Improving Deepfake Detection with Reinforcement Learning-Based Adaptive Data Augmentation

Yuxuan Chou[1,2*], **Tao Yu**[3], **Wen Huang**[1], **ZhangYuHeng**[1], **Tao Dai**[4†], **Shu-Tao Xia**[1,2]

[1]Tsinghua Shenzhen International Graduate School, Tsinghua University
[2]Pengcheng Laboratory
[3]Institute of Automation, Chinese Academy of Sciences
[4]College of Computer Science and Software Engineering, Shenzhen University

Abstract

The generalization capability of deepfake detectors is critical for real-world use. Data augmentation via synthetic fake face generation effectively enhances generalization, yet current SoTA methods rely on fixed strategies—raising a key question: *Is a single static augmentation sufficient, or does the diversity of forgery features demand dynamic approaches?* We argue existing methods overlook the evolving complexity of real-world forgeries (e.g., facial warping, expression manipulation), which fixed policies cannot fully simulate.To address this, we propose CRDA (Curriculum Reinforcement-Learning Data Augmentation), a novel framework guiding detectors to progressively master multi-domain forgery features from simple to complex. CRDA synthesizes augmented samples via a configurable pool of forgery operations and dynamically generates adversarial samples tailored to the detector's current learning state.Central to our approach is integrating reinforcement learning (RL) and causal inference. An RL agent dynamically selects augmentation actions based on detector performance to efficiently explore the vast augmentation space, adapting to increasingly challenging forgeries. Simultaneously, the agent introduces action space variations to generate heterogeneous forgery patterns, guided by causal inference to mitigate spurious correlations—suppressing task-irrelevant biases and focusing on causally invariant features. This integration ensures robust generalization by decoupling synthetic augmentation patterns from the model's learned representations.Extensive experiments show our method significantly improves detector generalizability, outperforming SOTA methods across multiple cross-domain datasets.

Introduction

Deepfake technology, which generates highly realistic visual content, has drawn significant attention but is frequently exploited for malicious purposes such as disinformation dissemination and fraud, posing severe societal harms (Zhang et al. 2025; Wang et al. 2025). Thus, developing robust and reliable deepfake detection systems is imperative.

While existing deepfake detection methods perform well on standardized benchmarks, real-world deployment requires handling forgeries from diverse techniques across heterogeneous environments, making generalization the key metric for robustness. Traditional methods relying solely on raw training data lack generalizability. Each forgery technique has unique artifacts, and over-optimizing for specific manipulations leads to bias. Synthetic face generation, as an effective data augmentation strategy, partially replicates real forgery processes, preserving and amplifying manipulation artifacts. However, current methods use fixed numbers of synthetic faces and static augmentation strategies during training.

This paper investigates a critical question in deepfake detectors training: **Is a single fixed augmentation strategy sufficient, or does the diversity of forgery features necessitate dynamic adaptation?** Empirical studies (Cheng et al. 2024) reveal that mixing synthetic samples from conflicting augmentation strategies—such as spatial-domain blurring and frequency-domain noise—often degrades performance. For instance, spatial-domain augmentations may introduce low-frequency blur that interferes with frequency forgery features, leading to suboptimal detection. Consequently, most prior work evaluates single-strategy augmentation in isolation. However, this approach contradicts the expectation that detectors trained on diverse forgery patterns should learn richer, more robust features. We identify the root cause as the absence of coordinated strategy scheduling, which induces incompatible feature conflicts across augmentation domains. Different augmentations inherently embed domain-specific artifacts (e.g., spatial warping patterns in FaceSwap (Kowalski 2018) vs. spectral inconsistencies in FreqBlender (Li et al. 2024)). Naively mixing these strategies during training will force detectors to develop competing feature preferences—oscillating between overfitting to dominant artifacts while neglecting causally stable forgery footprints. This bias becomes catastrophic when encountering unseen forgeries.

