouraging the encoder outputs to stay close to the selected codebook entries. By quantizing continuous representations into a finite codebook, VQ-VAE implicitly partitions inputs into K discrete latent groups that share common semantic prototypes.

2.3 Soft Mixture-of-Experts

A Mixture-of-Experts (MoE) layer combines multiple subnetworks (experts) $\{E_1, \ldots, E_M\}$ through a gating network G [Shazeer *et al.*, 2017]. While sparse MoE variants activate only a small subset of experts for computational efficiency, a Soft MoE [Puigcerver *et al.*, 2023] assigns continuous weights to all experts. This dense weighting allows the output to be expressed as a linear combination of multiple expert functions, providing a flexible function approximation form that supports smooth interpolation across experts. Given an input $\mathbf{x}$, the gating network produces a dense weight vector $\mathbf{w}(\mathbf{x}) = [w_1(\mathbf{x}), \ldots, w_M(\mathbf{x})] \in \mathbb{R}^M$, and the MoE output $\mathbf{y}$ is computed as

$$\mathbf{w}(\mathbf{x}) = \text{softmax}(G(\mathbf{x})), \quad \mathbf{y} = \sum_{m=1}^{M} w_m(\mathbf{x})\, E_m(\mathbf{x}). \tag{3}$$

This weighted aggregation allows the model to represent outputs as continuous combinations of multiple expert functions, while remaining fully differentiable.

3 Methodology

In this section, we present DECOR, a framework designed to address the tension between parameter sharing and role specialization in cooperative MARL. The central idea of DECOR is to move beyond static, monolithic role representations and instead construct agent roles through dynamic compositions of orthogonal behavioral traits. By decomposing role representations into orthogonal components, distinct agent capabilities are routed to separate experts, which explicitly alleviates gradient interference under shared parameters. At the same time, role reconstruction is guided by a learned group-level consensus, which provides a shared strategic context for agents within the same coordination group, thereby aligning their behaviors with the global team objective. Concretely, DECOR consists of two core components, as illustrated in Figure 2: an *Orthogonal Role Decomposition* module that learns a set of independent trait experts, and a *Group-Consensus Guided Role Reconstruction* module that dynamically composes these traits into agent roles.

3.1 Orthogonal Role Decomposition

Effective parameter sharing in cooperative MARL requires disentangling agent-specific capabilities that are otherwise coupled within monolithic role representations. When agent roles are modeled as single latent vectors, gradients corresponding to different functional requirements are prone to interfere with each other, leading to unstable optimization and limited specialization. To address this issue, we first introduce an orthogonal role decomposition mechanism that factorizes role representations into independent components, providing a structured basis for subsequent role construction.

Role Decomposition via Orthogonal MoE. To extract fundamental behavioral traits, we introduce an Orthogonal MoE. Let $\{E_m\}_{m=1}^M$ denote a set of M parallel expert networks. Given the latent embedding $\mathbf{h}_i$ for agent i, each expert produces a raw role basis vector $\mathbf{v}_{i,m} = E_m(\mathbf{h}_i) \in \mathbb{R}^d$. These vectors are stacked to form a raw basis matrix $\mathbf{V}_i = [\mathbf{v}_{i,1}, \ldots, \mathbf{v}_{i,M}] \in \mathbb{R}^{d \times M}$.

To encourage diverse and non-redundant trait representations, we aim to transform the raw basis $\mathbf{V}_i$ into an *orthogonal basis matrix* $\hat{\mathbf{B}}_i$ whose columns are mutually orthogonal. Specifically, $\hat{\mathbf{B}}_i$ is obtained by applying a deterministic yet differentiable orthogonalization transformation to $\mathbf{V}_i$, which enables gradients to be propagated back to the raw basis V_i during optimization. Conceptually, this objective can be expressed as maximizing the standard reinforcement learning objective subject to an orthogonality constraint:

$$\max_{\boldsymbol{\pi}} \mathbb{E}_{\boldsymbol{\pi}}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right] \quad \text{s.t.} \quad \hat{\mathbf{B}}_i^\top \hat{\mathbf{B}}_i = \mathbf{I}, \tag{4}$$

where $\mathbf{I}$ denotes the identity matrix. This constraint encourages the resulting $\hat{\mathbf{B}}_i$ to form an orthonormal system, thereby promoting diverse trait representations across experts.

