Deep Reinforcement Learning based Multi-UAV Collision Avoidance with Causal Representation Learning

1st Gaofei Han*
Department of Electronic Engineering
Shantou University
Shantou, China
22gfhan@stu.edu.cn

2nd Qingling Wu*
Department of Electronic Engineering
Shantou University
Shantou, China
22qlwu@stu.edu.cn

3rd Boxi Wang
Department of Electronic Engineering
Shantou University
Shantou, China
19bxwang@stu.edu.cn

4th Che Lin
Department of Electronic Engineering
Shantou University
Shantou, China
19clinl@stu.edu.cn

5th Jiafan Zhuang✉
Department of Electronic Engineering
Shantou University
Shantou, China
jfzhuang@stu.edu.cn

6th Wenji Li✉
Department of Electronic Engineering
Shantou University
Shantou, China
liwj@stu.edu.cn

7th Zhifeng Hao
Department of Electronic Engineering
Shantou University
Shantou, China
haozhifeng@stu.edu.cn

8th Zhun Fan✉
Shenzhen Institute for Advanced Study
University of Electronic Science and Technology of China
Shenzhen, China
zfan@stu.edu.cn

***Abstract*—The deep reinforcement learning-based multi-UAV collision avoidance and navigation methods have made significant progress. However, the fundamental challenges of those methods is their limited ability to generalize beyond the specific domains they are trained on. We find that the cause of the generalization failures is attributed to spurious correlation. To address this issue, we propose a causal representation learning method to identify the causal representations from images. Specifically, our method can ignore factors of variation that are irrelevant to the deep reinforcement learning task through causal intervention. Subsequently, the causal representations are fed into the policy network for action prediction. Experimental results show that our method exhibits better generalization performance compared to state-of-the-art method in different testing scenarios.**

***Index Terms*—Causal Representation Learning, Deep Reinforcement Learning, Multi-UAV Systems, Collision Avoidance and Navigation.**

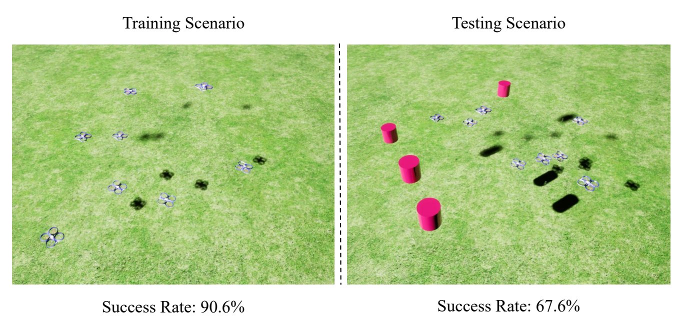


Fig. 1. **Illustration on the influence of spurious correlations**. Due to the presence of spurious correlations, deep reinforcement learning algorithms have a tendency to memorize solutions of seen obstacles in the training scenario, thereby achieving high training rewards with a brittle policy that will not generalize to unseen obstacles in the testing scenario.

I. Introduction

With the development of artificial intelligence and robotics, unmanned aerial vehicle systems (UAVs) [1]–[4] technology has been widely applied, such as aerial photography [5], [6], pesticide spraying [7], [8], search and rescue [9], [10]. Collision avoidance and navigation capability are crucial to ensure UAVs operate safely and effectively. The UAVs must have the ability to avoid obstacles in complex environments while identifying the optimal path from a starting point to a target point.

There has been a significant amount of research and related works on UAVs collision avoidance and navigation. The traditional approaches [11]–[13]for UAVs collision avoidance and navigation are based on simultaneous localization and mapping (SLAM) [14]. These approaches construct a local environment map based on SLAM for path planning to achieve collision avoidance and navigation. Another method is the path retracing [15], [16] which enables the UAVs to autonomously navigate along previously demonstrated or pre-planned paths.

✉ Corresponding author
* These authors contributed equally to this work

However, this method has limited adaptability. A common characteristic of these traditional methods is their reliance on prior maps of the environment and manually designed features. These methods have gradually become inadequate for practical navigation tasks in terms of robustness, generalization, and adaptability.

