Fourier Guided Adaptive Adversarial Augmentation for Generalization in Visual Reinforcement Learning

Jeong Woon Lee and Hyoseok Hwang*

Department of Software Convergence, Kyung Hee University, Republic of Korea
{everyman123,hyoseok}@khu.ac.kr

Abstract

Visual Reinforcement Learning (RL) facilitates learning directly from raw images; however, the domain gap between training and testing environments frequently leads to a decline in performance within unseen environments. In this paper, we propose Fourier Guided Adaptive Adversarial Augmentation (FGA3), a novel augmentation method that maintains semantic consistency. We focus on style augmentation in the frequency domain by keeping the phase and altering the amplitude to preserve the state of the original data. For adaptive adversarial perturbation, we reformulate the worst-case problem to RL by employing adversarial example training, which leverages value loss and cosine similarity within a semantic space. Moreover, our findings illustrate that cosine similarity is effective in quantifying feature distances within a semantic space. Extensive experiments on DMControl-GB and Procgen have shown that FGA3 is compatible with a wide range of visual RL algorithms, both off-policy and on-policy, and significantly improves the robustness of the agent in unseen environments.

Introduction

Reinforcement Learning (RL) integrated with Deep Neural Networks (DNNs) has shown outstanding achievements in a range of tasks challenging at a human level (Le et al. 2022). Visual RL, a sub-field of RL that processes complex visual information instead of relying on pre-defined state representations, opens avenues for learning directly from unprocessed sensory data (Mnih et al. 2013). This approach minimizes the need for manually engineered features and enhances the adaptability of RL (Ma et al. 2022). Visual RL has demonstrated success in intricate and dynamic settings such as robotics (Nair et al. 2018), navigation (Zhu et al. 2017), video gaming (Kanervisto et al. 2021), and autonomous driving (Cai et al. 2021).

Despite the promising potential, visual RL experiences a notable decline in performance when applied to unseen environments (Huang et al. 2022). The principal contributor to this discrepancy, the domain gap, is attributed to limited data diversity stemming from variations between the training and testing environments (Wang et al. 2020). To address this challenge, data augmentation has been extensively employed to broaden training data distribution to encompass the unseen domain (Zhou et al. 2022). However, subjecting a model to a wide range of augmentations without thorough consideration of semantic consistency may not guarantee generalization in visual RL. This is because some augmentation methods, such as flipping and rotation, may inadvertently alter the state information and conflict with environmental dynamics. This discrepancy could lead to training the model toward an unintended direction (Huang et al. 2022). Therefore, an augmentation method that preserves the original image's semantics is crucial for generalization in visual RL.

Several studies have explored augmentation strategies to maintain semantic consistency to improve generalization (Lee, Bae, and Kim 2023). Among them, normalization techniques like adaptive instance normalization (AdaIN) are commonly employed (Huang and Belongie 2017; Jin et al. 2020). AdaIN encapsulates style through normalization parameters, comprising the features' channel-wise mean and standard deviation. In other methods, some approaches give variation to style in the frequency domain by preserving the phase and altering the amplitude (Yang et al. 2020; Xu et al. 2021) leveraging a key aspect of the Fourier transform: the phase retains the original signal's high-level semantic information, while the amplitude holds low-level statistics corresponding to the style (Xu et al. 2021). In addition to studying *what* causes style transfer, there has also been research on *how* to apply them to broaden the range of a distribution, and there are two main streams: mixed- and perturbation-based approaches.

In the mixed-based approach, diversity in style is enhanced by mixing the data styles through the normalization parameters or Fourier amplitude (Hong, Choi, and Kim 2021; Zhou, Qi, and Shi 2022). On the other hand, the perturbation-based approach increases the style diversity by adding a fixed or randomized magnitude of perturbation to the style (Zhang et al. 2019; Chattopadhyay et al. 2023). The perturbation-based approach offers the advantage of allowing control over the extent of covariate shifts by adjusting the magnitude of perturbations enabling the introduction of diversity beyond the original training data distribution.

In recent studies, there has been a growing emphasis on generating adaptive perturbations through adversarial training rather than relying on fixed or randomly created perturbations (Fan et al. 2021; Qiao and Peng 2021). One ap-

*Corresponding author (hyoseok@khu.ac.kr)

proach trains the additional model to generate the adaptive perturbation from training data (Lee, Ahn, and Park 2022). Another approach formulates the worst-case problem and generates adaptive perturbation by modifying the perturbation directly (Volpi et al. 2018). The perturbation is crafted to increase the loss while decreasing the semantic distance between the adversarial example and the original in the semantic space. This approach does not require the additional cost of training a model to make perturbations. Inspired by this advantage, Policy-Aware Adversarial Data Augmentation (PAADA) (Zhang and Guo 2022) applies the worst-case problem to visual RL in a policy gradient-based algorithm and improves the generalization. However, this method's superior performance depended on contributing when combined with the Mixup, not just adversarial augmentation alone. This limitation opens up opportunities for improved application of the worst-case problem in visual RL.

Therefore, in this study, we propose **F**ourier **G**uided **A**daptive **A**dversarial **A**ugmentation (**FGA3**), a novel augmentation method that preserves semantic consistency. To achieve this goal, we apply adversarial perturbation in the Fourier amplitude domain to extend the distribution while preserving semantics. Then, we reformulate the worst-case problem in two phases to make visual RL more efficient. First, we utilize the value loss to generate an adversarial example assuming that training with semantically similar samples with significant differences in value loss makes the model more robust to unseen environments. Also, we employ cosine similarity as the distance metric motivated by research showing the usefulness of cosine similarity in feature space (Chung, Kim, and Kwak 2022). We evaluate the effectiveness of FGA3 on generalization in DMControl Generalization Benchmark (DMControl-GB) (Hansen, Su, and Wang 2021) and Procgen (Cobbe et al. 2020). Extensive experiments have demonstrated that FGA3 can integrate with various on/off-policy algorithms and significantly enhance the robustness of agents in unseen environments. The main contributions of this study can be summarized as follows:

- We introduce FGA3, an effective frequency-based adaptive adversarial augmentation method that preserves semantic information of original data and learns from vulnerability, thus enhancing the model's generalization capabilities in unseen environments.
- We propose an efficient approach to applying the worst-case problem to visual RL by applying value-based loss and cosine similarity, which are validated to improve performance more efficiently than existing methods.
- We demonstrate that the proposed method significantly enhances the performance of the integrated base algorithm on generalization in previously unseen environments.