However, directly solving this constrained optimization is not amenable to standard gradient-based training. Instead, we realize this objective constructively by employing the Gram-Schmidt (GS) [Björck, 1967] process as a deterministic and differentiable orthogonalization operator. The GS process sequentially projects each raw vector $\mathbf{v}_{i,m}$ onto the orthogonal complement of the subspace spanned by the previously constructed basis vectors. The resulting orthogonal basis vector $\hat{\mathbf{b}}_{i,m}$ is computed recursively as:

$$\mathbf{b}_{i,m} = \mathbf{v}_{i,m} - \sum_{j=1}^{m-1} \langle \mathbf{v}_{i,m}, \hat{\mathbf{b}}_{i,j} \rangle \hat{\mathbf{b}}_{i,j}, \quad \hat{\mathbf{b}}_{i,m} = \frac{\mathbf{b}_{i,m}}{\|\mathbf{b}_{i,m}\|_2 + \epsilon}, \tag{5}$$

where ϵ is a small constant introduced for numerical stability. Since the preceding basis vectors $\{\hat{\mathbf{b}}_{i,j}\}$ are already normalized, the projection reduces to a scalar inner-product coefficient. The final orthogonal basis matrix is assembled as $\hat{\mathbf{B}}_i = [\hat{\mathbf{b}}_{i,1}, \ldots, \hat{\mathbf{b}}_{i,M}]$.

Geometrically, this orthogonalization ensures that $\hat{\mathbf{B}}_i$ forms a linearly independent coordinate system in the role representation space. By enforcing $\langle \hat{\mathbf{b}}_{i,m}, \hat{\mathbf{b}}_{i,n} \rangle = 0$ for $m \neq n$, the gradient updates associated with different role components are confined to mutually orthogonal subspaces. As a result, when heterogeneous agents update specific role components, their gradient flows remain disentangled, effectively mitigating the destructive interference commonly observed under naive parameter sharing.

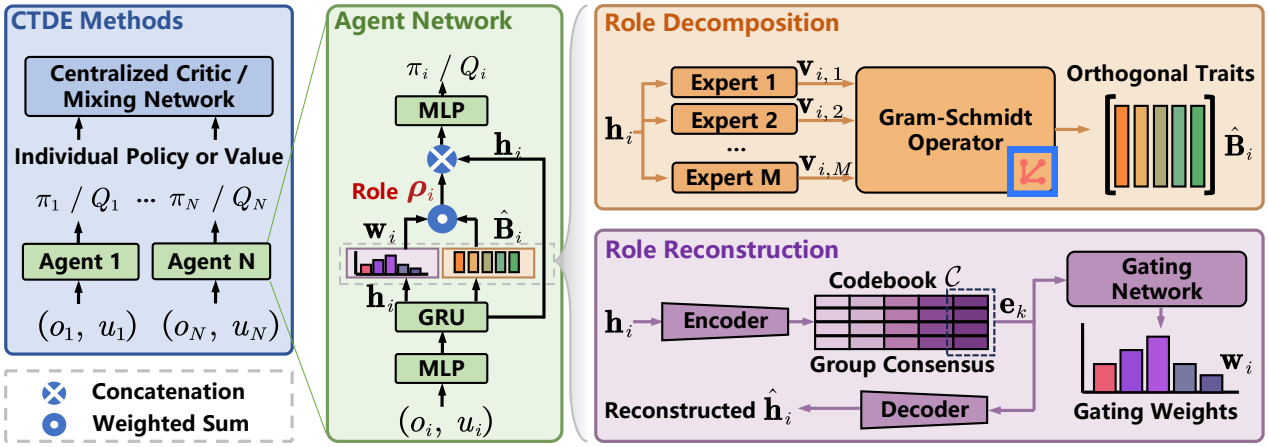

Figure 2: Overview of the proposed DECOR framework. Under the CTDE paradigm, each agent first decomposes its latent embedding $\mathbf{h}$ into a set of orthogonal behavioral traits $\hat{\mathbf{B}}$ via an orthogonal Mixture-of-Experts. A group consensus $\mathbf{e}_k$ is then extracted from agent embeddings and used to generate gating weights $\mathbf{w}$, which reconstruct agent roles $\boldsymbol{\rho}$ as weighted compositions of orthogonal traits. The reconstructed role $\boldsymbol{\rho}$ is integrated into the agent network to guide decision-making.

3.2 Group-Consensus Guided Role Reconstruction

While orthogonal role decomposition provides a set of disentangled trait representations, effective coordination further requires that these traits be composed in a manner consistent with the global team objective. To this end, we introduce a *Group Consensus Builder* that extracts shared coordination signals from decentralized agent experiences, dynamically forming coordination groups and providing semantic guidance for role reconstruction.

The group consensus builder is implemented using a VQ-VAE, whose discrete latent space serves as a shared dictionary of group-level coordination consensus, where each codebook entry represents a recurring coordination pattern observed across agents and timesteps. Compared to continuous latent representations, the discrete codebook introduces an information bottleneck that suppresses task-irrelevant variability while preserving identifiable and stable team-level structures [Van Den Oord *et al.*, 2017]. By mapping continuous and potentially noisy local representations into a shared discrete codebook, the consensus builder encourages agents with similar underlying coordination states to be associated with the same consensus, thereby implicitly aligning their local representations.