To explore more adaptable solutions in the field of machine learning, deep reinforcement learning (DRL) [17]–[19] have received significant attention. DRL can learn and optimize strategies adaptively in complex and dynamic environments without the need for manual parameter tuning and rule adjustment. Specifically, DRL extracts low-dimensional state representations from raw observational data. Then the DRL model selects the optimal action based on the current state representation and updates the policy or value function based on the rewards or penalties received from the environment. This process aims to maximize long-term rewards, ultimately deriving the optimal strategy to accomplish the desired task.

However, one of the fundamental challenges with DRL approaches is their limited ability to generalize beyond the specific domains they are trained on. Consider the example shown in Fig. 1. The UAVs are trained in a playground environment with only UAV-like obstacles and must learn to navigate to a goal located from a start located. In the testing scenario, we introduce several unseen obstacles. The policy trained using DRL techniques demonstrates strong performance when deployed in training scenario with UAV-like obstacles. However, its performance significantly degrades in testing scenario with unseen obstacles. This example illustrates the difficulty of existing deep reinforcement learning methods to generalize to scenarios with unseen obstacles.

To reveal the reasons for the generalization failures in reinforcement learning, we review pioneering work in the field of multi-UAV collision avoidance based on deep reinforcement learning, i.e.,SAC+RAE) [20]. We find that the cause of the generalization failures is attributed to spurious correlations [21], [22]. Specifically, DRL algorithms have a tendency to memorize solutions of seen obstacles during the training phase, thereby achieving high training rewards with a brittle policy that will not generalize to unseen obstacles. Moreover, learned policies often fail to ignore non-causal factors (e.g.,the obstacle shape) in their sensor observations and are highly sensitive to changes in such non-causal factors.

Recent studies also show that the generalization of DRL policies can be improved substantially by causal representation learning. Causal representation learning [23]–[25] aims to find the high level causal variables from low-level observations. It can effectively extract causal factors that impact the task, reducing the influence of spurious correlations on the generalization ability of DRL. In this paper, we propose a Causal Representation Learning (CRL) method to identify the causal representations from images. Specially, we only intervene the shape of obstacles within deep images, keeping the rest content of the image unchanged. Then the images of spherical obstacles and cubic obstacle are used solely as auxiliary task for causal representation learning. Through its auxiliary task, CRL learns to generalize by maximizing the mutual information between latent representations of different obstacles images. In addition, we apply supervision signals to the latent representations to reduce redundancy between dimensions. Intuitively, this encourages different dimensions to capture distinct information. After that, CRL learns to ignore factors of variation that are irrelevant to the DRL task, which greatly enhance generalization ability of DRL model.

To evaluate the generalization ability of CRL, we conduct several testing scenarios across different obstacles. In comparison with previous state-of-the-art (SOTA) methods, the results demonstrate that CRL outperforms the SOTA across different testing scenarios and prove the effectiveness of our method.

II. RELATED WORK

A. DRL-based Collision Avoidance Navigation

The multi-UAV collision avoidance navigation based on DRL is a process of training UAVs to navigate to a goal located from a start located without collision through an environment with actions and corresponding rewards. Pham *et al*. [26] discretize the UAV's flight plane into a series of grids and then use deep q-network (DQN) [27] to decide the actions that the UAV should take at a given position. However, they only consider the problem of flying in an environment without obstacles and do not include high-dimensional inputs such as images. Walvekar *et al*. [28] use DQN to address the collision avoidance problem for UAV in a 3D environment. The authors use first-person view images from the UAV as input to the DQN network, but they do not explicitly provide navigation goals, so their approach cannot solve navigation problems with random targets. To address the issue of random target navigation, Kersandt *et al*. [29] concatenate the goals and observation and input these into the DQN algorithm. This method maps the information to left turn, right turn, and forward movement actions, enabling the UAV to reach the random target points. However, the above methods are unable to capture causal representations in visual observation, making it difficult to generalize to unseen obstacles scenarios.

B. Causal Representation Learning

Our work builds on the nascent field of causal representation learning. Causal representation learning [23]–[25] is a method to model the causal relationships within the data. It can reduce the effect of spurious correlations and transform the data into a structured representation that aligns with physical laws. Yang *et al*. [30] propose a method called CausalVAE, which is the first to introduce structural causal models into representation learning. This approach considers the relationships between generative factors in the data from a causal perspective. To address the problem of learning causal representation from multiple distributions, Zhang *et al*. [31] propose applying sparsity constraints to the latent variable graph structure. This method can recover the latent causal variables and their relationships. To address the sparsity of supervisory signals and the long-tail problem in causal representation learning, Zhao *et al*. [32] propose a causal representation decoupling learning