Related Work

Augmentation in Visual Reinforcement Learning

RL agents often suffer from overfitting to training environments and result in poor generalization performance to unseen environments (Cobbe et al. 2019). To tackle this issue, augmentation has been widely utilized within the realm of visual RL for generalization in recent years (Lee et al. 2019; Laskin et al. 2020; Raileanu et al. 2021). Ma *et al.* (Ma et al. 2022) argue that augmentation encourages agents to comprehend inherently invariant representations and stabilizes models. Lee *et al.* (Lee et al. 2019) demonstrate that augmenting images with randomized networks can improve the generalization capabilities of agents. Reinforcement Learning with Augmented Data (RAD) (Laskin et al. 2020) implements pixel-level transformations to add diversity to the images. Raileanu *et al.* (Raileanu et al. 2021) emphasize the importance of choosing a data augmentation in a given task and propose Data-regularized Actor-Critic (DrAC) to choose the proper data augmentation. In contrast to previous studies that attempted augmentation of images in the spatial domain, Spectrum Random Masking (SRM) proposed by Huang *et al.* (Huang et al. 2022) randomly masks frequency components to preserve the structure of the images. However, SRM masks even the phase component, where the semantic information of the image is encoded. Alternative approaches (Zhou et al. 2021; Lee, Ahn, and Park 2022) have explored augmentation strategies that maintain semantic consistency through normalization. MixStyle (Zhou et al. 2021) uses AdaIN to separate content and style and mixes the styles while keeping the underlying semantic content intact. However, Lee *et al.* (Lee, Bae, and Kim 2023) argue that normalization techniques, like instance normalization, can distort the semantic information in images by altering the phase of data. Compared to previous approaches, FGA3 maintains semantic consistency with the original data by limiting augmentation to the Fourier amplitude. Furthermore, unlike Fourier augmentation for domain generalization proposed by Xu *et al.* (Xu et al. 2021), our method is specifically tailored to enhance generalization in visual RL.

Adversarial Training for Domain Generalization

Szegedy *et al.* (Szegedy et al. 2014) uncover the vulnerability to subtle perturbations in DNNs. After that, adversarial training is introduced as a strategy to train the robust models. Recently, there has been a trend of applying adversarial training for domain generalization that treats adversarial examples as augmented instances (Fan et al. 2021; Qiao and Peng 2021). Volpi *et al.* (Volpi et al. 2018) formulate domain generalization as a worst-case problem and attempt to achieve it by training on adversarial examples that increase the loss in the current model. Advstyle (Zhong et al. 2022) separates content and style using instance normalization(Huang and Belongie 2017) and generates adversarial examples by modifying the style for semantic segmentation. In visual RL, Oikarinen *et al.* (Oikarinen et al. 2021) propose Robust ADversarIAl Loss (RADIAL) to promote robust training against adversarial perturbation. RADIAL calculates the lower bounds of the value function to ensure robust action selection under the perturbed observations. Zhang *et al.* (Zhang and Guo 2022) propose PAADA that applies the worst-case problem to a policy gradient-based algorithm to enhance generalization. Style-Agnostic RL (SAR) (Lee, Ahn, and Park 2022) generates adversarial examples by modifying the normalization parameters through a generative model. However, SAR requires relatively more training steps because it in-

volves training a model to generate adversarial examples. In contrast, our method directly modifies perturbations and does not require the additional cost of training a model.

Background

Reinforcement Learning

Reinforcement Learning (RL) is a subset of machine learning where agents learn decision-making through interaction with the environment. RL is formulated on Markov Decision Process (MDP) $\mathcal{M} = \langle \mathcal{S}, \mathcal{A}, \mathcal{P}, r, \gamma \rangle$, where $\mathcal{S}$ is the state space, $\mathcal{A}$ is the action space, $\mathcal{P} : \mathcal{S} \times \mathcal{A} \to \mathcal{S}$ models the probability distribution of transitioning to different states in response to the action $a_t \in \mathcal{A}$ taken at time t given the current state $s_t \in \mathcal{S}$, r is a reward, and $\gamma \in [0, 1)$ is a discount factor. In visual RL, s_t represents the image observed by the agent. The primary goal of RL is to train agents to take actions that maximize the expected return expressed as $\mathbb{E}\left[\sum_{t=0}^{\infty} \gamma^t r_t\right]$, where r_t is a reward at time t.

Adversarial Training

Adversarial training enhances the robustness of the model by incorporating adversarially perturbed examples into the training set (Goodfellow, Shlens, and Szegedy 2015). Adversarial examples are crafted by inducing a specific perturbation to the input data, defined as

$$x' = x + \epsilon \text{ sign}(\nabla_x J(\theta, x, y)), \tag{1}$$

where x' is the adversarial example for the input x, by adding perturbations with magnitude ϵ. The sign of the perturbation is determined from the sign of the gradient of the loss function $\nabla_x J(\theta, x, y)$. Training the model with these adversarial examples can significantly improve its robustness.

Fourier Transform

Fourier transform is a critical method in analyzing image frequency components. Consider an image $o \in \mathbb{R}^{H \times W \times C}$ where H, W, and C represent height, width, and channel, respectively. $\mathcal{F}$ converts image to Fourier transformation $\mathcal{F}(o) \in \mathbb{C}^{H \times W \times C}$, expressed as

$$\mathcal{F}(o)(u, v) = \sum_{h=0}^{H-1} \sum_{w=0}^{W-1} o(h, w) e^{-j2\pi(\frac{hu}{H} + \frac{wv}{W})}, \tag{2}$$

where $u \in [0, H-1]$ and $v \in [0, W-1]$ are the horizontal and vertical components, respectively, and $o(h, w)$ represents the pixel value at position (h, w). $\mathcal{F}(o)(u, v)$ consists of a real part $\mathcal{R}(o)(u, v)$ and an imaginary part $\mathcal{I}(o)(u, v)$. $\mathcal{F}(o)(u, v)$ can be further decomposed into

$$\mathcal{F}(o)(u, v) = |\mathcal{F}(o)(u, v)| e^{j\angle \mathcal{F}(o)(u,v)}, \tag{3}$$

where the amplitude $|\mathcal{F}(o)(u, v)|$ is expressed as

$$|\mathcal{F}(o)(u, v)| = \sqrt{\mathcal{R}^2(o)(u, v) + \mathcal{I}^2(o)(u, v)}, \tag{4}$$

and the phase $\angle \mathcal{F}(o)(u, v)$ defined as

$$\angle \mathcal{F}(o)(u, v) = \arctan \frac{\mathcal{I}(o)(u, v)}{\mathcal{R}(o)(u, v)}. \tag{5}$$