Formally, we take the latent agent embedding $\mathbf{h}_i$, which summarizes the agent's action-observation history τ_i, as the input to the consensus builder. A learnable codebook $\mathcal{C} = \{\mathbf{e}_k\}_{k=1}^{K} \subset \mathbb{R}^D$ is maintained, where each embedding $\mathbf{e}_k$ corresponds to a distinct consensus. An encoder f_ϕ maps $\mathbf{h}_i$ to a continuous latent representation $z_e(\mathbf{h}_i)$, which is then quantized by retrieving the nearest codebook entry according to Equation (1):

$$z_q(\mathbf{h}_i) = \mathbf{e}_k, \quad k = \underset{j \in \{1,\dots,K\}}{\operatorname{argmin}} \|z_e(\mathbf{h}_i) - \mathbf{e}_j\|_2. \tag{6}$$

Through this quantization process, agents exhibiting similar underlying coordination patterns are mapped to the same consensus, even when their local observations differ. The selected codebook entry $\mathbf{e}_k$ thus represents a stable group-level intent and serves as a discrete semantic anchor for subsequent role reconstruction.

To prevent degenerate solutions and ensure that the consensus captures meaningful temporal information, a decoder f_ψ reconstructs the original latent embedding as $\hat{\mathbf{h}}_i = f_\psi(z_q(\mathbf{h}_i))$. The group consensus builder is trained using the standard VQ-VAE objective as given in Equation (2):

$$\mathcal{L}_{\text{VQ}}^i = \|\mathbf{h}_i - \hat{\mathbf{h}}_i\|_2^2 + \left\| \text{sg}[z_e(\mathbf{h}_i)] - \mathbf{e}_k \right\|_2^2 + \beta \left\| z_e(\mathbf{h}_i) - \text{sg}[\mathbf{e}_k] \right\|_2^2. \tag{7}$$

The reconstruction term preserves information encoded in $\mathbf{h}_i$, while the remaining terms align the continuous latent space with the discrete consensus.

Trait Composition via Consensus-Guided Gating. Conditioned on the group consensus $\mathbf{e}_k$, DECOR reconstructs an agent-specific role by composing the orthogonal trait basis $\hat{\mathbf{B}}_i$ through a consensus-guided gating mechanism. Intuitively, while $\hat{\mathbf{B}}_i$ characterizes the agent's available capabilities, the consensus $\mathbf{e}_k$ specifies how these capabilities should be combined under the current team strategy. To implement this composition, we employ a lightweight *gating network* that maps the consensus embedding $\mathbf{e}_k$ to a continuous importance distribution over the M orthogonal traits. The resulting gating weights $\mathbf{w}_i \in \mathbb{R}^M$ are computed as:

$$\mathbf{w}_i = \text{softmax}(\text{MLP}(\mathbf{e}_k)), \tag{8}$$

where each weight $w_{i,m}$ reflects the relevance of the m-th trait under the shared group consensus. Since the gating weights depend solely on the consensus, agents associated with the same consensus are encouraged to emphasize a consistent subset of traits, promoting coordinated behavior. The final role representation is then reconstructed as a weighted combination of orthogonal traits:

$$\boldsymbol{\rho}_i = \hat{\mathbf{B}}_i \, \mathbf{w}_i \in \mathbb{R}^d. \tag{9}$$

This formulation achieves a dual effect: orthogonality in $\hat{\mathbf{B}}_i$ mitigates gradient interference across trait subspaces during training, while the consensus-guided weights $\mathbf{w}_i$ align individual role compositions with the global team objective during execution.

It is worth noting that the reconstructed role representation $\boldsymbol{\rho}_i$ in DECOR can be flexibly integrated into different MARL backbones. For value-based methods, $\boldsymbol{\rho}_i$ is concatenated with the latent agent embedding $\mathbf{h}_i$ and provided as additional context to the individual utility network [Rashid *et al.*, 2020]:

$$Q_i(\mathbf{h}_i, u_i) = \mathrm{MLP}([\mathbf{h}_i; \boldsymbol{\rho}_i]). \tag{10}$$

For policy-based methods, $\boldsymbol{\rho}_i$ can be incorporated analogously as a context-modulating input to the decentralized policy network [Yu *et al.*, 2022].

3.3 Overall Learning Objective

Having established the mechanisms for orthogonal role decomposition and tactical intent alignment, we integrate these components into an end-to-end training scheme. DECOR is designed to be agnostic to the specific cooperative MARL backbone and can be trained under a general CTDE framework. The role decomposition and reconstruction operations are applied at the individual agent level and therefore do not impose additional structural constraints on the underlying reinforcement learning objective.