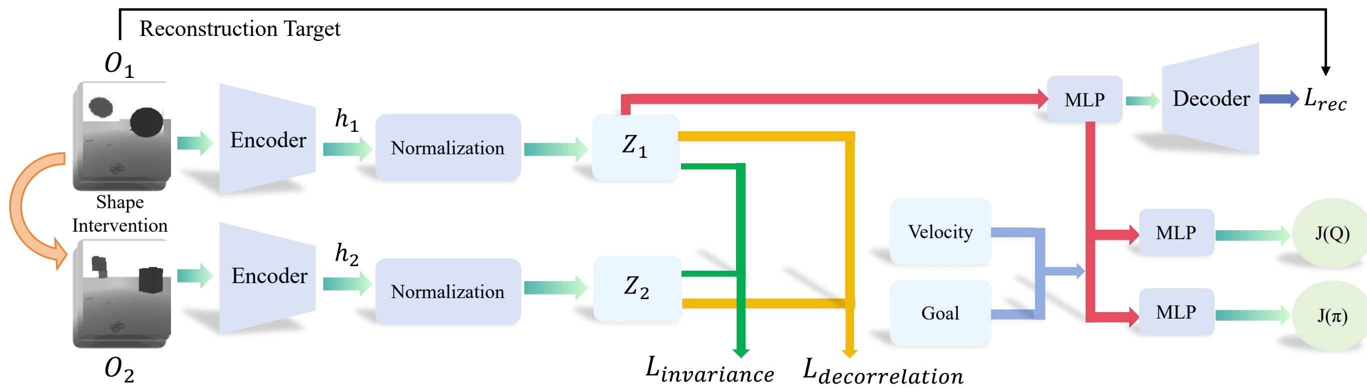


Fig. 2. **The illustration of causal representation learning method.** To extract invariant causal representation, we intervene the shape of obstacles and maximize the mutual information through the $\mathcal{L}_{invariance}$. In addition, we introduce the $\mathcal{L}_{decorrelation}$ to ensure that different dimensions capture different information.

based on contrastive learning. This method can effectively improve the accuracy and robustness of model prediction. In this paper, we conduct causal intervention on the shape of obstacles to extract invariant causal representations. This method can effectively reduce the influence of spurious correlations on the generalization ability of DRL.

III. APPROACH

A. Problem Formulation

DRL [17]–[19] is a method that studies how an intelligent agent interacts with environment to achieve the maximum reward. The agent formulates the strategy based on the current state of the environment, and this iterative process exhibits Markov properties. Therefore, DRL can be modeled as a Markov decision process. The POMDP can be described as a 5-tuple $(\mathcal{S}, \mathcal{A}, \mathcal{P}, \mathcal{R}, \gamma)$, where $\mathcal{S}$ is the set of agent state spaces and $\mathcal{A}$ denotes the set of agent action spaces. $\mathcal{P}$ denotes the conditional probability of state transitions $P^a_{ss'} = P[S_{t+1} = S' \mid S_t = s, A_t = a]$. γ is called the discount factor, used to describe the discounting of reward values over time. The ultimate goal of this process is to establish a policy that maps the state space to the optimal control actions.

1) Observation space: The observation space of the UAV can be defined as $O = [I, V, G]$, where I denotes the accumulation of three consecutive depth images. V denotes the speed of the UAVs at the current moment. G denotes the euclidean distance between the UAV's current position and the target position.

2) Action space: To maintain the continuity of the UAV's movement, we designed the UAV's continuous action space $A = [v_x, v_y, v_z]$. v_x represents the UAV's forward velocity. v_y denotes the turning velocity of the UAVs. And v_z denotes the climbing velocity of the UAVs.

3) Reward function: The purpose of collision avoidance navigation is to control the UAVs to reach the target position in a complex environment without collision. This process can be divided into two subtasks: target approach r_g and obstacle avoidance r_c. Therefore, when the UAVs safely reaches the target positions, it should be given positive feedback as a reward. Conversely, when a collision occurs, negative feedback should be given as a punishment.

$$r = r_g + r_c \tag{1}$$

$$r_g = \begin{cases} r_{arrival} & if\ d_t < 0.5 \\ \alpha_{goal} \cdot (d_t - d_{t-1}) & otherwise \end{cases} \tag{2}$$

where d_t is the distance between the UAV' position and the target positions at time t. $r_{arrival}$ is the reward for UAVs that have reached the target positions. And α_{goal} is the reward weight.

$$r_c = \begin{cases} r_{collision} & if\ crash \\ \alpha_{avoid} \cdot max(d_{safe} - d_{min}, 0) & otherwise \end{cases} \tag{3}$$

where $r_{collision}$ is the collision penalty. α_{avoid} is the penalty weight. And d_{safe} is the safe distance of UAVs.

