Learning Robust Policy for Multi-UAV Collision Avoidance via Compact Causal Feature

Zhun Fan
Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
Chengdu, China
Shenzhen Loop Area Institute
fanzhun@uestc.edu.cn

Gaofei Han
Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
Chengdu, China
202512281049@std.uestc.edu.cn

Che Lin
Shantou University
Shantou, China
19clin1@stu.edu.cn

Wenji Li
Shantou University
Shantou, China
liwj@stu.edu.cn

Jie Xu✉
Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ)
Shenzhen, China
xujie1@gml.ac.cn

Jiafan Zhuang✉
Shenzhen Institute for Advanced Study, University of Electronic Science and Technology of China
Chengdu, China
jfzhuang@uestc.edu.cn

ABSTRACT

Deep reinforcement learning (DRL)-based multi-UAV collision avoidance methods often exhibit limited generalization when deployed in unseen environments, primarily due to the reliance on non-causal and redundant visual features. Such overfitting to spurious correlations compromises both robustness and safety during real-world deployment. To address these limitations, this study proposes a novel Compact Causal Feature Learning (CCFL) framework that enables UAVs to learn compact and generalizable causal representations. Specifically, a Causal Feature Identification module is designed to disentangle input representations into causal and non-causal components, ensuring that the learned features preserve true environmental causality. Furthermore, a Redundancy Feature Compression module is introduced to remove redundant dependencies and compact the causal subspace, thereby enhancing generalization to previously unseen scenarios. Extensive experiments on a challenging UAV collision avoidance benchmark demonstrate that CCFL achieves substantial performance gains over state-of-the-art baselines, increasing individual success rates by 42.0% and swarm success rates by 61.6%. These results validate the effectiveness of compact causal feature learning for improving the adaptability, robustness, and safety of autonomous UAV systems operating in complex dynamic environments.

KEYWORDS

Multi-UAV Systems, Collision Avoidance, Deep Reinforcement Learning, Compact Causal Feature Learning

Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), C. Amato, L. Dennis, V. Mascardi, J. Thangarajah (eds.), May 25 – 29, 2026, Paphos, Cyprus. © 2026 International Foundation for Autonomous Agents and Multiagent Systems (www.ifaamas.org). https://doi.org/10.65109/AAFZ2582

TABLE 1: Experimental analysis of the generalization problem in SAC+RAE, evaluated using SSR (Swarm Success Rate) and ISR (Individual Success Rate).

Training	Testing	Finetuned	SSR (%)	ISR (%)
Playground	Playground	/	50.5	91.9
Playground	Forest	/	0.0	51.9
Playground	Forest	Visual Network	43.2	88.6
Playground	Forest	Policy Network	0.0	48.2

ACM Reference Format:
Zhun Fan, Gaofei Han, Che Lin, Wenji Li, Jie Xu✉, and Jiafan Zhuang✉. 2026. Learning Robust Policy for Multi-UAV Collision Avoidance via Compact Causal Feature. In *Proc. of the 25th International Conference on Autonomous Agents and Multiagent Systems (AAMAS 2026), Paphos, Cyprus, May 25 – 29, 2026*, IFAAMAS, 9 pages. https://doi.org/10.65109/AAFZ2582

1 INTRODUCTION

Collision avoidance [10, 21] for unmanned aerial vehicles (UAVs) has emerged as a central research topic in robotics and artificial intelligence, with broad applications in areas such as precision agriculture [17], search-and-rescue operations [29], mining [28], and infrastructure inspection [19]. In these domains, the ability of multi-UAV systems to perform robust and reliable collision avoidance is critical to ensure that each UAV can navigate safely from its origin to the intended destination. However, real-world environments [14] are inherently complex, dynamic, and uncertain. Even minor perception or decision errors can trigger cascading failures, leading to mission failure, property damage, or even threats to human safety. These factors make collision avoidance not only a fundamental technical challenge but also a crucial safety concern in the deployment of autonomous UAVs.

Over the past two decades, extensive research efforts have been devoted to developing effective collision avoidance algorithms. Traditional approaches [13, 23, 30] are primarily based on manually

designed rules or modular pipelines integrating prediction, planning, and control. Although these methods can achieve reasonable performance in structured settings, they often result in overly complex systems with limited scalability and parameter sensitivity issues. In contrast, deep reinforcement learning [1] provides an end-to-end framework for learning control and decision-making policies directly from high-dimensional sensory inputs. By leveraging large-scale environmental interactions, DRL enables UAVs to autonomously develop adaptive behaviors that generalize across varying obstacle-rich scenarios. Consequently, advancing DRL-based navigation and collision avoidance methods has become a key direction for enhancing the robustness, adaptability, and safety of multi-UAV systems in complex and unseen environments.

Despite their success, DRL methods remain inherently data-driven and typically rely on the assumption that the training and testing environments share a similar distribution [12]. In practice, this assumption seldom holds. Policies trained in simulation or limited environments are frequently deployed in real-world settings characterized by distinct visual appearances, dynamics, and obstacle configurations. Such distribution shifts can lead to substantial performance degradation [7, 10, 21, 34], severely constraining the reliability, robustness, and real-world applicability of DRL-based UAV systems.

To further examine the generalization capability of DRL, we revisit SAC+RAE [16], a representative method for multi-UAV collision avoidance. As illustrated in TABLE 1, the model is first trained in a controlled environment (*i.e.*, a playground) and subsequently deployed in an unseen scenario (*i.e.*, a dense forest). Experimental results demonstrate a pronounced decline in performance when the model encounters unseen environments, verifying the significant impact of domain shift. To identify the root cause of this degradation, we conducted comparative experiments focusing on two key components: the visual network for feature extractor and the policy network. Specifically, each module was fine-tuned in the target environment while keeping the other fixed, thereby ensuring exposure to identically distributed data and isolating the effect of representation versus policy adaptation. The results show that fine-tuning the visual feature extractor substantially improves performance, whereas fine-tuning the policy network yields negligible gains. These findings indicate that the limited domain generalization capability of current DRL-based methods primarily stems from the weak generalization of visual representations.

Learning generalizable features is essential for improving the generalization of UAV collision avoidance in unseen environments. Existing methods can be divided into causal feature learning and compact feature learning. Causal approaches [3, 26, 33] enhance generalization by uncovering underlying causal relationships but suffer from high computational cost and complexity, while compact approaches [11, 15, 25] reduce redundancy through low-dimensional encoding but often lose critical causal information. Integrating these two paradigms offers a promising direction capturing causally relevant factors while maintaining compactness to improve adaptability and robustness across domains.