To address these limitations, we propose Curriculum Reinforcement-Learning Data Augmentation (CRDA), an innovative framework integrating curriculum learning with adversarial reinforcement learning. CRDA's core idea is to guide the detector through a progressive learning process by designing increasingly difficult forged samples and optimizing strategies via dynamic feedback, augmented by causal learning theory to suppress task-irrelevant spurious correlations. It consists of three key components: 1) **RL-based dynamic augmentation strategy:** the agent dynamically se-

*zhouyuxuan25@mails.tsinghua.edu.cn
†Corresponding author
Copyright © 2026, Association for the Advancement of Artificial Intelligence (www.aaai.org). All rights reserved.

lects optimal, challenging data augmentation methods based on real-time model state perception, optimizing via multi-objective rewards. This forces the detector to learn universal features through the continuous evolution of forged samples in an adversarial game. 2) **Multi-environment causal invariance learning:** leveraging causal inference, we analyze bias sources from data augmentation feature conflicts, introducing Invariant Risk Minimization (IRM) for training on biased data. Auxiliary environments are constructed using historical entropy from the PolicyNet. 3) **Multi-dimensional curriculum scheduling:** the curriculum strategy is extended to three decoupled dimensions—controlling the proportion of augmented synthetic faces, regulating RL exploration intensity, and adjusting forgery region scale/area—all scheduled by training phase.

Briefly, the main contributions of this work are summarized as follows:

- To our best knowledge, CRDA is the first study in deepfake detection to systematically explore and integrate the interaction mechanisms of multiple data augmentations.
- We propose a strategy that combines curriculum learning and reinforcement learning to guide the detector in progressively mastering multi-domain forgery identification skills from simple to complex.
- We apply causal inference theory to deepfake detection, conduct an in-depth analysis of potential bias sources from the perspective of data augmentation feature conflicts, and propose corresponding solutions.

Related Works

Data Augmentation Towards Deepfake Detectors Generalization

Deepfake detection (Yan et al. 2023b; Kaur et al. 2024; Heidari et al. 2024; Nguyen et al. 2024; Ba et al. 2024) techniques identify manipulated images by analyzing discrepancies between real and synthetic facial data across multiple dimensions, including identity features (Zhou and Lim 2021), spatial artifacts (Haliassos et al. 2021; Li et al. 2020a; Zhao et al. 2021), frequency-domain differences (Li et al. 2024; Hasanaath et al. 2025), and architecture-specific patterns (Masi et al. 2020; Bai et al. 2023). Among various strategies, data augmentation (Li and Lyu 2018; Li et al. 2020a; Chen et al. 2022; Cheng et al. 2024; Shiohara 2022; Hasanaath et al. 2025; Li et al. 2024) has become a key approach to enhancing cross-domain robustness. Early approaches such as FWA (Li and Lyu 2018) employed a self-blending strategy through facial region downsampling and spatial warping to simulate deepfake artifacts. Face X-ray (Li et al. 2020a) explicitly trained detectors to identify blending boundaries, while I2G (Zhao et al. 2021) extended this concept with pairwise self-consistency learning to detect intra-image inconsistencies. SLADD (Chen et al. 2022) introduced adversarial training by dynamically generating challenging blending patterns. More recent advancements like SBI (Shiohara 2022) achieve high-fidelity augmentation through intra-identity face swapping, effectively mimicking state-of-the-art manipulation techniques. FSBI (Hasanaath et al. 2025) expands SBI into the frequency domain innovatively, and FreqBlender (Li et al. 2024) uses a specialized frequency parsing network to synthesize fake faces by leveraging inherent correlations among frequency knowledge.

Reinforcement Learning

Reinforcement learning (RL) (Kaelbling, Littman, and Moore 1996) is a machine learning category where an agent learns an optimal policy through trial-and-error interaction with the environment to maximize long-term cumulative rewards. It has proven effective in data augmentation in several tasks (Cubuk et al. 2019; Zhang et al. 2019; Liu et al. 2023; Reed et al. 2021): AutoAugment (Cubuk et al. 2019) by Google Brain uses RL to automatically search for optimal augmentation policies on the validation set, optimizing classification accuracy with strong transferability. Liu et al.(Liu et al. 2023) applied this to medical image segmentation, improving accuracy via Adaptive Sequence-length based Deep Reinforcement Learning. However, most RL-based augmentation schemes focus on general tasks, while ours is specifically designed for deepfake detection.