B. Architectural Overview

Our approach builds upon the previous work [20]. As shown in Figure 2, our framework consists of two parts: representation learning and policy learning. We construct a causal representation learning structure based auto-encode. It aims to learn representations that effectively reduce the information of obstacle shape. Subsequently, based on those causal representations along with the current velocity and the target position are fed into policy network for strategy learning. According to the current environment, SAC algorithm [33] returns a set of actions for UAV control, including velocity on three dimensions.

In the policy improvement step, we update the actor network by maximizing the loss function $J(\pi)$, which can be expressed as follows:

$$J(\pi) = \mathbb{E}_{o \sim \mathcal{B}} \left[D_{KL}(\pi(\cdot|o) \| \mathcal{Q}(o, \cdot)) \right] \tag{4}$$

We update the critic network by minimizing the loss function $J(Q)$, which can be expressed as follows:

$$J(Q) = \mathbb{E}_{(o,a,r,o') \sim \mathcal{B}} \left[(Q(o,a) - r - \gamma \bar{V}(o'))^2 \right] \tag{5}$$

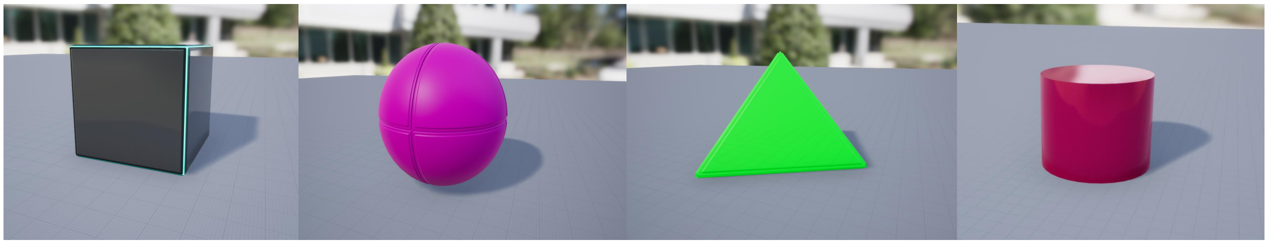


Fig. 3. **Simulation scenarios for evaluating the generalization ability.** We set up four different shapes of obstacles during the testing phase to evaluate the effectiveness of our method. All obstacles are set to the same size.

We adopt the loss function $\mathcal{L}(rec)$ to reconstruction image through updating the encoder [34] and decoder [35] with the following objective:

$$\mathcal{L}(rec) = \mathbb{E}_x \left[\log p_\phi(x|z) + \lambda_z \|z\|^2 + \lambda_\phi \|\phi\|^2 \right] \quad (6)$$

C. Extract Invariant Causal Representation

The key idea of this section is to learn causal representations from images that makes the optimal policy built on top of this representations invariant across training domains. Effectively, this approach attempts to learn and exploit the causes of successful actions. Some works [36], [37] has demonstrated that we need to discover invariant mechanisms from multiple source domain data and identify hidden causal variables. Therefore, in this work, we conduct causal intervention [38], [39] on the shape of obstacles to construct multiple source domain data with different obstacles. Specifically, we use Airsim's image rendering technology [40], [41] to change the shape of the obstacles while keeping other content unchanged. As shown in Fig. 2, we adopt a auto-encoder (AE) [42]–[44] for representation disentanglement. We use a depth camera to sample different observations$\mathcal{O}_1$, $\mathcal{O}_2$ from the environment. $\mathcal{O}_1$ denotes the images containing the UAVs and spherical obstacles. $\mathcal{O}_2$ denotes the images containing both the UAVs and cubic obstacles. We feed them into the encoder to extract representations h1, h2 respectively. Then we apply an instance-dimensional normalization to ensure each feature dimension has a 0-mean and $\frac{1}{\sqrt{N}}$-standard deviation distribution, which is implemented as:

$$z = \frac{h_i - \mu(h_i)}{\sigma(h_i) * \sqrt{N}} \quad (7)$$

The obtained normalized z_1, z_2 are further used to maximize the mutual information through the invariance term:

$$\mathcal{L}_{invariance} = ||z_1 - z_2||^2 \quad (8)$$

Intuitively, the invariance term is used to minimize the difference between two normalized representations.