From a theoretical perspective, compact causal features are determined by stable physical laws that remain invariant across domains and are highly relevant to UAV collision avoidance, while excluding irrelevant or redundant information. Such features inherently

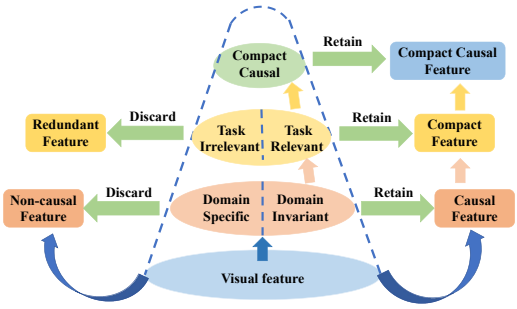

Figure 1: Hierarchical relationship among feature types. Non-causal features lie in the outer region, while causal features are divided into redundant and compact causal features. Compact causal features capture domain-invariant and task-relevant information, which are essential for robust and generalizable policy learning.

exhibit both domain invariance and task relevance. As illustrated in Figure 1, SAC+RAE employs a regularized autoencoder to encode depth images into compact visual representations. However, under purely reconstruction-based supervision [16], the encoder tends to capture all visual details without distinguishing between causal and non-causal factors (*e.g.*, obstacle distance, shape, and background texture). Among these, obstacle distance and UAV velocity constitute causal features that remain consistent across environments, whereas background textures and obstacle shapes are non-causal and environment-specific. Passing all features indiscriminately to the policy network introduces spurious correlations that impair generalization. Moreover, even causal features contain redundant components unrelated to decision-making further reducing efficiency and robustness (*e.g.*, UAV Texture and Crashed UAV). Consequently, when exposed to unseen environments containing novel obstacles or textures, such non-causal and redundant representations degrade policy performance and lead to frequent failures. This limitation underscores a critical challenge in enabling DRL-based UAV systems to achieve robust generalization under domain shift.

Therefore, effectively learning compact causal representations is crucial for enhancing the generalization ability of DRL-based UAV systems. In this work, we propose a novel framework, Compact Causal Feature Learning (CCFL), which captures causal features from data while eliminating redundancy, producing representations that are both generalizable and efficient. CCFL comprises two key modules: the Causal Feature Identification (CFI) module and the Redundancy Feature Compression (RFC) module. The CFI module employs causal representation learning to separate domain-invariant causal features from domain-specific non-causal ones, ensuring that downstream models rely solely on stable causal signals. The RFC module further compresses the causal representations, retaining task-relevant information while removing redundant components. Together, CFI guarantees the causality of the learned features, and RFC enhances their compactness, resulting in representations that are both concise and robust under domain shift.

To validate the effectiveness of CCFL, we develop a high-fidelity UAV collision avoidance benchmark that features unseen backgrounds and obstacles and supports causal feature learning and intervention. The environment allows flexible manipulation of scene elements while keeping others fixed, providing controlled conditions for identifying and validating causal relationships and systematically evaluating algorithmic robustness under diverse interventions. Experimental results demonstrate that CCFL significantly outperforms existing methods, improving individual and swarm success rates by 42.0% and 61.6%, respectively. These results confirm that CCFL effectively enhances the adaptability, robustness, and generalization of multi-UAV systems in complex and previously unseen environments.

Our main contributions are summarized as follows:

- We analyze the limitations of existing deep reinforcement learning methods in adapting to unseen environments and demonstrate the necessity of learning compact causal representations to enhance generalization.
- We propose a novel framework, Compact Causal Feature Learning, which extracts causal features from data while eliminating redundancy, resulting in representations that are both generalizable and efficient.
- We build a high-fidelity UAV simulation environment supporting causal feature learning and controlled interventions, enabling systematic evaluation under domain shifts. Experiments within this environment show that CCFL consistently outperforms state-of-the-art methods, significantly improving robustness and generalization in unseen scenarios.

2 RELATED WORK

2.1 Causal Representation Learning

Causal representation learning [4, 20, 26] aims to extract core features that reflect causal relationships from raw data and transform them into structured representations suitable for causal modeling, thereby enabling the construction of models with stronger robustness and generalization capability. SchÃűlkopf *et al.* [26] first systematically introduce the paradigm of causal representation learning, defining it as the task of discovering latent causal variables and reconstructing causal relationships from high-dimensional observational data. They emphasize that causal representation learning enhances model robustness and generalization under distribution shifts, task transfer, and counterfactual reasoning scenarios. Zhang *et al.* [33] propose ICRNet, a causal invariant representation method for handling style variations in remote sensing images. By introducing a style intervention mechanism, it disentangles stable content factors from variable style factors, thereby enhancing model robustness and generalization under non-stationary conditions. Brehmer *et al.* [3] propose a weakly supervised causal representation framework to handle unobserved causal variables in images. It employs a variational autoencoder to model latent causal structures from unpaired samples, achieving strong identifiability and interpretability in image and robotic tasks.

2.2 Compact Representation Learning

Compact Representation Learning *et al.* [32, 35] aims to reduce redundancy and complexity in the feature space while preserving essential information, thereby improving the model's generalization, efficiency, and robustness. Hinton *et al.* [11] propose to achieve information compression through feature dimensionality reduction and sparse coding techniques; However, these methods often overlook higher-order dependencies among features. Rakelly *et al.* [25] theoretically analyze the representation sufficiency of various mutual information objectives in reinforcement learning. They demonstrate that the forward mutual information objective ensures the learning of sufficient representations for control, providing a solid theoretical basis for compact representation learning. Huang *et al.* [15] propose a deductive reinforcement learning framework based on a semantic encoder, which achieves compact representation of visual inputs by extracting low-dimensional and robust features. However, this method relies on semantic segmentation priors and does not model feature redundancy or causal relationships, thus limiting its generalization ability in complex environments.

In this paper, we propose a method that integrates causal and compact representation learning to capture causal feature in data while reducing feature redundancy, thereby improving the model's generalization capability.

3 APPROACH

3.1 Problem Formulation and DRL Setting

The objective of our framework is to generate control actions based on the information available to the UAVs and to execute them continuously, ensuring that it can avoid collisions throughout the entire process. Therefore, designing appropriate observations, actions, and reward functions is crucial.