The entire framework is optimized by minimizing a composite objective that balances the primary reinforcement learning objective with an auxiliary role representation learning objective:

$$\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{RL}}(\theta) + \lambda_{\text{VQ}} \sum_{i=1}^{N} \mathcal{L}_{\text{VQ}}^{i}(\phi, \psi, \mathcal{C}), \tag{11}$$

where $\mathcal{L}_{\text{RL}}$ denotes a generic cooperative reinforcement learning loss. In value-based instantiations, $\mathcal{L}_{\text{RL}}$ corresponds to a temporal-difference objective defined on a centralized or factorized action-value function [Sunehag *et al.*, 2017]. For policy-based methods, $\mathcal{L}_{\text{RL}}$ can be instantiated as a policy gradient or actor-critic objective [Sutton *et al.*, 1998]. λ_{VQ} controls the contribution of the auxiliary role learning objective defined in Equation (7), and θ represents the parameters of the agent networks as well as any centralized components used during training. A detailed algorithmic description of DECOR is provided in Appendix A.

4 Related Work

Parameter Sharing and Gradient Conflict in MARL. Parameter sharing has emerged as a dominant paradigm in cooperative MARL for enhancing sample efficiency and scalability [Gupta *et al.*, 2017; Foerster *et al.*, 2016]. However, this paradigm implicitly assumes agent homogeneity. When applied to agents with distinct role specialization, naive sharing forces a single network to simultaneously satisfy contradictory optimization objectives, inducing gradient interference and behavioral homogenization [Qin *et al.*, 2025]. To mitigate this, distinct architectural strategies have been proposed. Static grouping methods like SePS [Christianos *et al.*, 2021] rely on manual heuristics, which are inflexible in dynamic tasks. Dynamic approaches, such as CDS [Li *et al.*, 2021] and recent pruning-based techniques [Kim and Sung, 2023; Li *et al.*, 2024a], adaptively generate agent-specific subnetworks. More recently, ADMN [Yu *et al.*, 2024] introduces routing mechanisms to select modules. However, these methods primarily focus on architectural variations, without explicitly regulating the semantics of the learned representations. Consequently, functional features corresponding to different roles may remain entangled within shared latent subspaces, allowing conflicting gradients to interfere during optimization. In contrast, DECOR explicitly enforces orthogonality at the level of behavioral traits, thereby disentangling gradient flows across roles and addressing gradient interference in a principled manner.

Role Emergence and Representation. Role-based methods seek to induce structured coordination by learning latent representations that guide policies [Zeng *et al.*, 2023]. Early works such as ROMA [Wang *et al.*, 2020b] and RODE [Wang *et al.*, 2020c] construct roles based on local observations or action spaces. ACORM [Hu *et al.*, 2023] further advances this by employing contrastive learning to extract trajectory-based embeddings. Despite these advances, existing methods commonly model roles as compact and monolithic embeddings that are inferred predominantly from individual experience. Such formulations make it difficult to incorporate explicit group-level coordination signals that should constrain and align individual roles. Moreover, encoding diverse and potentially conflicting functional requirements into a single latent vector can lead to entangled representations. In contrast, our framework adopts a *decompose-and-reconstruct* perspective, where roles are composed from orthogonal traits and guided by a learned group consensus.

5 Experiments

This section evaluates the proposed DECOR framework in cooperative MARL settings. Experiments are conducted on multiple cooperative MARL benchmarks, including SMAC [Samvelyan *et al.*, 2019], SMACv2 [Ellis *et al.*, 2023], and Google Research Football [Kurach *et al.*, 2020]. These experiments assess the effectiveness of DECOR in promoting coordinated behavior, its generality across different learning paradigms, and the contribution of its core components. In addition, visualization analyses are provided to offer intuitive insights into how DECOR improves learning efficiency. Unless otherwise specified, DECOR is instantiated on top of the classical value decomposition algorithm QMIX [Rashid *et al.*, 2020].

5.1 Performance on SMAC

To enable a systematic comparison with representative cooperative MARL baselines, we first evaluate DECOR on the StarCraft Multi-Agent Challenge (SMAC). This benchmark comprises a variety of cooperative tasks with heterogeneous agents, where effective role specialization and coordination are critical for success. Baselines that explicitly address agent heterogeneity or role learning are included, namely ADMN [Yu *et al.*, 2024], SePS [Christianos *et al.*,