Phase retains a high level of the semantic information of the image, while the amplitude primarily preserves its low-level statistics (Hansen and Hess 2007; Xu et al. 2021). The image is reconstructed by the Inverse Fourier Transform formulated as

$$o(h, w) = \frac{1}{HW} \sum_{u=0}^{H-1} \sum_{v=0}^{W-1} \mathcal{F}(o)(u, v) e^{j2\pi\left(\frac{uh}{H} + \frac{vw}{W}\right)}. \tag{6}$$

Proposed Method

Problem Definition

Consider the worst-case problem (Volpi et al. 2018) for the off-policy algorithm in the vicinity of the source domain S_0 formulated as

$$\underset{\theta \in \Theta}{\text{minimize}} \sup_{S: D(S, S_0) \leq \rho} \mathbb{E}_S \left[\mathcal{L}_Q(\theta; \mathcal{B})\right], \tag{7}$$

where $\theta \in \Theta$ is the parameter of model, $D(S, S_0)$ represents the domain distance between S_0 and S. The source domain S_0 and target domain S refer to the distribution of states. $\mathcal{L}_Q$ is the loss function, known as Temporal Difference (TD) loss, defined as

$$\mathcal{L}_Q(\theta; \mathcal{B}) := \mathbb{E}_{s_t, a_t, r_t, s_{t+1} \sim \mathcal{B}} \left[\frac{1}{2}(y_t - Q_\theta(s_t, a_t))^2\right], \tag{8}$$

where $\mathcal{B}$ is the replay buffer, Q_θ is the action value function parameterized by θ, and y_t is the TD target:

$$y_t := r_t + \gamma \max_{a'_t} Q_\theta(s_{t+1}, a'_t). \tag{9}$$

In actor-critic, TD loss corresponds to the critic loss, also known as value loss. For on-policy algorithms such as Proximal Policy Optimization (PPO) (Schulman et al. 2017), value function V_θ is used to calculate the value loss instead of Q_θ. Minimizing the supremum of TD loss across domains that are distance ρ away from the source domain S_0 ensures that we can get Q_θ consistent across those domains, equivalent to achieving the generalization of the agent.

Overview of FGA3

The overall framework of our method is depicted in Figure 1. FGA3 consists of two steps: Adversarial Example Learning and Robust Agent Training. In Adversarial Example Learning, the adversarial examples are generated by modifying the amplitude in the direction of increasing the value loss and decreasing the cosine similarity distance $d_\theta(z, \bar{z})$ between the original latent vector z and adversarial latent vector $\bar{z}$. In Robust Agent Training, the original and adversarial examples are used to train the value network.

Distance Between the Domains

We aim to measure the distance between the domains within a semantic space. Consequently, we define the target domain S such that $D(S, S_0) \leq \rho$, which reflects the feasible covariate shifts while maintaining the semantic content of the source domain (Volpi et al. 2018). We understand the distance in the semantic space as being determined by the distance of

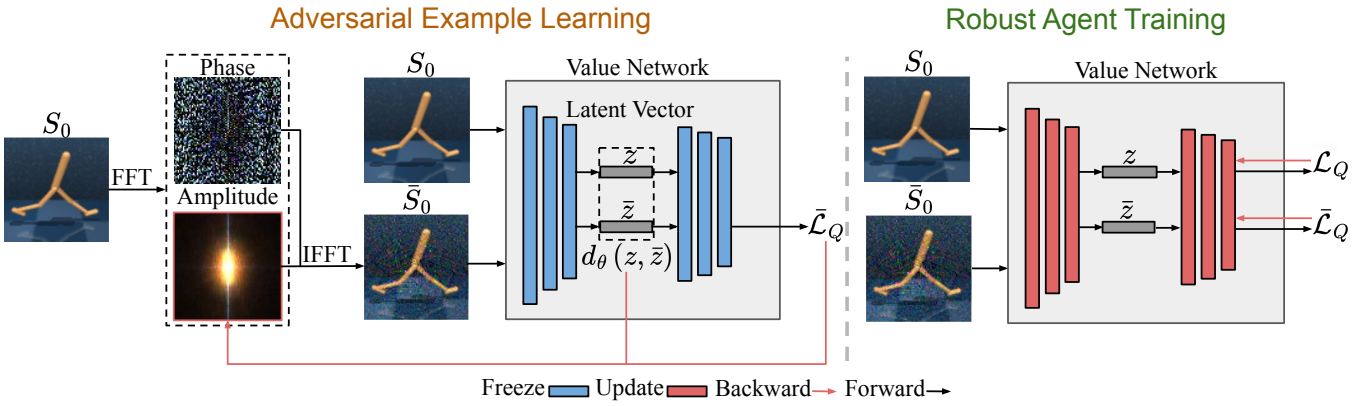

Figure 1: Framework of FGA3.

the learned representations from high-capacity models (Volpi et al. 2018). The semantic distance metric D_θ is defined as

$$D_\theta(S, S') := \mathbb{E}\left[d_\theta(s, s')\right], \tag{10}$$

where d_θ is cosine similarity distance (Grill et al. 2020) expressed as

$$d_\theta(s, s') := 2 - 2\frac{\langle g(\theta; s), g(\theta; s')\rangle}{\|g(\theta; s)\|_2 \, \|g(\theta; s')\|_2}, \tag{11}$$

where $g(\theta; s)$ is the latent vector from the encoder layer of the current model. We utilize the cosine similarity distance, as a higher cosine similarity between a target and source feature indicates shared semantic information. Consequently, we reformulate the worst-case scenario in terms of semantic distance as

$$\underset{\theta\in\Theta}{\text{minimize}} \sup_S \{\mathbb{E}_S\left[\mathcal{L}_Q(\theta; \mathcal{B})\right] : D_\theta(S, S_0) \le \rho\}. \tag{12}$$

Surrogate Loss

For the computational issue, we consider the following Lagrangian relaxation of the worst-case problem with the penalty parameter α formulated as

$$\underset{\theta\in\Theta}{\text{minimize}} \sup_S \{\mathbb{E}_S[\mathcal{L}_Q(\theta; \mathcal{B})] - \alpha D_\theta(S, S_0)\}. \tag{13}$$

The constraint $D_\theta(S, S_0) \le \rho$ is incorporated into the objective function through Lagrangian relaxation. This transformation eliminates the explicit appearance of ρ, replacing it with the Lagrange multiplier α. With the Lagrangian relaxation, we can define the robust surrogate loss (Volpi et al. 2018) defined as

$$\phi_\alpha(\theta; (t(s_0))) := \sup_{s\in S}\{\mathcal{L}_Q(\theta; t(s)) - \alpha d_\theta(s, s_0)\}, \tag{14}$$

where $t(s) = (s, a_0, r_0, s_1)$ represents transition. We can apply the stochastic gradient descent method to the robust surrogate loss, denoted as ϕ_α. Under suitable conditions (Boyd and Vandenberghe 2004; Volpi et al. 2018), we have

$$\nabla_\theta \phi_\alpha(\theta; t(s_0)) = \nabla_\theta \mathcal{L}_Q(\theta; t(s_\alpha^\star)), \tag{15}$$