The observation, serving as the input to the policy, is defined as $o = [o_z, o_g, o_v]$. Specifically, o_z denotes the depth observation obtained from onboard sensors, which provides distance information to surrounding obstacles. o_g represents the goal observation, capturing the UAV's relative position and orientation with respect to the target, and o_v corresponds to the velocity observation, describing the UAV's current linear and angular velocities.

At each time step, the UAV receives a three-dimensional velocity control vector $a = [v_x^{cmd}, v_z^{cmd}, v_y^{cmd}]$, which is used for navigation and collision avoidance. Specifically, v_x^{cmd} denotes the commanded forward velocity, v_z^{cmd} represents the commanded climbing velocity, and v_y^{cmd} corresponds to the commanded steering velocity.

The reinforcement learning reward function is formulated as a weighted combination of multiple components, each designed to guide the agent toward efficient and robust collision avoidance.

$$r^t = r^t_{goal} + r^t_{avoid} \tag{1}$$

The collision avoidance reward r^t_{avoid} penalizes the UAVs for proximity to obstacles, encouraging it to maintain safe distances and thereby reducing the risk of collisions.

$$r^t_{avoid} = \begin{cases} r_{collision} & if\ \ collision\ \ occurs \\ \omega_{avoid} \cdot max(d_{safe} - d^t_{min}, 0) & otherwise \end{cases} \tag{2}$$

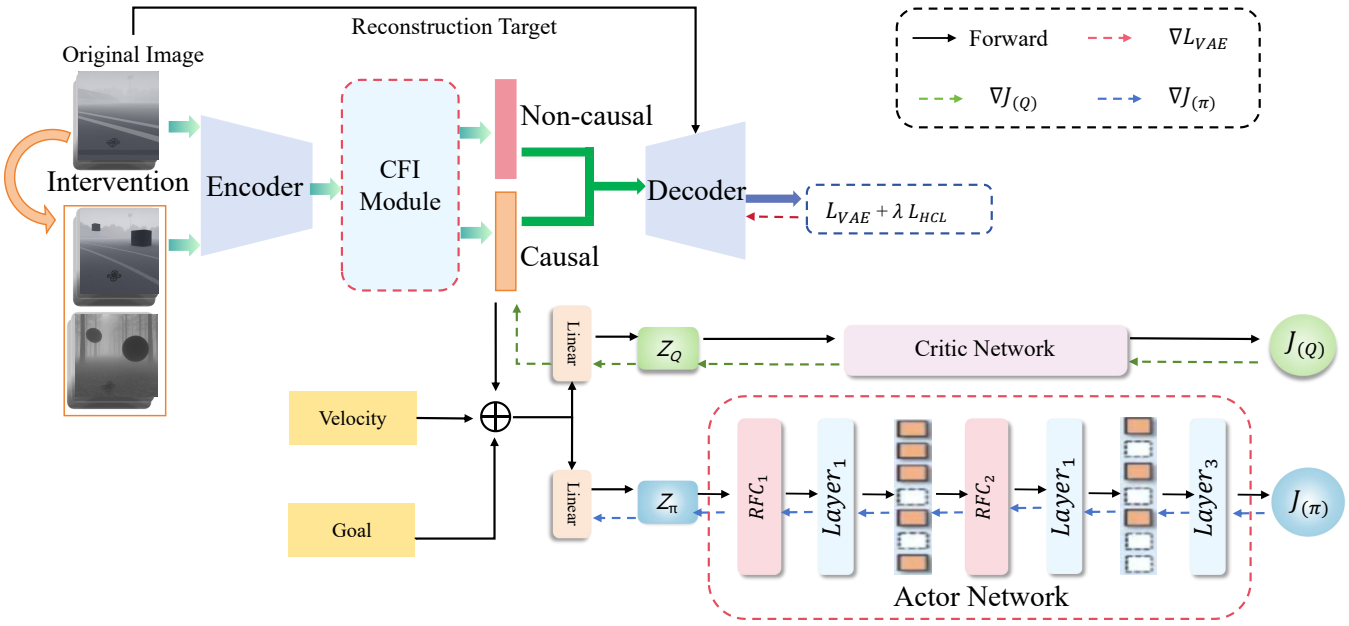

Figure 2: The architecture of our framework for multi-UAV collision avoidance.. During the representation learning stage, the CFI module disentangles visual features into causal and non-causal features. In the policy learning stage, the RFC module is integrated into the policy network to further remove redundant information, yielding compact and informative causal features.

In contrast, the navigation reward r^t_{goal} promotes progress toward the designated goal by rewarding velocities that align with the direction of the target and reduce the distance between the UAV and its destination.

$$r^t_{goal} = \begin{cases} r_{arrival} & if\ \|p^t_i - g_i\| < 0.5 \\ \omega_{goal} \cdot (\|p^{t-1}_i - g_i\| - \|p^t_i - g_i\|) & otherwise \end{cases} \tag{3}$$

By combining these two components, the UAV avoids hazardous regions while efficiently progressing toward its objective, balancing safety and task completion.

3.2 Overview

As illustrated in Figure 2, we introduce a CCFL framework to enhance generalization in UAVs collision avoidance under unseen scenarios. Based on prior work [16], our approach leverages a Variational Autoencoder (VAE) [22] to compress high-dimensional depth images into representations, while a Soft Actor-Critic (SAC) [9] agent learns end-to-end collision avoidance policies through reinforcement learning.

In our framework, the loss function integrates both representation learning from VAE and policy optimization from SAC. The VAE consists of an encoder-decoder trained with a reconstruction objective and a regularization term. The loss is defined as:

$$\mathcal{L}_{\text{VAE}} = \mathbb{E}_{q_\phi(z|x)}\left[\log p_\theta(x|z)\right] - D_{\text{KL}}\left[q_\phi(z|x) \,\|\, p(z)\right] \tag{4}$$

where the first term encourages accurate reconstruction of the input, and the KL divergence regularizes the latent distribution toward a prior $p(z)$, preventing overfitting and ensuring compact feature representations. Following SAC, the actor network is updated through using the loss function $J(\pi)$, which can be expressed as follows:

$$J_{(\pi)} = \mathbb{E}_{o\sim\mathcal{B}}\left[D_{KL}(\pi(\cdot|o)\|\boldsymbol{Q}(o,\cdot))\right] \tag{5}$$

where $\boldsymbol{Q}(o,\cdot) \propto \exp\{\frac{1}{\alpha}Q(o,\cdot)\}$. The parameters of critic network is updated through loss function $J(Q)$, which can be expressed as

$$J_{(Q)} = \mathbb{E}_{(o,a,r,o')\sim\mathcal{B}}\left[(Q(o,a) - r - \gamma\bar{V}(o'))^2\right] \tag{6}$$

The visual representations extracted by the vision module typically encode all information of scene, including causal, non-causal and redundant components. The spurious associations [6, 31, 34] between the visual features and predicted actions may arise, thereby degrading generalization. To address this issue, our CCFL framework employs a CFI module to disentangle input features into causal and non-causal features, and further integrates a RFC module to obtain compact causal features, thereby improving the model's generalization.