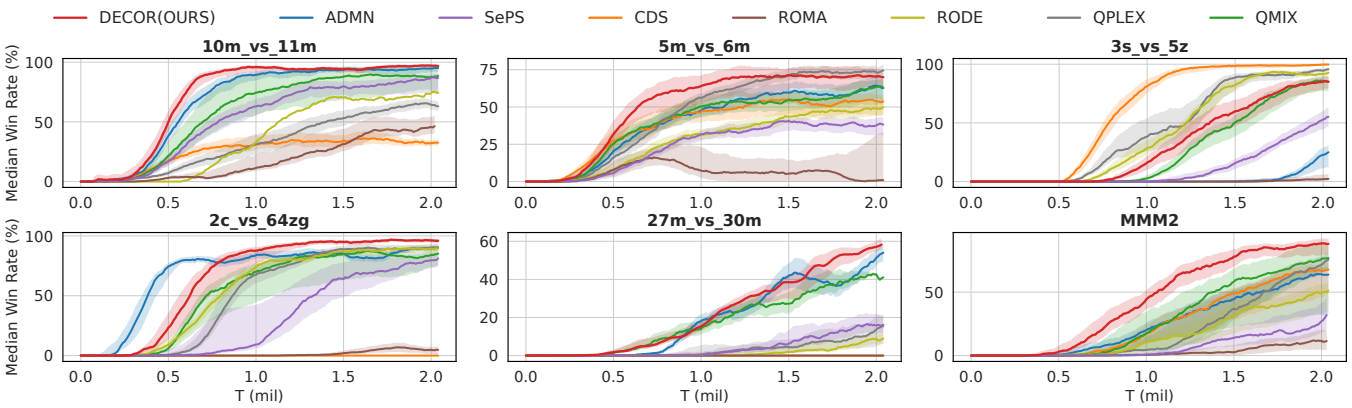

Figure 3: Performance comparison with baselines in different scenarios on SMAC.

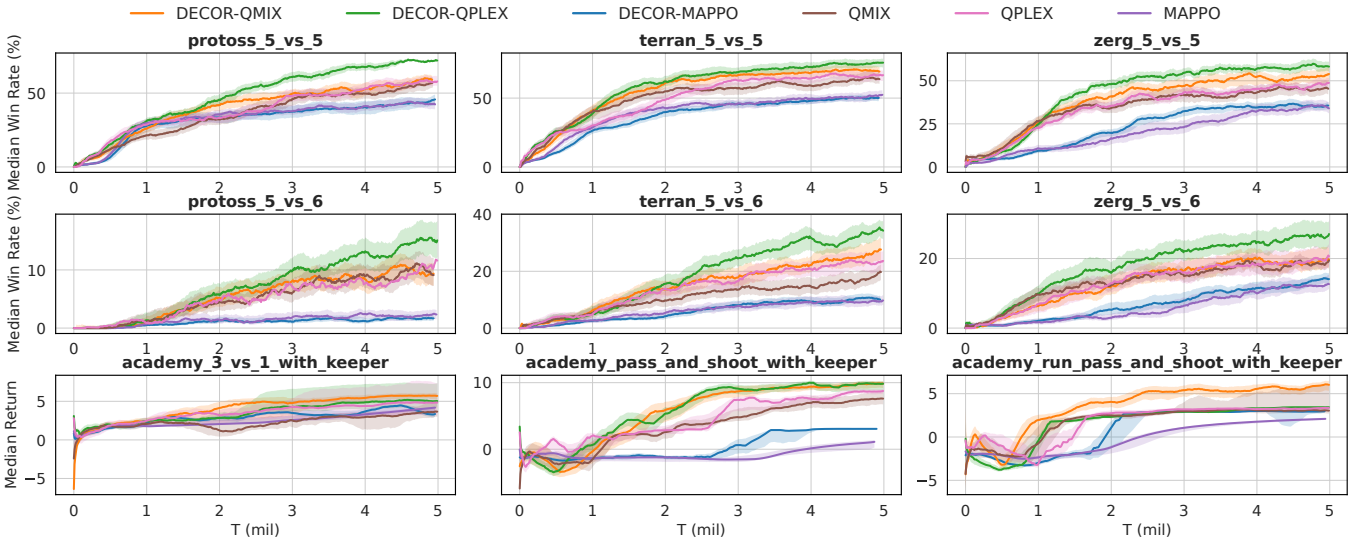

Figure 4: Performance comparison between DECOR-enhanced methods and baseline algorithms on SMACv2 and GRF.

2021], CDS [Li *et al.*, 2021], ROMA [Wang *et al.*, 2020b], and RODE [Wang *et al.*, 2020c]. In addition, QMIX and QPLEX [Wang *et al.*, 2020a] are considered as standard value decomposition methods for reference.

Figure 3 reports the median win rates across representative SMAC scenarios. Overall, DECOR demonstrates consistently superior performance and improved sample efficiency across most tasks. In highly heterogeneous scenarios such as *MMM2*, DECOR converges substantially faster than strong baselines. While ADMN adaptively routes information through different modules, it does not explicitly regulate the semantics of shared representations. In contrast, the orthogonal MoE in DECOR explicitly disentangles gradients associated with different unit types, effectively mitigating interference. In large-scale scenarios that demand complex division of labor, such as *27m_vs_30m*, DECOR maintains a clear performance advantage. This result indicates that the consensus-guided role reconstruction mechanism successfully breaks behavioral symmetry, enabling homogeneous agents to spontaneously group into specialized tactical roles, whereas competing methods tend to suffer from behavioral homogenization. A marginal performance gain is observed on *3s_vs_5z*. This behavior is primarily attributed to limitations of the underlying QMIX backbone on this specific map rather than the proposed framework itself. These empirical results validate the effectiveness of DECOR in resolving the tension between parameter sharing and role specialization, demonstrating clear advantages over existing role-based and selective-sharing approaches.