Furthermore, previous work [45] has shown that multiple dimensions in representations share overlapping information. To ensure that different dimensions capture different information, we introduce the following decorrelation term:

$$\mathcal{L}_{decorrelation} = \mathcal{F}(z_1, z_1^T, I) + \mathcal{F}(z_2, z_2^T, I) \quad (9)$$

where $\mathcal{F}(\cdot,\cdot) = \| \cdot - \cdot \|_F^2, \| \cdot \|_F^2$ denotes the Frobenius norm and I is an identity matrix. The decorrelation term can avoid the collapsed trivial solution outputting the same vector for all inputs by trying to equate the off-diagonal elements of the auto-correlation matrix of each representation to 0.

IV. EXPERIMENT AND RESULTS

A. Simulation Environment and Experiments Setup

Airsim [41], [46] is a high-fidelity software used for UAV simulation testing. Therefore, we choose Airsim as the basis for our simulation system to conduct research on UAV obstacle avoidance algorithms. The simulation scenario experiments are conducted on a simulation computer with the following software configuration: Ubuntu 20.04 operating system, Intel i9-12900k CPU, and a single NVIDIA RTX 3090 GPU.

B. Performance Metrics and Experiment Scenarios

To evaluate the effectiveness of our method, we define the following metrics:

- Success Rate: The percentage of UAVs that successfully reach their target within a limited time without any collisions.
- SPL (Success weighted by Path Length): This metric considers not only the success rate of the task but also the efficiency of the path.
- Extra Distance: The extra distance traveled by the UAVs to reach the target point compared to the straight-line distance between the initial and target positions.
- Average Speed: The average speed of all UAVs.

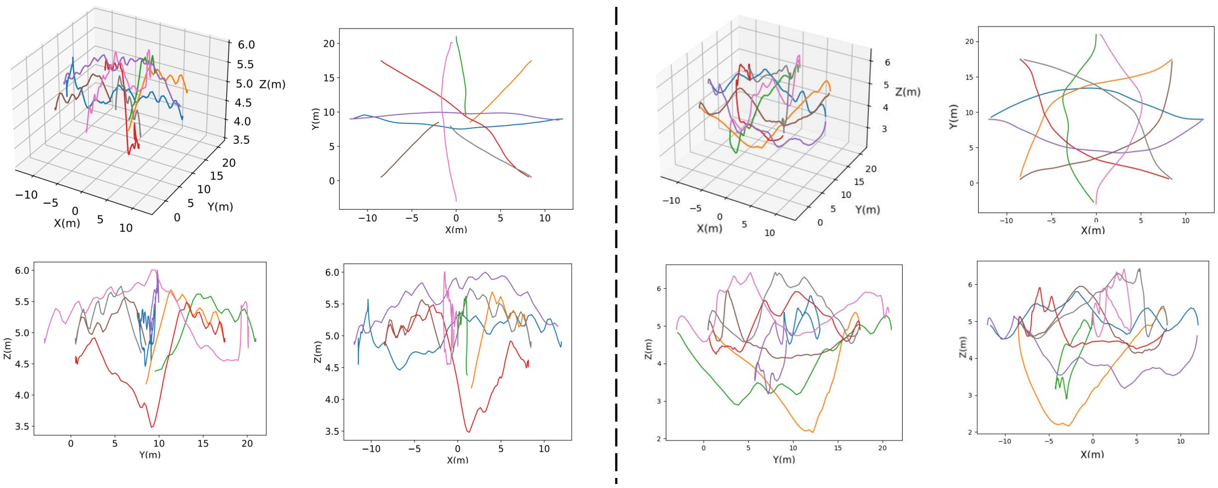
As shown in Fig. 3, to better illustrate the effectiveness of our proposed method in improving the generalization ability, we evaluate the CRL on the challenging shape of obstacle,such as: cube, sphere, triangle and cylinder. We have fixed the size of the simulation area by 16*16*4 and consider 8 UAVs besides four static obstacles. The initial and target positions of the UAVs and obstacles are randomly generated within this area.