Causal Inference

Causal inference (Pearl 2010) is a methodology that determines how one variable directly affects another, and quantifies such effects. Extensive research (Arjovsky et al. 2019; Lin et al. 2022; Lv et al. 2022) exists on eliminating dataset bias. Jones et al.(Jones et al. 2023) used causal inference to examine medical image biases, proposing three mechanisms for fairness. Arjovsky et al.(Arjovsky et al. 2019) derived feature invariance from causality and introduced Invariant Risk Minimization (IRM) to constrain models to learn stable correlations, addressing reliance on data biases. Lin et al.(Lin et al. 2022) explored environmental partitioning without information loss. We are the first to introduce causal inference to deepfake detection, guiding the model to learn unbiased multi-domain features (Zha et al. 2024b,a, 2025).

Methodology

Given the limited integration of diverse data augmentations in prior studies, our work aims to develop an effective framework capable of comprehensively and equitably learning multi-domain forgery information from various data augmentation approaches. **The overall architecture is illustrated in Figure 1 for the RL PolicyNet training methodology and Figure 2 for the detector training pipeline.** The following sections will provide detailed technical descriptions of our proposed methodology through its constituent modules.

RL-based Dynamic Augmentation Strategy

We start this section with a question: *What are the advantages of our RL-based data augmentation?* Previous work CDFA(Lin et al. 2024b) first used various data augmentation strategies to train deepfake detectors. It introduced a dynamic forgery search strategy (DFS) to optimize the strategy network, dynamically selecting forgery augmentation operations based on the validation set's performance and achieving satisfactory results. However, DFS's reliance on the validation

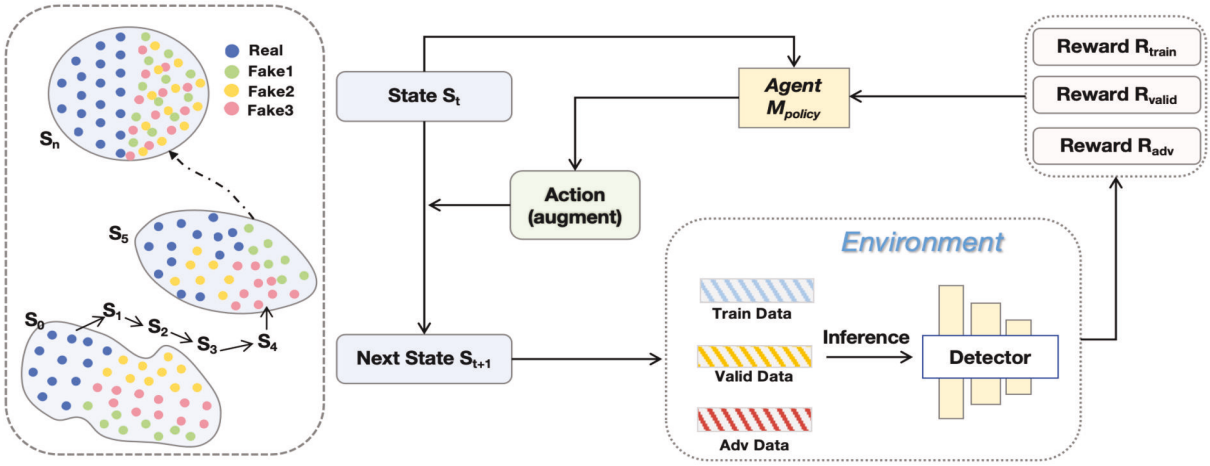

Figure 1: **PolicyNet Optimization via Multi-Rewards.** We train the RL PolicyNet using a multi-scale reward function (as shown in right panel) to guide the detector to learn features from easy to difficult. This strategy arranges different types of fake samples in a type-agnostic manner (as shown in left panel) within the detector's latent space S,thereby improving its generalization.