5.2 Performance on SMACv2 and GRF

To further evaluate generalization capability and robustness under environmental stochasticity, experiments are conducted on SMACv2 and Google Research Football (GRF). Compared to the static scenarios in SMAC, SMACv2 introduces substantial randomness by varying unit types and initial configurations across episodes, breaking the assumption of fixed team compositions and preventing agents from memorizing specific trajectories. GRF provides an additional challenging cooperative benchmark, characterized by dynamic multi-

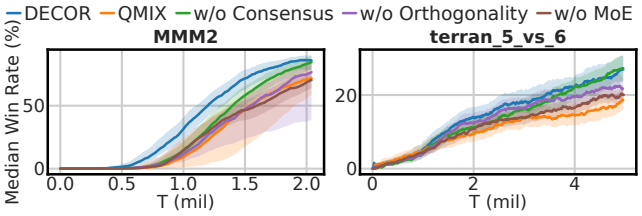

Figure 5: Ablation study comparing DECOR with ablated variants.

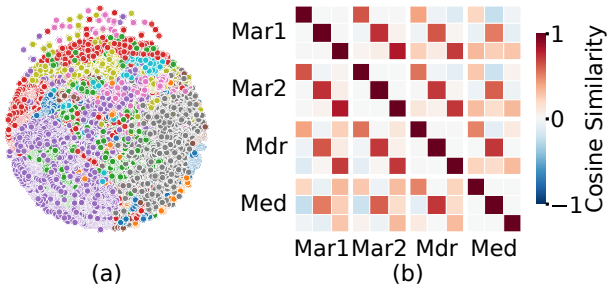

Figure 6: Visualizations. (a) Reconstructed role representations (t-SNE). (b) Orthogonal trait similarity across agents.

agent interactions and diverse tactical situations that require adaptive coordination. To assess architectural universality, DECOR is integrated into representative cooperative MARL baselines spanning different learning paradigms, including value decomposition methods (QMIX and QPLEX) and the policy-based approach MAPPO [Yu *et al.*, 2022]. Figure 4 shows the performance of the original base algorithms with their DECOR-augmented counterparts across representative scenarios. DECOR improves performance across heterogeneous backbone architectures, indicating that its benefits are not tied to a specific learning paradigm. Given the substantial stochasticity present in both SMACv2 and GRF, these results suggest that DECOR's dynamic role reconstruction enables robust generalization to diverse coordination settings.

5.3 Ablation Study

We conduct ablation studies to examine the contribution of each core component in DECOR. To isolate the effect of individual modules, the full DECOR framework is compared against the following ablated variants:

- **w/o Consensus**: removes the group consensus builder, with gating weights generated directly from the local hidden state $\mathbf{h}_i$, evaluating the necessity of global coordination signals and shared group-level consensus.
- **w/o Orthogonality**: retains the MoE architecture but removes the Gram-Schmidt orthogonalization, using standard soft gating to assess the importance of explicit gradient disentanglement.
- **w/o MoE**: replaces the orthogonal decomposition module with a standard shared MLP of comparable parameter count, evaluating whether performance gains arise merely from increased model capacity.

Ablation experiments are conducted on representative heterogeneous scenarios from both SMAC and SMACv2, including *MMM2* and *terran_5_vs_6*, respectively. Figure 5 presents the ablation results on these two scenarios. The full DECOR framework outperforms all ablated variants, demonstrating the complementary and synergistic effects of its components. Notably, w/o Orthogonality variant leads to a severe performance drop, degrading to a level comparable to the vanilla QMIX baseline. This provides empirical support for our claim: without explicit orthogonalization, different experts tend to learn redundant or entangled representations, failing to disentangle heterogeneous functional requirements. The w/o MoE variant also exhibits performance close to QMIX, indicating that simply increasing network capacity does not yield meaningful improvements. Although the w/o Consensus variant performs better than orthogonality-ablated settings, its gap with the full DECOR model highlights the importance of group-level consensus. These results confirm that the global consensus provided by the VQ-VAE codebook is essential for coherently composing orthogonal traits into effective team strategies.

5.4 Visualization

To examine whether the group consensus effectively guides role reconstruction, we visualize the learned role representations using t-SNE [Maaten and Hinton, 2008]. Each point corresponds to an agent's reconstructed role representation $\boldsymbol{\rho}_i$ and is colored according to the selected group consensus $k \in \{1, \ldots, K\}$ during reconstruction. As shown in Figure 6(a), role representations associated with the same consensus form well-separated clusters in the embedding space. This clustering pattern indicates that the proposed group consensus builder provides a coherent global coordination signal, aligning role representations across agents.

In addition, we visualize the similarity matrix of trait representations across different agents in the *MMM2* scenario in Figure 6(b). For clarity, only the first three orthogonal traits of each selected agent are shown. In the similarity matrix, rows and columns correspond to the concatenated trait vectors of four agents, where indices 0, 1, and 2 denote the corresponding orthogonal traits for each agent. The resulting matrix exhibits a block-wise periodic structure: traits with the same index across different agents consistently show high similarity, while traits with different indices remain well separated with low similarity values. This structured pattern demonstrates that trait alignment is preserved across agents, providing evidence that the proposed orthogonal role decomposition module learns a coherent and semantically consistent trait basis.