C. Performance Comparison

We compare our method with the SOTA method, i.e.,SAC+RAE, under different obstacles scenarios. As shown in TABLE I, our method have a higher success rate and SPL compared to SAC+RAE, which clearly shows that the

TABLE I
PERFORMANCE (AS MEAN/STD) COMPARISON UNDER DIFFERENT BACKGROUNDS.

Obstacle shape	Seen/Unseen	Method	Success Rate (%)	SPL (%)	Extra Distance (m)	Average Speed (m/s)
Cube	seen	SAC+RAE	67.6	58.3	**1.483/1.436**	0.771/0.158
		Our method	**74.4** (↑ 6.8)	**62.5** (↑ 4.2)	1.542/1.519	**0.806/0.125**
Sphere	seen	SAC+RAE	68.3	58.4	**1.680/1.709**	0.783/0.160
		Our method	**73.4** (↑ 5.1)	**62.2** (↑ 3.8)	1.735/1.718	**0.814/0.132**
Triangle	Unseen	SAC+RAE	72.5	62.1	**1.535/1.622**	0.775/0.164
		Our method	**76.3** (↑ 3.8)	**64.4** (↑ 2.3)	1.815/1.802	**0.809/0.135**
Cylinder	Unseen	SAC+RAE	67.6	58.6	**1.646/1.621**	0.783/0.155
		Our method	**72.1** (↑ 4.5)	**61.7** (↑ 3.1)	1.853/1.810	**0.810/0.138**



SAC+RAE

Our Method

Fig. 4. **Visualization of UAV trajectories in perspective drawing and three-view drawing**. We use different colors to represent trajectories of different UAVs.

proposed causal representation learning method can effectively improve the generalization ability of DRL model for unseen obstacles. At the same time, our method demonstrates a faster average speed, which reduce the time to reach the target point. However, our method have a longer extra distance compared to SAC+RAE. It indicates that each UAV need to adjusts its planned path more frequently to avoid collisions.

D. Visualization

To more intuitively demonstrate the collision avoidance performance of our method, we provide the visualization of UAVs trajectories in perspective drawing and three-view drawing. As shown in Fig. 4, our method demonstrates better path planning and collision avoidance performance compared to SAC+RAE.

V. CONCLUSIONS

In this paper, we propose a causal representation learning method to identify the causal variables from images to improve the generalization ability of DRL model. we intervene the shape of obstacles and maximize the mutual information through the invariance term. In addition, we introduce the decorrelation term to ensure that different dimensions capture different information. Experimental results show that our method exhibits better generalization performance compared to state-of-the-art method in different testing scenarios.

ACKNOWLEDGMENT

This work is supported in part by the National Science and Technology Major Project (grant numbers 2021ZD0111501, 2021ZD0111502), the National Natural Science Foundation of China (grant numbers 62176147, 51907112, U2066212, 61961036, 62162054), Science and Technology Planning Project of Guangdong Province of China (grant numbers 2023A1515011574, 2022A1515110566, 2022A1515110660), and the STU Scientific Research Foundation for Talents (grant numbers NTF21001, NTF21052, NTF22030).

REFERENCES

[1] H. Eisenbeiss *et al.*, “A mini unmanned aerial vehicle (uav): system overview and image acquisition,” *International Archives of Photogrammetry. Remote Sensing and Spatial Information Sciences*, 2004.

[2] C. Ju and H. I. Son, “A distributed swarm control for an agricultural multiple unmanned aerial vehicle system,” *Proceedings of the Institution of Mechanical Engineers, Part I: Journal of Systems and Control Engineering*, 2019.

[3] H. Li, Y. Cai, J. Hong, P. Xu, H. Cheng, X. Zhu, B. Hu, Z. Hao, and Z. Fan, “Vg-swarm: A vision-based gene regulation network for uavs swarm behavior emergence,” *IEEE Robotics and Automation Letters*, 2023.

[4] W. Li, P. Ren, Z. Wang, D. Wang, B. Xu, C. Guan, C. Chen, J. Zhuang, and Z. Fan, “A multi-agent control model based on gene regulatory network,” in *2023 9th International Conference on Big Data and Information Analytics (BigDIA)*, 2023.

[5] D. Gilvear and R. Bryant, “Analysis of aerial photography and other remotely sensed data,” *Tools in fluvial geomorphology*, 2003.