set means the selected operations are a local optimum for that specific set. Moreover, its augmentation strategies are all variants of spatial blending operations, and the repetitive features generated limit the detector's generalization capability. In contrast, RL can explore a broader range of optimal strategies through interaction with the environment. We expanded the data augmentation operations and designed three reward functions corresponding to different levels of forgery features, using their gradual integration to guide the detector in mastering multi-domain characteristics step by step. As shown in Figure 1, the core elements are defined as follows:

State. The state space of this RL is the latent space of the detector F. For each episode, the initial state s_0 is a k-dimensional standard normal random vector: $s_0 \sim \mathcal{N}_k(\mathbf{0}, \mathbf{1}), s_0 \in \mathbb{R}^k$, where k is the dimension of latent space. At every step t, the state s_t is updated by the action a_t.

Action. The action space is the output space of the PolicyNet P, a set of data augmentation operations representing the PolicyNet's guidance on the detector F. In this paper, the action space $\mathcal{A} = \{a_j\}_{j=1}^{7}$ corresponds to seven augmentation strategies by default:

$$\Pi_\theta(a|s_t) = \text{Softmax}(\mathbf{W}_2\text{ReLU}(\mathbf{W}_1 s_t + \mathbf{b}_1) + \mathbf{b}_2) \quad (1)$$

The PolicyNet uses weight matrices $\mathbf{W}_1, \mathbf{W}_2$ and bias terms $\mathbf{b}_1, \mathbf{b}_2$ to transform the latent state s_t into action probabilities, with the softmax function converting outputs into probabilities for the seven strategies. These include four benchmark forgery methods from FaceForensics++ (Dang et al. 2020): Deepfakes (Deepfakes 2019), Face2Face (Thies et al. 2016), FaceSwap (Kowalski 2018), NeuralTextures (Thies, Zollhöfer, and Nießner 2019); the advanced SBI (Shiohara 2022) method; and frequency-domain augmentations FSBI (Hasanaath et al. 2025) and FreqBlender (Li et al. 2024).

Reward. After state update via an action, the agent receives a reward to produce better actions. To balance training efficiency and generalization, the reward guides the model to learn basic features early, challenging features in the middle, and encourage exploration later, comprising:

1. **Training Stability Term** ($\lambda_1\mathbb{E}[1-\mathcal{C}_{tr}]$): $\mathbb{E}[1-\mathcal{C}_{tr}] = \frac{1}{N}\sum_{i=1}^{N}(1-|f_\theta(x_i)-y_i|)$ Encourages confidence in correct predictions while preventing overfitting, aiding the detector in rapidly establishing capabilities early.

2. **Validation Performance Improvement** ($\lambda_2\Delta\text{AUC}_{val}$): $\Delta\text{AUC}_{val} = \text{AUC}_{val}^{(t)} - \text{AUC}_{val}^{(t-1)}$ Measures incremental AUC improvement on the validation set (composed of complex augmented fake faces) to drive generalization.

3. **Adversarial Deception Metric** ($\lambda_3\mathcal{C}_{adv}$): $\mathcal{C}_{adv} = \mathbb{E}_{x\sim\mathcal{D}_{real}}[\sigma(g_\beta(\sum_{j=1}^{|\mathbb{T}|}\Pi_\theta(a_j|s_t)\cdot f_\alpha(\tau_j(x))))_{y=0}]$. Here, $\mathbb{T}$ is the set of augmentation strategies; $\Pi_\theta(a_j|s_t)$ is the selection probability of strategy τ_j; $\tau_j(x)$ applies the j-th operation to real sample x; g_β is the detector's classification head; f_α extracts features from $\tau_j(x)$. Weighted summed features input to g_β assess misclassification probability, guiding selection of adversarial strategies to construct a minimax game and expand the detector's knowledge boundary. Additionally, a regularization term ($-\lambda_4\text{KL}(\Pi_t\|\Pi_{t-1})$) enforces policy consistency:

$$\text{KL}(\Pi_t\|\Pi_{t-1}) = \mathbb{E}_{s\sim\mathcal{D}}[\Pi_t \log\frac{\Pi_t}{\Pi_{t-1}}] \quad (2)$$