3.3 Causal Feature Identification Module

To address the degradation of generalization caused by non-causal information in visual representations, we propose a Causal Feature Identification module that leverages a contrastive learning mechanism to effectively disentangle causal and non-causal features. As shown in Figure 3, CFI first employs an encoder to extract latent representations from input images and then constructs positive and negative sample pairs within a contrastive framework. Specifically, the positive pairs are formed from images that share the same underlying causal semantics such as sphere obstacles in playground environments and cube obstacles in forest environments. Although these images differ in obstacle shape or background, such variations do not alter the essential collision avoidance decision. Therefore,

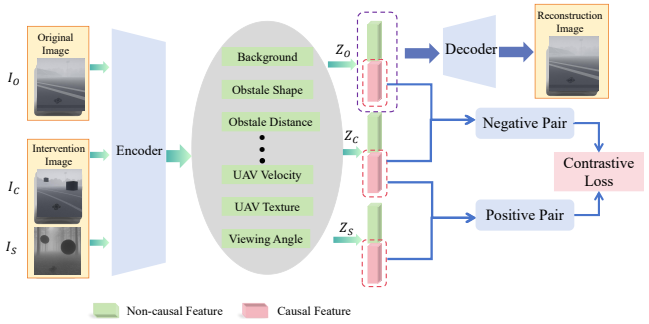

Figure 3: Causal Feature Identification Module. The encoder disentangles causal and non-causal features from original and intervention images, while the decoder reconstructs the input. Contrastive learning aligns causal representations across domains to ensure invariance and robust feature separation.

their features should remain similar to ensure that the downstream policy network outputs consistent actions, thereby decoupling non-causal features. In contrast, negative pairs are constructed from obstacle-free playground images and playground images containing sphere obstacles, as the presence or absence of obstacles directly affects the UAV's decision-making process.

The extracted features from images are decomposed into domain-invariant and domain-specific components. During training, the Hard Contrastive Learning (HCL) [18] loss $\mathcal{L}_{HCL}$ is applied to the domain-invariant features to enforce consistency across causal variations, while the reconstruction loss $\mathcal{L}_{VAE}$ is jointly imposed on both feature components to preserve the fidelity of the input image. Formally, the overall objective is expressed as:

$$\mathcal{L} = \mathcal{L}_{HCL} + \lambda \mathcal{L}_{VAE}, \tag{7}$$

where λ is a balancing hyperparameter.

The HCL loss is defined as:

$$\mathcal{L}_{HCL} = \mathbb{E}\left[-\log \frac{\exp\left(\text{sim}(z_c, z_s)/\tau\right)}{\exp\left(\text{sim}(z_c, z_s)/\tau\right) + \exp\left(\text{sim}(z_o, z_c)/\tau\right)}\right] \tag{8}$$

where $\text{sim}(\cdot)$ denotes the similarity function (*e.g.*, cosine similarity), τ is a temperature parameter, z_o denotes the causal feature from the obstacle-free playground image, z_c denotes the causal feature from the forest image with cube obstacles, and z_s denotes the causal feature from the playground image with spherical obstacles.

3.4 Redundancy Feature Compression Module

Although causal representations remove domain-invariant factors, they may still contain task-irrelevant redundancies that limit generalization. Inspired by prior work [34], we propose a RFC module that adaptively filters redundant features to obtain more compact causal representations. In contrast to previous work, the RFC module focuses on filtering redundant information within causal features, eliminating the need for hierarchical consistency constraints. As shown in Figure 4, the module generates a differentiable binary mask to selectively activate or suppress input channels, thereby preserving task-relevant information while suppressing redundant features.

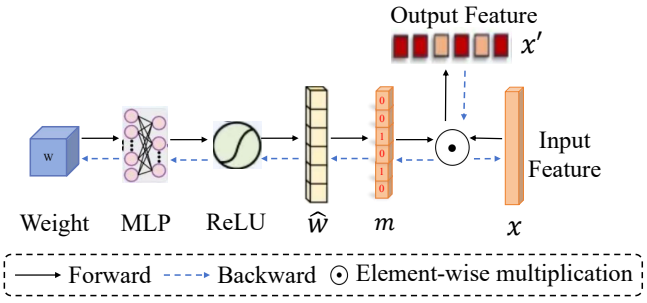

Figure 4: Redundancy Feature Compression Module. This module transforms a trainable weight into a binary mask for feature selection.

To achieve redundancy feature compression, the input feature $x \in \mathbb{R}^C$ is element-wise multiplied by a learnable binary mask m, producing a refined feature:

$$x' = x \odot m, \tag{9}$$

where C denotes the number of feature channels. This operation selectively preserves channels containing task-relevant information while suppressing task-irrelevant ones, thereby reducing interference from redundant factors.

The core of the RFC module is to construct a differentiable mask generation mechanism that can be trained jointly with the policy network. Specifically, for an intermediate vector $x \in \mathbb{R}^C$, we associate a set of trainable weights $w \in \mathbb{R}^C$, which are first transformed by a lightweight MLP followed by a ReLU activation to yield:

$$\hat{w} = ReLU(MLP(w)), \tag{10}$$

The final mask is computed as:

$$m = \frac{\hat{w}^2}{\hat{w}^2 + \epsilon}, \tag{11}$$

where ϵ is a small constant ensuring numerical stability. In this design, channels with near-zero $\hat{w}$ values are suppressed, while others remain active. This continuous approximation enables end-to-end optimization without manually defining thresholds, making the mask both trainable and differentiable.