[6] X. Li and L. Yang, “Design and implementation of uav intelligent aerial photography system,” in *2012 4th International Conference on Intelligent Human-Machine Systems and Cybernetics*, 2012.

[7] M. F. F. Rahman, S. Fan, Y. Zhang, and L. Chen, “A comparative study on application of unmanned aerial vehicle systems in agriculture,” *Agriculture*, 2021.

[8] P. Garre and A. Harish, “Autonomous agricultural pesticide spraying uav,” in *IOP Conference Series: Materials Science and Engineering*, 2018.

[9] Y. Naidoo, R. Stopforth, and G. Bright, “Development of an uav for search & rescue applications,” in *IEEE Africon’11*, 2011.

[10] E. T. Alotaibi, S. S. Alqefari, and A. Koubaa, “Lsar: Multi-uav collaboration for search and rescue missions,” *IEEE Access*, 2019.

[11] R. Li, H. Shang, M. Shen, and Y. Xiao, “Vision slam-based uav obstacle avoidance system,” in *AIP Conference Proceedings*, 2024.

[12] R. Li, J. Liu, L. Zhang, and Y. Hang, “Lidar/mems imu integrated navigation (slam) method for a small uav in indoor environments,” in *2014 DGON inertial sensors and systems (ISS)*, 2014.

[13] A. Yusefı, A. Durdu, and C. Sungur, “Orb-slam-based 2d reconstruction of environment for indoor autonomous navigation of uavs,” *Avrupa Bilim ve Teknoloji Dergisi*, 2020.

[14] T. Taketomi, H. Uchiyama, and S. Ikeda, “Visual slam algorithms: A survey from 2010 to 2016,” *IPSJ transactions on computer vision and applications*, 2017.

[15] J. Lee, R. Huang, A. Vaughn, X. Xiao, J. K. Hedrick, M. Zennaro, and R. Sengupta, “Strategies of path-planning for a uav to track a ground vehicle,” in *AINS Conference*, 2003.

[16] F. L. Junior, L. A. Moreira, E. M. Moreira, T. J. Baldivieso, M. S. Brunaes, and P. F. Rosa, “Uav path automation using visual waypoints acquired from the ground,” in *2020 IEEE 29th International Symposium on Industrial Electronics (ISIE)*, 2020.

[17] Y. Li, “Deep reinforcement learning: An overview,” *arXiv preprint arXiv:1701.07274*, 2017.

[18] C. Wang, J. Wang, Y. Shen, and X. Zhang, “Autonomous navigation of uavs in large-scale complex environments: A deep reinforcement learning approach,” *IEEE Transactions on Vehicular Technology*, 2019.

[19] C. Wang, J. Wang, J. Wang, and X. Zhang, “Deep-reinforcement-learning-based autonomous uav navigation with sparse rewards,” *IEEE Internet of Things Journal*, 2020.

[20] H. Huang, G. Zhu, Z. Fan, H. Zhai, Y. Cai, Z. Shi, Z. Dong, and Z. Hao, “Vision-based distributed multi-uav collision avoidance via deep reinforcement learning for navigation,” in *2022 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*, 2022.

[21] A. Ward, “Spurious correlations and causal inferences,” *Erkenntnis*, 2013.

[22] C. Liu, X. Sun, J. Wang, H. Tang, T. Li, T. Qin, W. Chen, and T.-Y. Liu, “Learning causal semantic representation for out-of-distribution prediction,” *Advances in Neural Information Processing Systems*, 2021.

[23] B. Schölkopf, F. Locatello, S. Bauer, N. R. Ke, N. Kalchbrenner, A. Goyal, and Y. Bengio, “Toward causal representation learning,” *Proceedings of the IEEE*, 2021.

[24] J. Brehmer, P. De Haan, P. Lippe, and T. S. Cohen, “Weakly supervised causal representation learning,” *Advances in Neural Information Processing Systems*, 2022.

[25] J. Mitrovic, B. McWilliams, J. Walker, L. Buesing, and C. Blundell, “Representation learning via invariant causal mechanisms,” *arXiv preprint arXiv:2010.07922*, 2020.

[26] H. X. Pham, H. M. La, D. Feil-Seifer, and L. V. Nguyen, “Autonomous uav navigation using reinforcement learning,” *arXiv preprint arXiv:1801.05086*, 2018.