Algorithm 1: Fourier Guided Adaptive Adversarial Augmentation

Input: Collection of transitions $\mathcal{C} = \{(s_t, a_t, r_t, s_{t+1})\}_{t=0}^T$
Output: learned parameter θ
Init: $\theta \leftarrow \theta_0, \bar{\mathcal{C}} \leftarrow \emptyset$

for $t = 1, \ldots, T$ **do**
  # Get the Fourier transformation by Eq. 2.
  $\mathcal{F}(s_t) \leftarrow s_t$
  # Calculate amplitude and phase by Eq. 4, 5.
  $|\mathcal{F}(s_t)|, \angle\mathcal{F}(s_t) \leftarrow \mathcal{F}(s_t)$
  $|\bar{\mathcal{F}}| \leftarrow |\mathcal{F}(s_t)|$
  $\bar{s}_t \leftarrow s_t$
  for $k = 1, \ldots, K$ **do**
    $\mathcal{L}_{adv} = \mathcal{L}_Q(\theta; (\bar{s}_t, a_t, r_t, s_{t+1})) - \alpha d_\theta(\bar{s}_t, s_t)$
    $|\bar{\mathcal{F}}| \leftarrow |\bar{\mathcal{F}}| + \eta_{adv}\nabla_{|\bar{\mathcal{F}}|}\mathcal{L}_{adv}$
    # Generate the new Fourier transformation by Eq. 3.
    $\bar{\mathcal{F}}(\bar{s}_t) \leftarrow |\bar{\mathcal{F}}|, \angle\mathcal{F}(s_t)$
    # Apply Inverse Fourier Transform by Eq. 6.
    $\bar{s}_t \leftarrow \bar{\mathcal{F}}(\bar{s}_t)$
  end for
  $\bar{\mathcal{C}} \leftarrow \bar{\mathcal{C}} \cup (\bar{s}_t, a_t, r_t, s_{t+1})$
end for
$\theta \leftarrow \theta - \eta\nabla_\theta\{\mathcal{L}_Q(\theta; \mathcal{C}) + \mathcal{L}_Q(\theta; \bar{\mathcal{C}})\}$
return θ

where $s_\alpha^\star$ is an adversarial example of s_0 at the current model expressed as

$$s_\alpha^\star = \arg\max_{s\in S}\{\mathcal{L}_Q(\theta; t(s)) - \alpha d_\theta(s, s_0)\}. \tag{16}$$

Consequently, we undertake two steps: generating the adversarial example $s_\alpha^\star$, and training the parameter of model θ through gradient descent on $\nabla_\theta \mathcal{L}_Q(\theta; t(s_\alpha^\star))$.

Adversarial Example Learning

First, we extract a collection of transitions $\mathcal{C} = \{(s_t, a_t, r_t, s_{t+1})\}_{t=0}^T$ from the replay buffer $\mathcal{B}$. Then, for the state s_t of each transition, we apply the Fourier Transform and calculate the amplitude $|\mathcal{F}(s_t)(u, v)|$ and the phase $\angle\mathcal{F}(s_t)(u, v)$. After that, we initialize the amplitude $|\bar{\mathcal{F}}|$ by $|\mathcal{F}(s_t)(u, v)|$ which is regarded as learnable parameters and

	Random Colors						Natural Videos					
	WW	WS	CS	BC	FS	Avg ↑	WW	WS	CS	BC	FS	Avg ↑
SAC	410 (44)	557 (34)	637 (29)	275 (184)	600 (152)	496	350 (46)	510 (14)	492 (30)	189 (122)	362 (119)	381
PAD	512 (65)	732 (46)	515 (51)	566 (38)	797 (35)	624	716 (17)	934 (30)	520 (114)	444 (74)	540 (7)	631
DrQ	638 (29)	810 (54)	574 (37)	591 (141)	749 (98)	672	776 (44)	910 (36)	455 (71)	440 (43)	562 (31)	629
SVEA	812 (127)	446 (374)	846 (12)	963 (9)	983 (2)	810	536 (94)	385 (265)	616 (78)	544 (28)	886 (14)	593
SAC+SRM	505 (22)	691 (125)	719 (28)	692 (62)	802 (34)	682	485 (31)	630 (102)	518 (15)	434 (15)	566 (33)	527
DrQ+SRM	733 (67)	834 (51)	670 (2)	543 (98)	889 (20)	734	867 (20)	932 (43)	489 (36)	520 (100)	734 (35)	<u>708</u>
SVEA+SRM	840 (86)	702 (360)	836 (16)	968 (1)	946 (41)	<u>858</u>	700 (113)	649 (335)	630 (28)	704 (54)	752 (126)	687
SAC+FGA3	713 (46)	893 (9)	549 (47)	767 (44)	864 (94)	757	688 (64)	876 (20)	475 (40)	567 (44)	622 (25)	646
DrQ+FGA3	699 (5)	865 (30)	592 (56)	772 (64)	869 (72)	759	769 (24)	959 (3)	533 (4)	517 (44)	557 (45)	667
SVEA+FGA3	759 (66)	928 (31)	847 (14)	949 (18)	912 (29)	**879**	749 (59)	921 (24)	736 (45)	692 (93)	528 (36)	**725**

Table 1: Comparison with other methods on Random Colors and Natural Videos benchmark in DMControl-GB after training on 500K frames. We provide the mean and standard deviation of episode return trained with three different random seeds. $(\cdot)$ represents the standard deviation.

the state $\bar{s}_t$ by s_t. $|\bar{\mathcal{F}}|$ is updated iteratively for exploring $s^\star_\alpha$ by

$$|\bar{\mathcal{F}}| \leftarrow |\bar{\mathcal{F}}| + \eta_{adv} \nabla_{|\bar{\mathcal{F}}|} \mathcal{L}_{adv}, \tag{17}$$

where η_{adv} is the learning rate for adversarial example learning and $\mathcal{L}_{adv}$ is the adversarial loss defined as

$$\mathcal{L}_{adv} = \mathcal{L}_Q(\theta; (\bar{s}_t, a_t, r_t, s_{t+1})) - \alpha d_\theta(\bar{s}_t, s_t). \tag{18}$$

$\bar{s}_t$ is also generated iteratively through Inverse Fourier Transform each time the amplitude is updated. Finally, we can get the collection of adversarial transitions $\bar{\mathcal{C}} = \{(\bar{s}_t, a_t, r_t, s_{t+1})\}_{t=0}^{T}$.