After integrating the RFC module into the actor network, it is essential to ensure that it effectively removes redundant channels while preserving task-relevant ones. To this end, the actor loss $J_{(\pi)}$ is used as a supervisory signal, guiding the RFC module to retain feature channels critical for obstacle avoidance. In addition, an L_1 regularization term is imposed on the mask m to promote sparsity. During binary mask generation, the trainable weights of the RFC module are optimized via $J_{(\pi)}$. If the RFC module mistakenly suppresses task-relevant channels, the actor loss $J_{(\pi)}$ will significantly increase, thereby driving the module to retain task-relevant channels while eliminating redundant ones.

4 EXPERIMENT AND RESULTS

4.1 Simulation Environment for Causal Intervention

To investigate causal feature learning in UAV control tasks, we develop a high-fidelity simulation environment based on Unreal

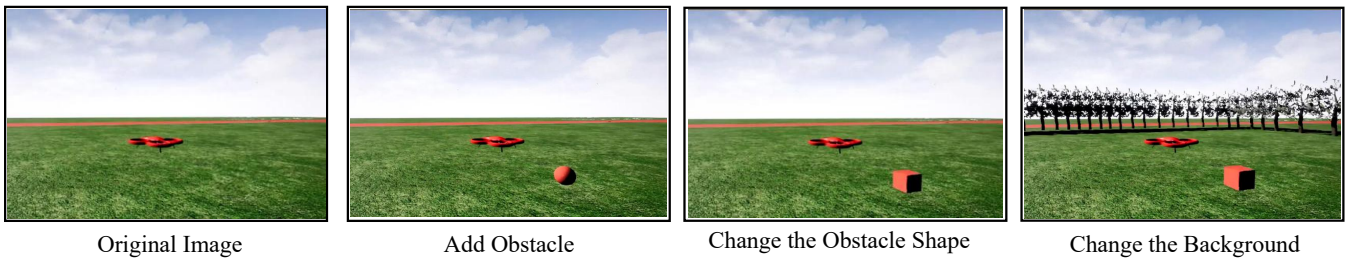

Figure 5: Illustration of the causal intervention mechanism in the simulation. The first image shows the original scene. The second image adds an obstacle, the third changes the obstacle shape, and the fourth modifies the background. All other factors remain identical across the four images.

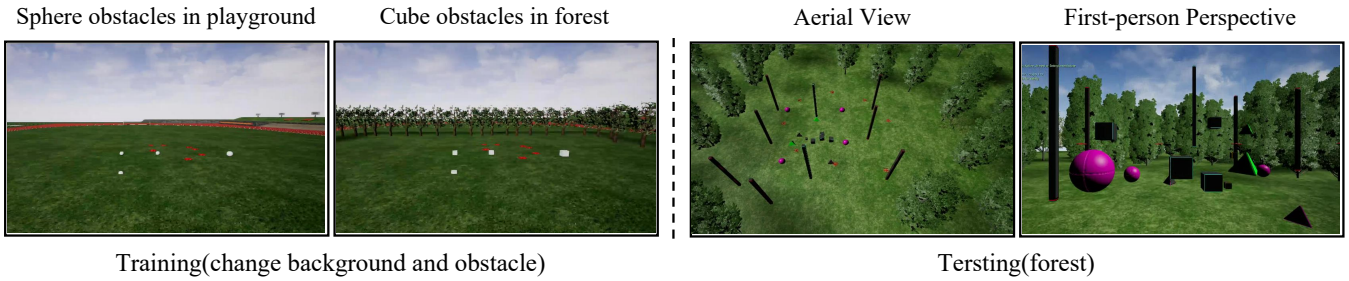

Figure 6: Simulation scenarios for model training and testing. Specifically, the training environment enables causal intervention on obstacle shapes and backgrounds, while the testing environment is a complex forest scene with diverse obstacles.

Engine [2] and AirSim [27]. It enables fine-grained manipulation of scene elements, allowing causal interventions to learn and evaluate generalizable representations. It simulates realistic flight scenarios, including playground, forest, canyon, and snow mountain.

A key feature of the simulation is its support for scene-level causal interventions. Specifically, the simulation environment allows modification of a single element such as obstacle shape or environmental background, while all others remain fixed. This capability enables researchers to directly analyze the causal effects of specific variables on UAV perception and policy learning. For instance, by maintaining identical UAV's poses and physical states while altering only the obstacle shape and background, one can evaluate the model's causal sensitivity and representational robustness to environmental variations.

As shown in Figure 5, by keeping all other factors completely unchanged while sequentially adding an obstacle, changing its shape, and modifying the background, we can systematically analyze the model's causal responses and generalization capability with respect to obstacle shape and background. The environment interfaces seamlessly with reinforcement learning frameworks through AirSim. UAVs can autonomously navigate through the environment, avoiding obstacles and reaching designated targets. All multimodal data, including RGB, depth, and semantic segmentation images, are logged at each time step.

4.2 Experiment Metrics and Setup

4.2.1 Experiment Metrics. Building upon previous work [7, 16, 34], several quantitative metrics are employed to evaluate the performance of multi-UAV collision avoidance in our experiments:

- Swarm Success Rate (SSR): The proportion of simulation episodes in which all UAVs successfully reach their designated goals without collision, reflecting the overall coordination and safety of the swarm.
- Individual Success Rate (ISR): The ratio of UAVs that successfully complete the navigation task, indicating the robustness of each agent's decision-making under cooperative conditions.
- Success weighted by Path Length (SPL): A success rate weighted by individual path lengths, which accounts for both task completion and path efficiency, providing a more balanced evaluation across diverse trajectories.

$$SPL = \frac{1}{N}\sum_{i=1}^{N} S_i \frac{l_i}{max(p_i, l_i)} \tag{12}$$

  where l_i denotes the shortest possible distance. p_i is the actual flight distance covered, and S_i is a binary success flag (1 if the UAV reaches the goal, otherwise 0).
- Extra Distance: The average relative increase in traveled distance compared to the shortest feasible path, measuring path optimality and navigation efficiency.
- Average Speed: The mean velocity of UAVs during successful episodes, assessing the trade-off between navigation efficiency and safety.

4.2.2 Experiment Setup. To verify the effectiveness of CCFL in improving the generalization capability, as illustrated in Fig 6, we designed a series of cross-scenario experiments by altering background and introducing unseen obstacles.

TABLE 2: Performance comparison with existing methods.