[27] M. Roderick, J. MacGlashan, and S. Tellex, “Implementing the deep q-network,” *arXiv preprint arXiv:1711.07478*, 2017.

[28] A. Walvekar, Y. Goel, A. Jain, S. Chakrabarty, and A. Kumar, “Vision based autonomous navigation of quadcopter using reinforcement learning,” in *2019 IEEE 2nd International Conference on Automation, Electronics and Electrical Engineering (AUTEEE)*, 2019.

[29] K. Kersandt, G. Muñoz, and C. Barrado, “Self-training by reinforcement learning for full-autonomous drones of the future,” in *2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC)*, 2018.

[30] M. Yang, F. Liu, Z. Chen, X. Shen, J. Hao, and J. Wang, “Causalvae: Disentangled representation learning via neural structural causal models,” in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2021.

[31] K. Zhang, S. Xie, I. Ng, and Y. Zheng, “Causal representation learning from multiple distributions: A general setting,” *arXiv preprint arXiv:2402.05052*, 2024.

[32] W. Zhao, D. Tang, X. Chen, D. Lv, D. Ou, B. Li, P. Jiang, and K. Gai, “Disentangled causal embedding with contrastive learning for recommender system,” in *Companion Proceedings of the ACM Web Conference 2023*, 2023.

[33] T. Haarnoja, A. Zhou, K. Hartikainen, G. Tucker, S. Ha, J. Tan, V. Kumar, H. Zhu, A. Gupta, P. Abbeel *et al.*, “Soft actor-critic algorithms and applications,” *arXiv preprint arXiv:1812.05905*, 2018.

[34] G. Zhu, J. Liu, Z. Fan, D. Yuan, P. Ma, M. Wang, W. Sheng, and K. C. Wang, “A lightweight encoder–decoder network for automatic pavement crack detection,” *Computer-Aided Civil and Infrastructure Engineering*, 2024.

[35] K. Hu, H. Li, J. Zhuang, Z. Hao, and Z. Fan, “Efficient focus autoencoders for fast autonomous flight in intricate wild scenarios,” *Drones*, 2023.

[36] P. Stojanov, Z. Li, M. Gong, R. Cai, J. Carbonell, and K. Zhang, “Domain adaptation with invariant representation learning: What transformations to learn?” *Advances in Neural Information Processing Systems*, 2021.

[37] K. Zhang, M. Gong, and B. Schölkopf, “Multi-source domain adaptation: A causal view,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2015.

[38] D. Zhang, H. Zhang, J. Tang, X.-S. Hua, and Q. Sun, “Causal intervention for weakly-supervised semantic segmentation,” *Advances in Neural Information Processing Systems*, 2020.

[39] Y. Hagmayer, S. A. Sloman, D. A. Lagnado, and M. R. Waldmann, “Causal reasoning through intervention,” *Causal learning: Psychology, philosophy, and computation*, 2007.

[40] B. Alvey, D. T. Anderson, A. Buck, M. Deardorff, G. Scott, and J. M. Keller, “Simulated photorealistic deep learning framework and workflows to accelerate computer vision and unmanned aerial vehicle research,” in *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 2021.

[41] S. Shah, D. Dey, C. Lovett, and A. Kapoor, “Airsim: High-fidelity visual and physical simulation for autonomous vehicles,” in *Field and Service Robotics: Results of the 11th International Conference*, 2018.

[42] M. Tschannen, O. Bachem, and M. Lucic, “Recent advances in autoencoder-based representation learning,” *arXiv preprint arXiv:1812.05069*, 2018.

[43] Q. Meng, D. Catchpoole, D. Skillicom, and P. J. Kennedy, “Relational autoencoder for feature extraction,” in *2017 International joint conference on neural networks (IJCNN)*, 2017.

[44] Z. Fan, C. Li, Y. Chen, J. Wei, G. Loprencipe, X. Chen, and P. Di Mascio, “Automatic crack detection on road pavements using encoder-decoder architecture,” *Materials*, 2020.

[45] Q. Ji, J. Li, J. Hu, R. Wang, C. Zheng, and F. Xu, “Rethinking dimensional rationale in graph contrastive learning from causal perspective,” in *Proceedings of the AAAI Conference on Artificial Intelligence*, 2024.

[46] J. Hong, D. Chen, W. Li, and Z. Fan, “Trajectory planner for uavs based on potential field obtained by a kinodynamic gene regulation network,” *Sensors*, 2023.