Robust Agent Training

Given the original collection of transitions $\mathcal{C}$ and the collection of adversarial transitions $\bar{\mathcal{C}}$, we train the agent for optimization formulated as

$$\underset{\theta}{\text{minimize}}\, \mathcal{L}_Q(\theta; \mathcal{C}) + \mathcal{L}_Q(\theta; \bar{\mathcal{C}}). \tag{19}$$

The overall procedure of FGA3 is illustrated in Algorithm 1.

Experiments

Experimental Setup

Simulation. We demonstrated the generalization performance of FGA3 using two distinct benchmarks: **DMControl-GB** (Hansen, Su, and Wang 2021) for assessing the performance of FGA3 combined with an off-policy algorithm, and **Procgen** (Cobbe et al. 2020) for evaluating the performance when integrated with the on-policy algorithm. DMControl-GB involves continuous control tasks using vision-based methods, while Procgen focuses on discrete control tasks. In DMControl-GB, the agent was trained in a stable environment and tested for generalization across environments with potential distribution shifts. Test environments include Random Colors and Natural Videos. Procgen, a robust testbed for visual RL, offers 16 tasks across procedurally generated games, similar to the ALE benchmark. Generalization performance was evaluated on unseen level distributions.

Implementation Detail. For DMControl-GB, we implemented all methods according to the guidelines in (Huang et al. 2022). For our evaluation, we selected Walker walk (WW), Walker stand (WS), Cartpole swingup (CS), Ball in cup catch (BC), and Finger spin (FS). We incorporated our method with Soft Actor Critic (SAC) (Haarnoja et al. 2018) without any augmentation, DrQ (Yarats, Kostrikov, and Fergus 2021), and Stabilized Q-Value Estimation under Augmentation (SVEA) (Hansen, Su, and Wang 2021) in the role of an agent and evaluated its performance through comparisons with SAC, DrQ, and SVEA with SRM (Huang et al. 2022) and original SAC, Policy Adaptation during Deployment (PAD) (Hansen et al. 2021), DrQ, and SVEA. In Procgen, we trained agents on the initial 200 levels and evaluated their generalization performance on the easy distribution mode. We choose every environment in Procgen. FGA3 was integrated with PPO (Schulman et al. 2017) and DrAC (Raileanu et al. 2021) as agents and its effectiveness was assessed by comparing it with methods such as PPO, RAD (Laskin et al. 2020), MixStyle (Zhou et al. 2021), DrAC, RADIAL (Oikarinen et al. 2021), SAR (Lee, Ahn, and Park 2022), and PAADA (Zhang and Guo 2022). PAADA was combined with PPO and DrAC without Mixup as our method for a fair comparison. Each method ran for 500K frames in DMControl-GB and 25M frames in Procgen, with experiments repeated three times using different random seeds. We used the average episode return as the evaluation metric for DMControl-GB based on (Hansen, Su, and Wang 2021) and the average rank as the evaluation metric for Procgen according to (Lee, Ahn, and Park 2022).

Selection Criteria for Comparison Targets. We compared FGA3 against SRM, the state-of-the-art Fourier-based augmentation for Visual RL. RADIAL and SAR were excluded from DMControl-GB due to the unavailability of their codes or results, and PAADA was excluded because its code is not publicly accessible. Instead, we reimplemented the Procgen in the framework of our method to ensure reproducibility. We utilized DrQ and SVEA in DMControl-GB and DrAC in Procgen, as these algorithms are designed to effectively lever-

	PPO	PPO+PAADA	PPO+FGA3	RAD	MixStyle	SAR	RADIAL	DrAC	DrAC+PAADA	DrAC+FGA3
Bigfish	3.00 (1.71)	3.37 (1.56)	6.87 (4.07)	7.53 (4.76)	5.47 (2.41)	1.37 (0.45)	1.22 (0.56)	5.60 (3.83)	9.80 (4.04)	12.60 (1.77)
Starpilot	27.53 (1.02)	29.23 (3.48)	27.57 (11.76)	25.07 (7.45)	28.07 (2.49)	31.70 (11.39)	12.43 (0.69)	32.07 (4.98)	24.10 (4.45)	30.10 (8.19)
Fruitbot	26.60 (2.57)	29.70 (1.04)	29.80 (0.96)	24.20 (0.96)	26.37 (3.73)	23.70 (4.82)	21.54 (0.98)	25.37 (5.47)	28.30 (1.96)	26.03 (3.68)
BossFight	8.77 (0.86)	7.73 (1.64)	9.23 (1.53)	0.20 (0.08)	7.93 (3.50)	7.53 (0.62)	4.32 (0.86)	7.13 (0.57)	5.20 (1.53)	7.77 (1.30)
Ninja	5.33 (1.25)	7.00 (0.82)	5.67 (1.70)	3.67 (0.47)	5.67 (0.94)	5.00 (0.00)	3.67 (0.53)	4.67 (1.25)	7.67 (0.47)	5.67 (1.89)
Plunder	4.97 (0.54)	4.90 (1.19)	5.20 (1.10)	2.60 (0.45)	5.90 (0.65)	4.40 (1.16)	5.21 (0.27)	4.53 (0.50)	8.73 (1.90)	11.77 (3.07)
CaveFlyer	5.83 (1.30)	4.20 (2.36)	6.53 (1.52)	3.43 (0.42)	4.83 (1.93)	3.43 (2.03)	4.17 (0.21)	4.53 (1.52)	6.10 (0.14)	5.67 (0.94)
CoinRun	9.00 (0.00)	8.67 (0.94)	8.67 (0.47)	4.67 (1.89)	8.33 (0.47)	8.33 (1.25)	7.43 (0.26)	8.67 (0.47)	7.67 (0.47)	8.33 (1.25)
Jumper	5.67 (1.25)	6.67 (1.25)	7.00 (1.41)	4.00 (2.45)	5.67 (0.47)	4.00 (1.63)	5.60 (0.65)	5.00 (1.63)	6.33 (1.70)	7.67 (0.47)
Chaser	6.50 (1.64)	2.83 (0.46)	4.69 (0.68)	2.87 (0.48)	4.35 (0.85)	2.48 (0.59)	1.76 (0.74)	8.53 (1.08)	5.11 (0.53)	7.05 (1.28)
Climber	7.00 (1.61)	5.87 (1.31)	5.90 (1.77)	4.40 (1.47)	6.03 (2.03)	4.00 (0.36)	2.81 (0.07)	6.33 (1.31)	5.10 (0.50)	6.87 (1.68)
Dodgeball	1.20 (0.59)	2.20 (0.98)	1.40 (0.43)	2.47 (0.74)	1.73 (0.98)	1.27 (0.34)	1.43 (0.12)	5.87 (2.56)	3.13 (2.16)	6.67 (1.24)
Heist	0.67 (0.47)	2.67 (1.25)	1.67 (1.25)	4.00 (0.82)	2.33 (1.25)	3.67 (0.94)	2.90 (0.29)	3.67 (1.25)	3.67 (1.70)	3.67 (0.47)
Leaper	6.33 (2.05)	5.33 (2.49)	4.67 (1.25)	3.33 (1.25)	2.67 (2.49)	1.67 (0.47)	3.97 (0.75)	3.00 (2.94)	5.33 (1.89)	4.00 (2.16)
Maze	4.67 (1.25)	7.33 (0.47)	6.33 (2.36)	5.67 (0.47)	7.67 (0.47)	5.67 (2.05)	5.23 (0.26)	5.67 (1.25)	7.67 (0.47)	6.67 (1.70)
Miner	10.10 (0.92)	9.57 (1.64)	9.63 (1.84)	8.07 (3.07)	9.83 (0.40)	4.27 (1.02)	6.36 (0.48)	10.43 (1.37)	11.20 (0.92)	10.80 (0.83)
Avg. Rank ↓	5.06	4.81	4.19	7.31	4.94	7.50	8.31	4.81	<u>3.69</u>	**2.88**