Method	SSR (%)	ISR (%)	SPL (%)	Extra Distance (m)	Average Speed (m/s)
SAC+RAE	0.0	51.9	44.0	6.427/5.652	0.781/0.182
+ AutoAugment [11]	4.7	57.6	53.3	1.970/1.120	0.828/0.095
+ $L1$ Norm [5, 24]	4.0	56.5	52.3	**1.965/1.162**	0.828/0.098
+ MaDi [8]	0.0	55.1	50.3	2.323/0.987	**1.017/0.103**
+ CRD [7]	1.6	69.6	62.6	2.888/2.677	0.835/0.132
+ CFS [34]	0.0	61.3	42.9	13.822/2.761	0.778/0.121
+ Our CCFL	**61.6**	**93.9**	**79.9**	4.258/1.165	0.972/0.080

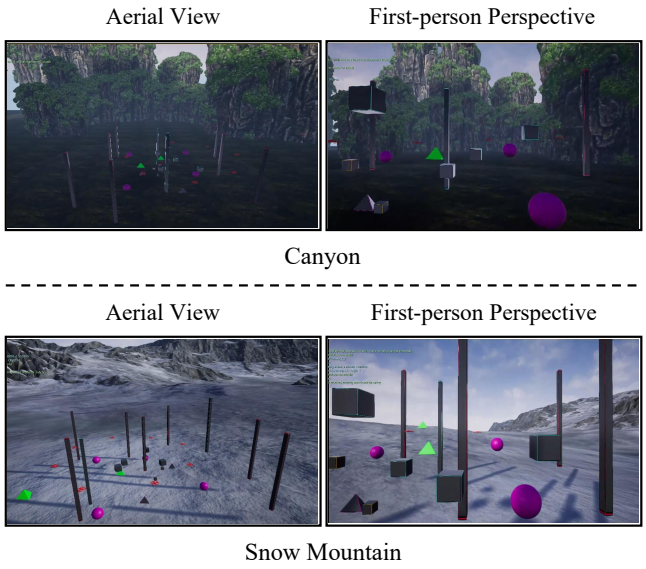

Figure 7: More Evaluation Scenes. We design two additional typical scenes (*i.e.*, Canyon and Snow mountain) for evaluation. Best viewed in zoom and color.

During training, we construct a controllable simulation environment with causal intervention capability (Section 4.1). It allows flexible switching of obstacle shapes and backgrounds, where the obstacle alternates between sphere and cube. The background can also dynamically switch between playground and forest. In addition, the initial and target positions of UAVs are randomly sampled in each episode, and flight constrained to a $16 \times 16 \times 4$ 3D space to promote exploration and improve policy stability and adaptability.

During testing, we modify scenarios along three dimensions to evaluate generalization: (1) the background is changed from a playground to a forest, (2) 22 unseen obstacle types are introduced with four randomly placed per episode to create an out-of-distribution environment, and (3) UAV initialization is changed from random placement to a circular formation with targets on the opposite side, forming a more challenging setting.

4.3 Performance Comparison

The proposed method will be compared to the baseline and existing approaches that address the generalization issue in DRL models. Specifically, our approach is compared with five representative methods: AutoAugment [11] (augmentation-based), L1 Norm [5, 24] (regularization-based), MaDi [8] (attention-based), CRD [7] (causal representation disentanglement-based), and CFS [34] (causal feature selection-based). For a fair comparison, all methods are built upon the SAC+AE and share identical test environment and reward function configurations.

As shown in Table 2, although SAC+RAE achieves a moderate ISR, its SSR remains 0.0, indicating that while some UAVs can complete the task individually, they fail to achieve swarm success simultaneously. In contrast, the CCFL method achieves substantial improvements across all evaluation metrics (SSR, ISR, and SPL), clearly demonstrating its superiority. Moreover, CCFL not only achieves a high success rate but also maintains superior trajectory efficiency, characterized by a smaller extra distance and a higher average speed, demonstrating its ability to achieve both reliable and collision avoidance efficiency.

4.4 More Evaluation Scenes

To further evaluate the generalization ability of the proposed method, two additional environments named snow mountain and canyon are designed. As shown in Figure 7, both environments are constructed with the same obstacle configuration as the forest scenario to ensure consistent evaluation conditions. The results in TABLE 3 show that CCFL outperforms the baseline on all metrics, achieving significant gains in SSR and ISR, which demonstrates its strong effectiveness and robustness in unseen environments.

4.5 Ablation Study

The ablation experiments are conducted to assess the individual contributions of the CFI and RFC modules within the proposed framework. As shown in TABLE 4, the results indicate that removing either module leads to a noticeable performance decline in SSR and ISR. When both modules are integrated, the model achieves the highest performance, confirming that CFI and RFC work synergistically to improve generalization and collision avoidance efficiency.

TABLE 3: Performance comparison under different backgrounds.

Scene	Seen/Unseen	Method	SSR (%)	ISR (%)	SPL (%)	Extra Distance (m)	Average Speed (m/s)
Forest	Unseen	SAC+RAE	0.0	51.9	44.0	6.427/5.652	0.781/0.182
		Our method	**61.6** (↑ 61.6)	**93.9** (↑ 42.0)	**79.9**	**4.258/1.165**	**0.972/0.080**
Snow Mountain	Unseen	SAC+RAE	0.0	40.0	33.0	5.977/5.797	0.757/0.233
		Our method	**38.0** (↑ 38.0)	**90.1** (↑ 50.1)	**74.2**	**5.251/2.087**	**0.984/0.115**
Canyon	Unseen	SAC+RAE	0.0	52.1	44.8	4.575/5.172	0.898/0.107
		Our method	**50.9** (↑ 50.9)	**92.3** (↑ 40.2)	**78.4**	**4.332/1.731**	**0.932/0.084**