6 Conclusion

In this paper, we proposed DECOR, a trait-inspired framework for cooperative MARL that reconciles parameter sharing with role specialization. DECOR decomposes agent roles into orthogonal behavioral traits and reconstructs them through group-consensus guided gating. Experiments on SMAC, SMACv2, and GRF demonstrate that DECOR consistently improves performance across value-based and policy-based CTDE backbones. Ablation and visualization results further verify the complementary roles of orthogonal decomposition and group consensus.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (Grant No.62506210, Grant No.62622210, Grant No.62472261).

References

[Björck, 1967] Åke Björck. Solving linear least squares problems by gram-schmidt orthogonalization. *BIT Numerical Mathematics*, 7(1):1–21, 1967.

[Briggs, 1976] Katharine C Briggs. *Myers-Briggs type indicator*. Consulting Psychologists Press Palo Alto, CA, 1976.

[Christianos *et al.*, 2021] Filippos Christianos, Georgios Papoudakis, Muhammad A Rahman, and Stefano V Albrecht. Scaling multi-agent reinforcement learning with selective parameter sharing. In *International Conference on Machine Learning*, pages 1989–1998. PMLR, 2021.

[Chu *et al.*, 2019] Tianshu Chu, Jie Wang, Lara Codecà, and Zhaojian Li. Multi-agent deep reinforcement learning for large-scale traffic signal control. *IEEE transactions on intelligent transportation systems*, 21(3):1086–1095, 2019.

[Ellis *et al.*, 2023] Benjamin Ellis, Jonathan Cook, Skander Moalla, Mikayel Samvelyan, Mingfei Sun, Anuj Mahajan, Jakob Nicolaus Foerster, and Shimon Whiteson. SMACv2: An improved benchmark for cooperative multi-agent reinforcement learning. In *Thirty-seventh Conference on Neural Information Processing Systems Datasets and Benchmarks Track*, 2023.

[Feng *et al.*, 2024] Pu Feng, Junkang Liang, Size Wang, Xin Yu, Xin Ji, Yiting Chen, Kui Zhang, Rongye Shi, and Wenjun Wu. Hierarchical consensus-based multi-agent reinforcement learning for multi-robot cooperation tasks. In *2024 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, pages 642–649. IEEE, 2024.

[Foerster *et al.*, 2016] Jakob Foerster, Ioannis Alexandros Assael, Nando De Freitas, and Shimon Whiteson. Learning to communicate with deep multi-agent reinforcement learning. *Advances in neural information processing systems*, 29, 2016.

[Gupta *et al.*, 2017] Jayesh K Gupta, Maxim Egorov, and Mykel Kochenderfer. Cooperative multi-agent control using deep reinforcement learning. In *International conference on autonomous agents and multiagent systems*, pages 66–83. Springer, 2017.

[Hu *et al.*, 2023] Zican Hu, Zongzhang Zhang, Huaxiong Li, Chunlin Chen, Hongyu Ding, and Zhi Wang. Attention-guided contrastive role representations for multi-agent reinforcement learning. *arXiv preprint arXiv:2312.04819*, 2023.

[Jung and Beebe, 2016] Carl Jung and John Beebe. *Psychological types*. Routledge, 2016.

[Kim and Sung, 2023] Woojun Kim and Youngchul Sung. Parameter sharing with network pruning for scalable multi-agent deep reinforcement learning. *arXiv preprint arXiv:2303.00912*, 2023.

[Kurach *et al.*, 2020] Karol Kurach, Anton Raichuk, Piotr Stańczyk, Michał Zajkac, Olivier Bachem, Lasse Espeholt, Carlos Riquelme, Damien Vincent, Marcin Michalski, Olivier Bousquet, et al. Google research football: A novel reinforcement learning environment. In *Proceedings of the AAAI conference on artificial intelligence*, volume 34, pages 4501–4510, 2020.

[Li *et al.*, 2021] Chenghao Li, Tonghan Wang, Chengjie Wu, Qianchuan Zhao, Jun Yang, and Chongjie Zhang. Celebrating diversity in shared multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 34:3991–4002, 2021.

[Li *et al.*, 2024a] Dapeng Li, Na Lou, Bin Zhang, Zhiwei Xu, and Guoliang Fan. Adaptive parameter sharing for multi-agent reinforcement learning. In *ICASSP 2024-2024 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pages 6035–6039. IEEE, 2024.

[Li *et al.*, 2024b] Xinran Li, Ling Pan, and Jun Zhang. Kaleidoscope: Learnable masks for heterogeneous multi-agent reinforcement learning. *Advances in Neural Information Processing Systems*, 37:22081–22106, 2024.