With λ varying by training phase, overall reward function is:

$$\begin{aligned} r_t = \lambda_1\mathbb{E}[1-\mathcal{C}_{tr}] + \lambda_2\Delta\text{AUC}_{val} \\ + \lambda_3\mathcal{C}_{adv} - \lambda_4\text{KL}(\Pi_t|\Pi_{t-1}) \end{aligned} \quad (3)$$

Given significant distributional differences among augmented forged samples, the environment is non-stationary. We use the PPO-Clip algorithm (Schulman et al. 2017) for policy optimization, which enforces hard clipping to limit policy updates (ensuring stability) and incorporates importance sampling with multiple updates to reduce sample demand and accelerate convergence.

Multi-environment Causal Invariance Learning

Guided by the causal theory, we create diverse environments based on the policy entropy and utilize Invariant Risk Min-

imization (IRM) to guide the detector in learning unbiased forgery features.

Causal Analysis of Data Augmentation Bias. In deepfake detection, data augmentation essentially constitutes human intervention in the causal mechanisms of data generation. We model augmentation operators as intervention variables A, which are governed by the data construction strategy $M(M \rightarrow A)$. From a causal perspective, features are categorized into causal features X_v and spurious features X_s.

Policy Entropy-driven Environment Variation. Building on the above causal analysis, we propose a policy entropy-driven environment partitioning mechanism that explicitly distinguishes the decision confidence of different data augmentation strategies. This mechanism constructs training environments with significant conditional distribution discrepancies while preserving causally consistent cross-domain features. Confidence divergence in augmentation strategies across environments forces the model to decouple superficial correlations tied to specific tactics (e.g., blending-induced blur artifacts), enhancing its ability to capture environment-stable causal features. This environmental contrast fundamentally overcomes the feature disentanglement limitations of conventional augmentation. For a training batch with N samples, the environment construction process is as follows: $\mathcal{H}_t = -\sum_{i=1}^{N}\sum_{j=1}^{K}\Pi_\phi(a_j|s_i)\log\Pi_\phi(a_j|s_i)$ where K denotes the number of augmentation strategies. The environment construction consists of two core parts:Dominant environment: Constructed by selecting samples with the absolute minimum entropy values across the entire batch to create a low-uncertainty environment: $\mathcal{E}_d^{(t)} = \{x_{i^*} \mid i^* = \text{argmin}_{1\le i\le N}\,\mathcal{H}_t(i)\}$ Adversarial environment ensemble: Employing quantile partitioning to capture regions of high uncertainty.

$$\begin{aligned}
\mathcal{E}^1_{\text{adv}} &= \{x_i \mid \mathcal{H}_t(i) \in [0.75, 1.0]\cdot\mathcal{H}_t^{\max}\},\\
\mathcal{E}^2_{\text{adv}} &= \{x_i \mid \mathcal{H}_t(i) \in [0.50, 0.75)\cdot\mathcal{H}_t^{\max}\},\\
\mathcal{E}^3_{\text{adv}} &= \{x_i \mid \mathcal{H}_t(i) \in [0.25, 0.50)\cdot\mathcal{H}_t^{\max}\}.
\end{aligned} \tag{4}$$

We also introduce a memory mechanism that maintains a first-in-first-out historical queue $Q_m$for each environment m, with capacity equal to the batch size:

$$\begin{aligned}
Q_m^{(t)} &= \text{FIFO}(Q_m^{(t-1)} \cup \mathcal{E}_m^{(t)} C)\\
\text{where}\quad \text{FIFO}(S,C) &\triangle \begin{cases} S, \text{if } |S| \le C \\ S_{|S|-C+1:|S|}, \text{if } |S| > C \end{cases}
\end{aligned} \tag{5}$$

Here, $Q_m^{(t)}$ is the historical queue for the m-th environment at time step t, $\mathcal{E}_m^{(t)}$ is the set of samples selected at time step t, and C is the capacity of the queue. The FIFO mechanism ensures that the queue's capacity does not exceed C, and when the queue is full, the oldest sample is removed.