Table 2: Comparison with other methods in Procgen after training on 25M frames. We provide the mean and standard deviation of episode return trained with three different random seeds. (·) represents the standard deviation. Avg. Rank is the average of the rankings assigned for each task.

age augmented data. To isolate and assess the augmentation performance of our method, we applied it to SAC and PPO.

Evaluation on DMControl-GB

In the Random Colors Benchmark Table 1, algorithms combined with FGA3 showed the best generalization in environments with color variation. SAC with FGA3 improved performance by 52.6% over its baseline, achieving the highest score among original algorithms except for SVEA. DrQ gained a 12.9% boost, while SVEA achieved an 8.5% improvement, marking the best performance overall. Across all cases, FGA3 outperformed SRM, with SAC showing an 11.0% improvement, DrQ a 3.4% gain, and SVEA a 2.4% increase. In the Natural Videos Benchmark, FGA3 demonstrated superior adaptability to background changes. SAC achieved a 69.6% improvement over its baseline, the highest among original algorithms. DrQ demonstrated a 6.0% increase, while SVEA improved by 22.3% and achieved the best overall performance. Compared to SRM, FGA3 boosted SAC by 22.6% and SVEA by 5.5%, consistently outperforming SRM in all cases except for DrQ.

Evaluation on Procgen

As shown in Table 2, algorithms combined with FGA3 achieved overall better performance than other methods. For PPO, our method not only outperformed RAD, MixStyle, and RADIAL, which utilized PPO as a baseline but also showed better performance compared to SAR and DrAC. For DrAC, FGA3 demonstrated the best overall performance. PPO with our method achieved superior performance over the baseline in 11 of the 16 games, and DrAC with the proposed method outperformed the baseline algorithm, *i.e.*, DrAC, in 12 of the 16 games. Additionally, our method combined with PPO and DrAC outperformed MixStyle and SAR. Unlike SAR, which requires training an additional model to create the adversarial perturbation, the proposed method directly adjusts the perturbations and achieves superior performance. Compared with PAADA, our method achieved better performance in 9 of the 16 games in both PPO and DrAC, respectively. These results demonstrate that our approach to creating perturbations in the direction of increasing the value loss is more effective than PAADA's method of generating perturbations in the direction of decreasing the policy objective.

	NoAug	Flip	Rotation	Mixup	Gaussian	Random Conv	FGA3
RC	496	609	472	216	658	699	**757**
NV	381	453	354	130	537	552	**646**
Avg ↑	439	531	413	173	598	626	**702**

Table 3: Comparison with other augmentation methods. The results show the overall performance of the augmentations for each benchmark. RC and NV represent Random Colors and Natural Videos respectively.

Comparison with Other Augmentation Methods

We conducted a comparative analysis of the proposed method against image-based augmentations: Flip (Laskin et al. 2020), Rotation (Gidaris, Singh, and Komodakis 2018), Mixup (Zhang et al. 2018), Gaussian (Laskin et al. 2020), and Random Convolution (Lee et al. 2019). The experiment was conducted in DMControl-GB and used SAC as the baseline algorithm. Image-based augmentations were applied to both the current state and the next state, following the approach used in SRM. As indicated in Table 3, FGA3 consistently outperformed other augmentation methods. The results suggested that, despite its proven efficacy in various computer vision tasks, Mixup diminished performance in our visual RL setting. We attribute this to the superposition of two states

Phase	Amplitude	Random Colors	Natural Videos	Avg ↑
		409	348	379
✓		120	205	163
	✓	**681**	**660**	**671**
✓	✓	101	394	248

Table 4: Results on ablation study.

Distance Term	Random Colors	Natural Videos	Avg ↑
None	563	551	557
Wasserstein	500	476	488
Cosine Similarity	**688**	**662**	**675**

Table 5: Results on effectiveness of distance term.

	K=1	K=2	K=4	K=8
Random Colors	443	454	606	**777**
Natural Videos	443	504	605	**744**
Avg ↑	443	479	606	**761**

Table 6: Results of FGA3 comparison according to changes in K.

caused by mixing two images.

Ablation Study

We conducted an ablation study to determine in which domain applying perturbation is the most effective strategy. Specifically, we examined whether focusing solely on amplitude, rather than phase or both phase and amplitude, yielded the best results. Using SAC as the baseline algorithm and Walker walk as the environmental task. As shown in Table 4, the strategy of applying perturbations exclusively to the amplitude resulted in optimal performance. On the other hand, applying perturbation to the phase decreased the performance of the baseline model.

Effectiveness of Distance Term

To validate the effectiveness of cosine similarity as the distance metric, we conducted comparisons against scenarios where semantic distance is either not considered or is measured using the Wasserstein distance (Volpi et al. 2018). We employed SAC as the baseline algorithm and used the Walker walk as the experimental task. We trained the agent up to 300K steps. As shown in Table 5, using Cosine Similarity improved generalization performance, while Wasserstein distance performed worse when semantic distance was ignored. Feature norms were consistent within the same environment, making Wasserstein distance proportional to Cosine similarity's feature norm. This amplifies the weight of semantic distance in adversarial loss, potentially limiting worst-case domain scenarios. We further examined the relationship between Wasserstein and Cosine Similarity distance in Appendix F.