[Lowe *et al.*, 2017] Ryan Lowe, Yi I Wu, Aviv Tamar, Jean Harb, OpenAI Pieter Abbeel, and Igor Mordatch. Multi-agent actor-critic for mixed cooperative-competitive environments. *Advances in neural information processing systems*, 30, 2017.

[Maaten and Hinton, 2008] Laurens van der Maaten and Geoffrey Hinton. Visualizing data using t-sne. *Journal of machine learning research*, 9(Nov):2579–2605, 2008.

[Oliehoek *et al.*, 2016] Frans A Oliehoek, Christopher Amato, et al. *A concise introduction to decentralized POMDPs*, volume 1. Springer, 2016.

[Puigcerver *et al.*, 2023] Joan Puigcerver, Carlos Riquelme, Basil Mustafa, and Neil Houlsby. From sparse to soft mixtures of experts. *arXiv preprint arXiv:2308.00951*, 2023.

[Qin *et al.*, 2025] Haoyuan Qin, Zhengzhu Liu, Chenxing Lin, Chennan Ma, Songzhu Mei, Siqi Shen, and Cheng Wang. Gradps: Resolving futile neurons in parameter sharing network for multi-agent reinforcement learning. In *Forty-second International Conference on Machine Learning*, 2025.

[Rashid *et al.*, 2020] 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. *Journal of Machine Learning Research*, 21(178):1–51, 2020.

[Samvelyan *et al.*, 2019] Mikayel Samvelyan, Tabish Rashid, Christian Schroeder de Witt, Gregory Farquhar, Nantas Nardelli, Tim G. J. Rudner, Chia-Man Hung, Philiph H. S. Torr, Jakob Foerster, and Shimon Whiteson. The StarCraft Multi-Agent Challenge. *CoRR*, abs/1902.04043, 2019.

[Shazeer *et al.*, 2017] Noam Shazeer, Azalia Mirhoseini, Krzysztof Maziarz, Andy Davis, Quoc Le, Geoffrey Hinton, and Jeff Dean. Outrageously large neural networks: The sparsely-gated mixture-of-experts layer. *arXiv preprint arXiv:1701.06538*, 2017.

[Sunehag *et al.*, 2017] Peter Sunehag, Guy Lever, Audrunas Gruslys, Wojciech Marian Czarnecki, Vinicius Zambaldi, Max Jaderberg, Marc Lanctot, Nicolas Sonnerat, Joel Z Leibo, Karl Tuyls, et al. Value-decomposition networks for cooperative multi-agent learning. *arXiv preprint arXiv:1706.05296*, 2017.

[Sutton *et al.*, 1998] Richard S Sutton, Andrew G Barto, et al. *Reinforcement learning: An introduction*, volume 1. MIT press Cambridge, 1998.

[Van Den Oord *et al.*, 2017] Aaron Van Den Oord, Oriol Vinyals, et al. Neural discrete representation learning. *Advances in neural information processing systems*, 30, 2017.

[Wang *et al.*, 2020a] Jianhao Wang, Zhizhou Ren, Terry Liu, Yang Yu, and Chongjie Zhang. Qplex: Duplex dueling multi-agent q-learning. *arXiv preprint arXiv:2008.01062*, 2020.

[Wang *et al.*, 2020b] Tonghan Wang, Heng Dong, Victor Lesser, and Chongjie Zhang. Roma: multi-agent reinforcement learning with emergent roles. In *Proceedings of the 37th International Conference on Machine Learning*, pages 9876–9886, 2020.

[Wang *et al.*, 2020c] Tonghan Wang, Tarun Gupta, Anuj Mahajan, Bei Peng, Shimon Whiteson, and Chongjie Zhang. Rode: Learning roles to decompose multi-agent tasks. *arXiv preprint arXiv:2010.01523*, 2020.

[Yu *et al.*, 2022] 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. *Advances in neural information processing systems*, 35:24611–24624, 2022.

[Yu *et al.*, 2024] Yang Yu, Qiyue Yin, Junge Zhang, Pei Xu, and Kaiqi Huang. Admn: Agent-driven modular network for dynamic parameter sharing in cooperative multi-agent reinforcement learning. In *Proceedings of the Thirty-Third International Joint Conference on Artificial Intelligence, IJCAI-24*, pages 302–310, 2024.

[Zeng *et al.*, 2023] Xianghua Zeng, Hao Peng, and Angsheng Li. Effective and stable role-based multi-agent collaboration by structural information principles. In *Proceedings of the AAAI conference on artificial intelligence*, volume 37, pages 11772–11780, 2023.

[Zhao *et al.*, 2024] Yang Zhao, Chang Zhou, Jin Cao, Yi Zhao, Shaobo Liu, Chiyu Cheng, and Xingchen Li. Multi-scenario combination based on multi-agent reinforcement learning to optimize the advertising recommendation system. In *2024 5th International Conference on Artificial Intelligence and Electromechanical Automation (AIEA)*, pages 190–194. IEEE, 2024.