Invariant Risk Minimization. Building upon the aforementioned environment variation, we formulate an enhanced IRM objective that adapts to dynamically evolving training environments. The core innovation is the establishment of a bi-level optimization between environment-aware feature learning and augmentation policy adaptation:

$$\begin{aligned}
\min_\theta \sum_{m=1}^{M} w_m \mathbb{E}(x,y) \sim Q_m[\mathcal{L}(f\theta(x),y)]\\
+ \Omega|\nabla_{w|w=1}\mathcal{L}m(w\circ f\theta)|^2
\end{aligned} \tag{6}$$

where θ denotes the parameters of the shared backbone and domain bias classification heads. The backbone extracts features from input data, while the domain bias classification heads predict domain-specific biases. M represents the number of environments; Q_m stands for the historical queue of the m-th environment; $\mathcal{L}$ is the loss function; and Ω is the regularization weight that balances the main loss and gradient penalty. Additionally, w_m denotes the importance weight of environment m, calculated as:

$$\begin{aligned}
w_m &= \frac{\exp(-\bar{\mathcal{H}}_m)}{\sum_{k=1}^{M}\exp(-\bar{\mathcal{H}}_k)},\\
\text{where}\quad \bar{\mathcal{H}}_m &= \mathbb{E}_{x\sim Q_m}\left[\mathcal{H}(\Pi_\phi(a|x))\right].
\end{aligned} \tag{7}$$

Multi-dimensional Curriculum Scheduling

Data Course: The proportion of augmented forged samples reflects the complexity of forgery features. In the early training stage, as the model has not yet mastered basic forgery identification, introducing a large number of advanced samples with complex features may slow down convergence due to the difficulty in processing such samples. Thus, we gradually incorporate augmented samples and design a sine-based proportion scheduling function.

$$q(t) = 0.5 + 0.5\cdot\max(\min(\sin(\pi\cdot\frac{t-\tau/4}{\tau/2}),1),-1) \tag{8}$$

τ denotes total epochs, and t represents the current epoch.

Exploration Course: Entropy regularization incorporates an entropy regularization term into the reinforcement learning objective function to balance the exploration-exploitation trade-off in RL policies. Herein, a dynamic entropy regularization mechanism is proposed. $\beta(t) = \beta_{\max}\cdot[\sigma(k(\frac{t}{\tau}-\mu)) - \sigma(-k\mu)]$ Here, σ is the sigmoid function, $k = 5$ controls the steepness of the transition, and $\mu = 0.3$ sets the peak phase of exploration intensity. The function maintains low exploration intensity in the early stage, linearly increases it in the middle stage to promote exploration of new strategies, and automatically anneals in later stage to stabilize the policy.

Region Course: We propose a facial curriculum template learning method based on dynamic region sampling. The method first constructs a region pool composed of 15 regions based on four basic facial organs (left eye, right eye, nose, mouth) and their combinations using facial landmark technology. It then dynamically controls the proportion of forgery region area using the exponential decay function: $A(t) = A_{\text{full}}e^{-\lambda t} + A_{\min}$ where $A_{\text{full}} = 1.0$, $A_{\min} = 0.3$, and $\lambda = 2/\tau$. During training, at each time step t, candidate regions are selected from the region pool according to the current target area $A(t)$, and then sampled randomly with Gaussian weights: $p_i \propto \exp(-\frac{(A_i - A(t))^2}{2\sigma^2})$ This allows