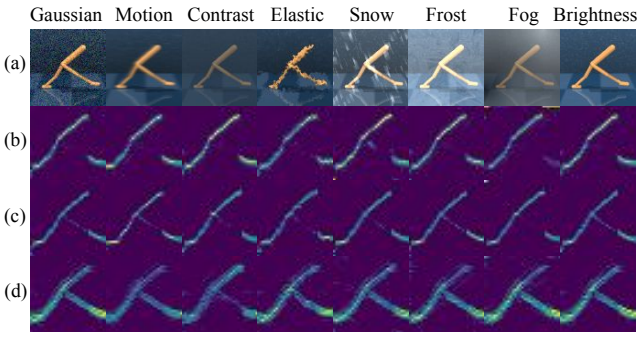

Figure 2: Spatial attention maps generated by an encoder. (a) Corrupted Images, Spatial Attention Map of (b) SAC, (c) SAC+SRM, (d) SAC+FGA3.

Effects of Iteration for Generating Perturbation

To evaluate the effectiveness of iteration in enhancing generalization performance, we compared the iteration parameter K with values of 1, 2, 4, and 8 using SAC as the baseline algorithm. We selected Walker walk as the experimental task. As shown in Table 6, we observed an improvement in generalization performance as the number of iterations for creating perturbations increases. These results suggested that more iterations could create worst-case scenarios enhancing the robustness of agents. A detailed analysis of the iteration's upper bound is provided in the Appendix D.

Comparison Through Spatial Attention Map

To ensure the agent consistently concentrates on task-relevant robot body under diverse image distributions and weather settings during training, we generated spatial attention maps (Laskin et al. 2020) for images corrupted under the Common Corruptions Benchmark (Hendrycks and Dietterich 2019): 4 distributions of images (Gaussian Noise, Motion Blur, Contrast, Elastic Transform) and four weather-related settings (Snow, Frost, Fog, Brightness). Following the approach in (Laskin et al. 2020), we selected a CNN layer, applied a 2-dimensional softmax across each channel, and then averaged these across the channel. Using SAC as the baseline algorithm, we compared no augmentation, SRM, and FGA3. As shown in Figure 2, the agent trained with our method focused more on the robot body relevant to the task, which suggested that our method could provide more comprehensive and robust representations in diverse image distributions.

Conclusion

This study introduces FGA3 as a novel augmentation method in visual RL. We emphasize the improved generalization through adversarial strategy, Fourier amplitude transformation, and cosine similarity distance. Our method effectively preserves semantic fidelity with exposure to diverse, challenging scenarios, strengthening model robustness in unseen environments. This study highlights the potential of FGA3 to enhance the robustness and adaptability representing a step forward in the practical application of visual RL.

Acknowledgements

This work was supported in part by the National Research Foundation of Korea (NRF) grant funded by the Korean government (MSIT) under Grant NRF-2022R1C1C1008074, and in part by the Institute of Information and Communications Technology Planning and Evaluation (IITP) grant funded by the Korean government (MSIT) under Grant RS-2022-00155911, Artificial Intelligence Convergence Innovation Human Resources Development (Kyung Hee University), and in part by Convergence security core talent training business support program under Grant IITP-2023-RS-2023-00266615).

References

Boyd, S. P.; and Vandenberghe, L. 2004. *Convex optimization*. Cambridge university press.

Cai, P.; Wang, H.; Huang, H.; Liu, Y.; and Liu, M. 2021. Vision-based autonomous car racing using deep imitative reinforcement learning. *IEEE Robotics and Automation Letters*, 6(4): 7262–7269.

Chattopadhyay, P.; Sarangmath, K.; Vijaykumar, V.; and Hoffman, J. 2023. Pasta: Proportional amplitude spectrum training augmentation for syn-to-real domain generalization. In *Proceedings of the IEEE/CVF International Conference on Computer Vision*, 19288–19300.

Chung, I.; Kim, D.; and Kwak, N. 2022. Maximizing cosine similarity between spatial features for unsupervised domain adaptation in semantic segmentation. In *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, 1351–1360.

Cobbe, K.; Hesse, C.; Hilton, J.; and Schulman, J. 2020. Leveraging procedural generation to benchmark reinforcement learning. In *International conference on machine learning*, 2048–2056. PMLR.

Cobbe, K.; Klimov, O.; Hesse, C.; Kim, T.; and Schulman, J. 2019. Quantifying generalization in reinforcement learning. In *International conference on machine learning*, 1282–1289. PMLR.

Fan, X.; Wang, Q.; Ke, J.; Yang, F.; Gong, B.; and Zhou, M. 2021. Adversarially adaptive normalization for single domain generalization. In *Proceedings of the IEEE/CVF conference on Computer Vision and Pattern Recognition*, 8208–8217.

Gidaris, S.; Singh, P.; and Komodakis, N. 2018. Unsupervised Representation Learning by Predicting Image Rotations. In *ICLR 2018*.

Goodfellow, I. J.; Shlens, J.; and Szegedy, C. 2015. EXPLAINING AND HARNESSING ADVERSARIAL EXAMPLES. *stat*, 1050: 20.

Grill, J.-B.; Strub, F.; Altché, F.; Tallec, C.; Richemond, P.; Buchatskaya, E.; Doersch, C.; Avila Pires, B.; Guo, Z.; Gheshlaghi Azar, M.; et al. 2020. Bootstrap your own latent-a new approach to self-supervised learning. *Advances in neural information processing systems*, 33: 21271–21284.

Haarnoja, T.; Zhou, A.; Abbeel, P.; and Levine, S. 2018. Soft actor-critic: Off-policy maximum entropy deep reinforcement learning with a stochastic actor. In *International conference on machine learning*, 1861–1870. PMLR.

Hansen, B. C.; and Hess, R. F. 2007. Structural sparseness and spatial phase alignment in natural scenes. *JOSA A*, 24(7): 1873–1885.

Hansen, N.; Jangir, R.; Alenyà Ribas, G.; Abbeel, P.; Efros A, A.; Pinto, L.; and Wang, X. 2021. Self-supervised policy adaptation during deployment. In *International Conference on Learning Representations, ICLR 2021: Vienna, Austria, May 04 2021*, 1–18. OpenReview. net.

Hansen, N.; Su, H.; and Wang, X. 2021. Stabilizing deep q-learning with convnets and vision transformers under data augmentation. *Advances in neural information processing systems*, 34: 3680–3693.

Hendrycks, D.; and Dietterich, T. 2019. Benchmarking Neural Network Robustness to Common Corruptions and Perturbations. *Proceedings of the International Conference on Learning Representations*.