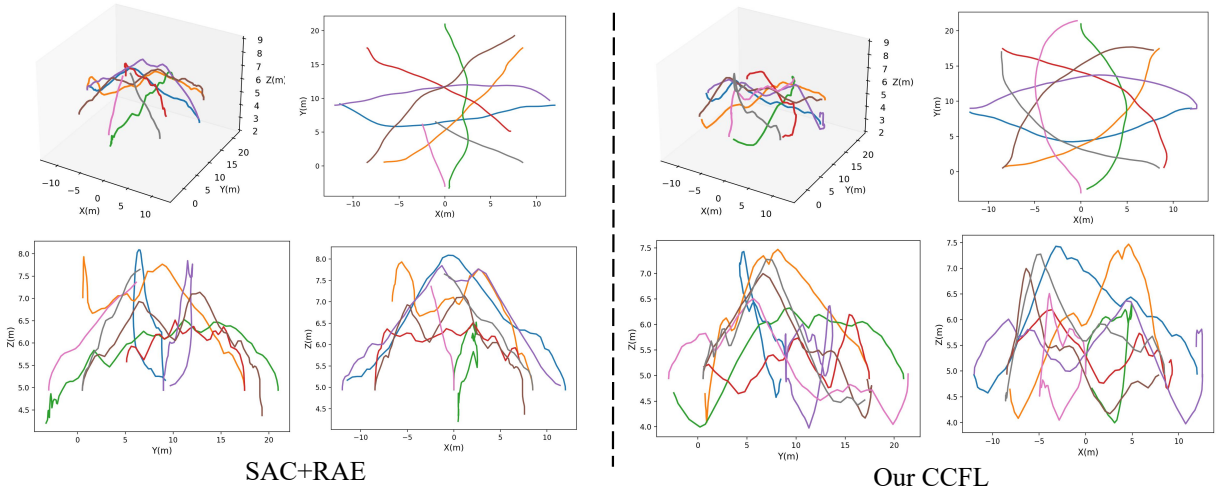

Figure 8: Visualization of UAV trajectories in perspective and three-view drawings. The trajectories of different UAVs are represented in distinct colors. Best viewed in color.

TABLE 4: Ablation study on two proposed modules.

CFI	RFC	SSR (%)	ISR (%)
✓		15.2	81.8
	✓	1.0	64.5
✓	✓	**61.6**	**93.9**

4.6 Trajectory Visualization

To further evaluate the performance of CCFL, the trajectories of UAVs are visualized in a forest environment with densely distributed obstacles. As shown in Figure 8, the SAC+RAE exhibits irregular and dispersed flight paths, indicating unstable navigation and inefficient collision avoidance. In contrast, the proposed CCFL method produces smoother, more consistent trajectories with fewer abrupt changes and better spatial coordination among UAVs. These results demonstrate that CCFL enables UAVs to maintain stable flight and effective collision avoidance even in unseen environments.

5 CONCLUSIONS

In this article, we propose a novel CCFL framework. The framework employs a CFI module to disentangle input features into causal and non-causal components, and further integrates a RFC module to compact causal representations, yielding features that are both causal and compact. To verify the effectiveness of compact causal representations in improving the generalization capability of UAVs, we design a series of cross-scenario experiments by altering background and introducing unseen obstacles. The experimental results demonstrate that the CCFL method significantly improves generalization, enabling effective adaptation to complex and unseen environments. Furthermore, it achieves superior collision avoidance performance compared with existing methods addressing the DRL generalization problem, confirming its effectiveness and robustness.

ACKNOWLEDGMENTS

This work is supported in part by the National Science and Technology Major Project (grant number 2021ZD0111502), the National Natural Science Foundation of China (grant numbers 62406186, 62476163), the Natural Science Foundation of Guangdong Province (grant number 2025A1515010800), the Guangdong Basic and Applied Basic Research Foundation (grant number 2023B1515120020), the State Key Laboratory of Autonomous Intelligent Unmanned Systems (grant number ZZKF2025-3-4), the GBA Ascend Application Innovation Institute, Guangdong Laboratory of Artificial Intelligence and Digital Economy (SZ) (grant number GML-ST-2026-02).

REFERENCES

[1] Kai Arulkumaran, Marc Peter Deisenroth, Miles Brundage, and Anil Anthony Bharath. 2017. Deep reinforcement learning: A brief survey. *IEEE Signal Processing Magazine* 34, 6 (2017), 26–38.

[2] Reece A Boyd and Salvador E Barbosa. 2017. Reinforcement learning for all: An implementation using unreal engine blueprint. In *2017 International Conference on Computational Science and Computational Intelligence (CSCI)*. IEEE, Las Vegas, USA, 787–792.

[3] Johann Brehmer, Pim De Haan, Phillip Lippe, and Taco S Cohen. 2022. Weakly supervised causal representation learning. *Advances in Neural Information Processing Systems* 35, 5 (2022), 38319–38331.

[4] Ruichu Cai, Zijian Li, Pengfei Wei, Jie Qiao, Kun Zhang, and Zhifeng Hao. 2019. Learning disentangled semantic representation for domain adaptation. In *IJCAI*, Vol. 2019. NIH Public Access, Macau, China, 2060.

[5] Yikun Cheng, Pan Zhao, Fanxin Wang, Daniel J Block, and Naira Hovakimyan. 2022. Improving the Robustness of Reinforcement Learning Policies With ℓ_1 Adaptive Control. *IEEE Robotics and Automation Letters* 7, 3 (2022), 6574–6581.

[6] Pim De Haan, Dinesh Jayaraman, and Sergey Levine. 2019. Causal confusion in imitation learning. *NeurIPS* 32 (2019), 11698–11709.

[7] Zhun Fan, Zihao Xia, Che Lin, Gaofei Han, Wenji Li, Dongliang Wang, Yindong Chen, Zhifeng Hao, Ruichu Cai, and Jiafan Zhuang. 2024. UAV Collision Avoidance in Unknown Scenarios with Causal Representation Disentanglement. *Drones* 9, 1 (2024), 10.

[8] Bram Grooten, Tristan Tomilin, Gautham Vasan, Matthew E Taylor, A Rupam Mahmood, Meng Fang, Mykola Pechenizkiy, and Decebal Constantin Mocanu. 2024. MaDi: Learning to Mask Distractions for Generalization in Visual Deep Reinforcement Learning. In *Proceedings of the 23rd International Conference on Autonomous Agents and Multiagent Systems*. International Foundation for Autonomous Agents and Multiagent Systems, Auckland, New Zealand, 733–742.

[9] Tuomas Haarnoja, Aurick Zhou, Pieter Abbeel, and Sergey Levine. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*. PMLR, Stockholm, Sweden, 1861–1870.

[10] Gaofei Han, Qingling Wu, Boxi Wang, Che Lin, Jiafan Zhuang, Wenji Li, Zhifeng Hao, and Zhun Fan. 2024. Deep Reinforcement Learning Based Multi-UAV Collision Avoidance with Causal Representation Learning. In *2024 10th International Conference on Big Data and Information Analytics (BigDIA)*. IEEE, Chiang Mai, Thailand, 833–839.

[11] Nicklas Hansen and Xiaolong Wang. 2021. Generalization in reinforcement learning by soft data augmentation. In *2021 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, Xi'an, China, 13611–13617.