Hong, M.; Choi, J.; and Kim, G. 2021. Stylemix: Separating content and style for enhanced data augmentation. In 2021 IEEE. In *CVF Conference on Computer Vision and Pattern Recognition (CVPR)*, 14857–14865.

Huang, X.; and Belongie, S. 2017. Arbitrary style transfer in real-time with adaptive instance normalization. In *Proceedings of the IEEE international conference on computer vision*, 1501–1510.

Huang, Y.; Peng, P.; Zhao, Y.; Chen, G.; and Tian, Y. 2022. Spectrum Random Masking for Generalization in Image-based Reinforcement Learning. *Advances in Neural Information Processing Systems*, 35: 20393–20406.

Jin, X.; Lan, C.; Zeng, W.; Chen, Z.; and Zhang, L. 2020. Style normalization and restitution for generalizable person re-identification. In *proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 3143–3152.

Kanervisto, A.; Scheller, C.; Schraner, Y.; and Hautamäki, V. 2021. Distilling reinforcement learning tricks for video games. In *2021 IEEE Conference on Games (CoG)*, 01–04. IEEE.

Laskin, M.; Lee, K.; Stooke, A.; Pinto, L.; Abbeel, P.; and Srinivas, A. 2020. Reinforcement learning with augmented data. *Advances in neural information processing systems*, 33: 19884–19895.

Le, N.; Rathour, V. S.; Yamazaki, K.; Luu, K.; and Savvides, M. 2022. Deep reinforcement learning in computer vision: a comprehensive survey. *Artificial Intelligence Review*, 1–87.

Lee, J.; Ahn, S.; and Park, J. 2022. Style-Agnostic Reinforcement Learning. In *European Conference on Computer Vision*, 604–620. Springer.

Lee, K.; Lee, K.; Shin, J.; and Lee, H. 2019. Network randomization: A simple technique for generalization in deep reinforcement learning. *International Conference on Learning Representations, ICLR 2019:*.

Lee, S.; Bae, J.; and Kim, H. Y. 2023. Decompose, Adjust, Compose: Effective Normalization by Playing with Frequency for Domain Generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 11776–11785.

Ma, G.; Wang, Z.; Yuan, Z.; Wang, X.; Yuan, B.; and Tao, D. 2022. A comprehensive survey of data augmentation in visual reinforcement learning. *arXiv preprint arXiv:2210.04561*.

Mnih, V.; Kavukcuoglu, K.; Silver, D.; Graves, A.; Antonoglou, I.; Wierstra, D.; and Riedmiller, M. 2013. Playing atari with deep reinforcement learning. *Advances in neural information processing systems*.

Nair, A. V.; Pong, V.; Dalal, M.; Bahl, S.; Lin, S.; and Levine, S. 2018. Visual reinforcement learning with imagined goals. *Advances in neural information processing systems*, 31.

Oikarinen, T.; Zhang, W.; Megretski, A.; Daniel, L.; and Weng, T.-W. 2021. Robust deep reinforcement learning through adversarial loss. *Advances in Neural Information Processing Systems*, 34: 26156–26167.

Qiao, F.; and Peng, X. 2021. Uncertainty-guided model generalization to unseen domains. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 6790–6800.

Raileanu, R.; Goldstein, M.; Yarats, D.; Kostrikov, I.; and Fergus, R. 2021. Automatic Data Augmentation for Generalization in Reinforcement Learning. *Advances in Neural Information Processing Systems*, 34: 5402–5415.

Schulman, J.; Wolski, F.; Dhariwal, P.; Radford, A.; and Klimov, O. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.

Szegedy, C.; Zaremba, W.; Sutskever, I.; Bruna, J.; Erhan, D.; Goodfellow, I.; and Fergus, R. 2014. Intriguing properties of neural networks. In *2nd International Conference on Learning Representations, ICLR 2014*.

Volpi, R.; Namkoong, H.; Sener, O.; Duchi, J. C.; Murino, V.; and Savarese, S. 2018. Generalizing to unseen domains via adversarial data augmentation. *Advances in neural information processing systems*, 31.

Wang, K.; Kang, B.; Shao, J.; and Feng, J. 2020. Improving generalization in reinforcement learning with mixture regularization. *Advances in Neural Information Processing Systems*, 33: 7968–7978.

Xu, Q.; Zhang, R.; Zhang, Y.; Wang, Y.; and Tian, Q. 2021. A fourier-based framework for domain generalization. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 14383–14392.

Yang, Y.; Lao, D.; Sundaramoorthi, G.; and Soatto, S. 2020. Phase consistent ecological domain adaptation. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 9011–9020.

Yarats, D.; Kostrikov, I.; and Fergus, R. 2021. Image augmentation is all you need: Regularizing deep reinforcement learning from pixels. In *International conference on learning representations*.

Zhang, H.; Cisse, M.; Dauphin, Y. N.; and Lopez-Paz, D. 2018. mixup: Beyond Empirical Risk Minimization. In *International Conference on Learning Representations*.

Zhang, H.; and Guo, Y. 2022. Generalization of Reinforcement Learning with Policy-Aware Adversarial Data Augmentation. In *Decision Awareness in Reinforcement Learning Workshop at ICML 2022*.

Zhang, S.; Su, S.; Li, L.; Zhou, Q.; Lu, J.; and Chang, C.-C. 2019. An image style transfer network using multilevel noise encoding and its application in coverless steganography. *Symmetry*, 11(9): 1152.

Zhong, Z.; Zhao, Y.; Lee, G. H.; and Sebe, N. 2022. Adversarial style augmentation for domain generalized urban-scene segmentation. *Advances in Neural Information Processing Systems*, 35: 338–350.

Zhou, K.; Liu, Z.; Qiao, Y.; Xiang, T.; and Loy, C. C. 2022. Domain generalization: A survey. *IEEE Transactions on Pattern Analysis and Machine Intelligence*.

Zhou, K.; Yang, Y.; Qiao, Y.; and Xiang, T. 2021. Domain Generalization with MixStyle. In *International Conference on Learning Representations 2021*.

Zhou, Z.; Qi, L.; and Shi, Y. 2022. Generalizable medical image segmentation via random amplitude mixup and domain-specific image restoration. In *European Conference on Computer Vision*, 420–436. Springer.

Zhu, Y.; Mottaghi, R.; Kolve, E.; Lim, J. J.; Gupta, A.; Fei-Fei, L.; and Farhadi, A. 2017. Target-driven visual navigation in indoor scenes using deep reinforcement learning. In *2017 IEEE international conference on robotics and automation (ICRA)*, 3357–3364. IEEE.