[12] Bruce Hoadley. 1971. Asymptotic properties of maximum likelihood estimators for the independent not identically distributed case. *The Annals of mathematical statistics* 42, 6 (1971), 1977–1991.

[13] Stefan Hrabar. 2011. Reactive obstacle avoidance for rotorcraft uavs. In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, San Francisco, USA, 4967–4974.

[14] Kaiyu Hu, Huanlin Li, Jiafan Zhuang, Zhifeng Hao, and Zhun Fan. 2023. Efficient Focus Autoencoders for Fast Autonomous Flight in Intricate Wild Scenarios. *Drones* 7, 10 (2023), 609.

[15] Changxin Huang, Ronghui Zhang, Meizi Ouyang, Pengxu Wei, Junfan Lin, Jiang Su, and Liang Lin. 2021. Deductive reinforcement learning for visual autonomous urban driving navigation. *IEEE Transactions on Neural Networks and Learning Systems* 32, 12 (2021), 5379–5391.

[16] Huaxing Huang, Guijie Zhu, Zhun Fan, Hao Zhai, Yuwei Cai, Ze Shi, Zhaohui Dong, and Zhifeng Hao. 2022. Vision-based Distributed Multi-UAV Collision Avoidance via Deep Reinforcement Learning for Navigation. In *IEEE/RSJ International Conference on Intelligent Robots and Systems*. IEEE, Kyoto, Japan, 13745–13752.

[17] Chanyoung Ju and Hyoung Il Son. 2019. Modeling and control of heterogeneous agricultural field robots based on Ramadge–Wonham theory. *IEEE Robotics and Automation Letters* 5, 1 (2019), 48–55.

[18] Yannis Kalantidis, Mert Bulent Sariyildiz, Noe Pion, Philippe Weinzaepfel, and Diane Larlus. 2020. Hard negative mixing for contrastive learning. *Advances in neural information processing systems* 33 (2020), 21798–21809.

[19] San Kim, Donggeun Kim, Siheon Jeong, Ji-Wan Ham, Jae-Kyung Lee, and Ki-Yong Oh. 2020. Fault diagnosis of power transmission lines using a UAV-mounted smart inspection system. *IEEE access* 8 (2020), 149999–150009.

[20] Zijian Li, Ruichu Cai, Guangyi Chen, Boyang Sun, Zhifeng Hao, and Kun Zhang. 2024. Subspace Identification for Multi-Source Domain Adaptation. *NeurIPS* 36 (2024), 34504–34518.

[21] Che Lin, Gaofei Han, Qingling Wu, Boxi Wang, Jiafan Zhuang, Wenji Li, Zhifeng Hao, and Zhun Fan. 2025. Improving Generalization in Collision Avoidance for Multiple Unmanned Aerial Vehicles via Causal Representation Learning. *Sensors* 25, 11 (2025), 3303.

[22] Romain Lopez, Jeffrey Regier, Michael I Jordan, and Nir Yosef. 2018. Information constraints on auto-encoding variational bayes. *Advances in neural information processing systems* 31 (2018), 787–792.

[23] Yuncheng Lu, Zhucun Xue, Gui-Song Xia, and Liangpei Zhang. 2018. A survey on vision-based UAV navigation. *Geo-spatial information science* 21, 1 (2018), 21–32.

[24] Jun Ma, Liming Yang, and Qun Sun. 2020. Capped L1-norm distance metric-based fast robust twin bounded support vector machine. *Neurocomputing* 412 (2020), 295–311.

[25] Kate Rakelly, Abhishek Gupta, Carlos Florensa, and Sergey Levine. 2021. Which mutual-information representation learning objectives are sufficient for control? *Advances in Neural Information Processing Systems* 109, 5 (2021), 612–634.

[26] Bernhard Schölkopf, Francesco Locatello, Stefan Bauer, Nan Rosemary Ke, Nal Kalchbrenner, Anirudh Goyal, and Yoshua Bengio. 2021. Toward causal representation learning. *Proc. IEEE* 109, 5 (2021), 612–634.

[27] Shital Shah, Debadeepta Dey, Chris Lovett, and Ashish Kapoor. 2017. Airsim: High-fidelity visual and physical simulation for autonomous vehicles. In *Field and service robotics: Results of the 11th international conference*. Springer, Zurich, Switzerland, 621–635.

[28] Javad Shahmoradi, Elaheh Talebi, Pedram Roghanchi, and Mostafa Hassanalian. 2020. A comprehensive review of applications of drone technology in the mining industry. *Drones* 4, 3 (2020), 34.

[29] Yulun Tian, Katherine Liu, Kyel Ok, Loc Tran, Danette Allen, Nicholas Roy, and Jonathan P How. 2020. Search and rescue under the forest canopy using multiple UAVs. *The International Journal of Robotics Research* 39, 10-11 (2020), 1201–1221.

[30] Juan-Carlos Trujillo, Rodrigo Munguia, Edmundo Guerra, and Antoni Grau. 2018. Cooperative monocular-based SLAM for multi-UAV systems in GPS-denied environments. *Sensors* 18, 5 (2018), 1351.

[31] Andrew Ward. 2013. Spurious correlations and causal inferences. *Erkenntnis* 78 (2013), 699–712.

[32] Zhenda Xie, Zheng Zhang, Yue Cao, Yutong Lin, Jianmin Bao, Zhuliang Yao, Qi Dai, and Han Hu. 2022. Simmim: A simple framework for masked image modeling. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*. IEEE, New Orleans, USA, 9653–9663.

[33] Yunsheng Zhang, Fanfan Liu, Jia Zhang, and Haifeng Li. 2025. Causal Invariant Representation Learning Based on Style Intervention Identity Regularization for Remote Sensing Image. *IEEE Geoscience and Remote Sensing Letters* 109, 5 (2025), 612–634.

[34] Jiafan Zhuang, Gaofei Han, Zihao Xia, Che Lin, Boxi Wang, Dongliang Wang, Wenji Li, Zhifeng Hao, Ruichu Cai, and Zhun Fan. 2025. Robust Policy Learning for Multi-UAV Collision Avoidance with Causal Feature Selection. In *Proceedings of the 24th International Conference on Autonomous Agents and Multiagent Systems*. IEEE, Detroit, USA, 2392–2401.

[35] Qiming Zou and Einoshin Suzuki. 2024. Compact goal representation learning via information bottleneck in goal-conditioned reinforcement learning. *IEEE Transactions on Neural Networks and Learning Systems* 7, 3 (2024), 6574–6581.