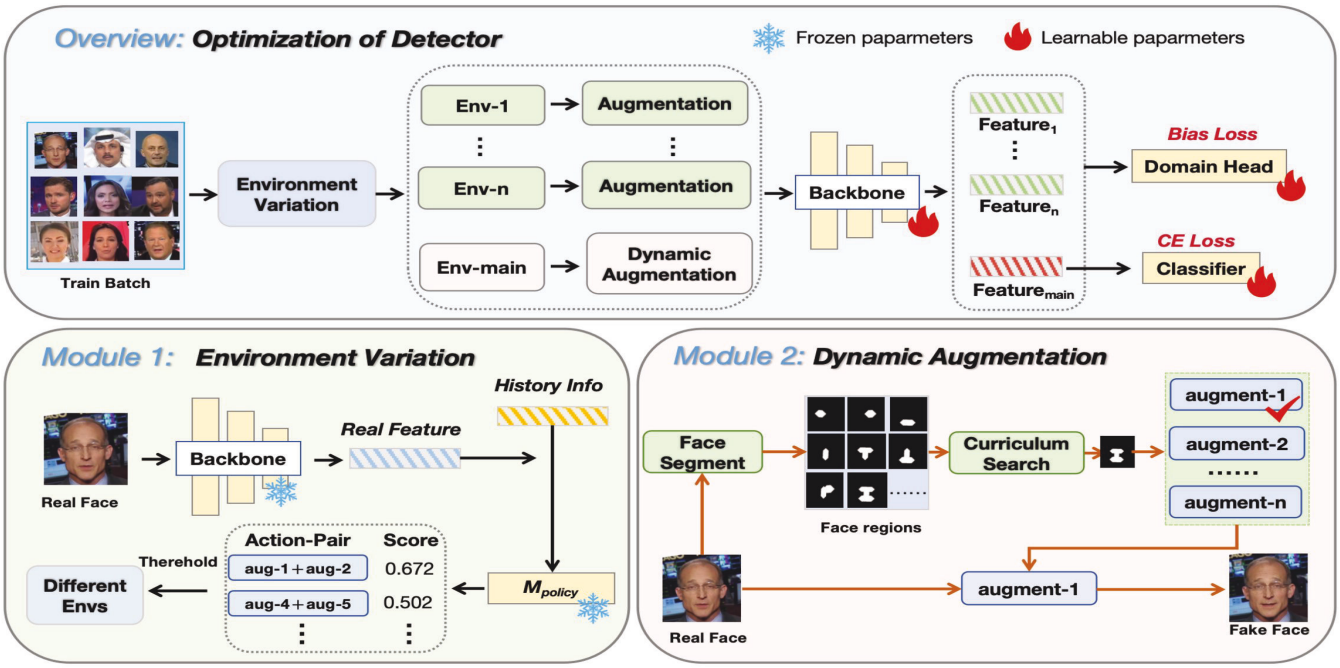

Figure 2: **Training process of CRDA detector:** In each round of optimization, CRDA generates several environments through Module 1: Environment Variation based on the PolicyNet. The main environment and auxiliary environments, after different data augmentations, optimize the detector via $\mathcal{L}_{\text{CE}}$ and $\mathcal{L}_{\text{bias}}$ losses respectively. In this process, the PolicyNet remains frozen.

the forgery region to gradually transition from multi-organ combinations in the early stages ($t < 0.3\tau$, on average, 3.2 organs) to single-organ regions in the later stages.

Loss Function

The overall loss function comprises two key components, designed to optimize both classification performance and feature invariance: $\mathcal{L}_{\text{total}} = \mathcal{L}_{\text{CE}} + \gamma\mathcal{L}_{\text{bias}}$.

Cross-Entropy Loss ($\mathcal{L}_{\text{CE}}$): As the primary classification loss, it is computed on augmented samples:

$$\mathcal{L}_{\text{CE}} = -\frac{1}{N}\sum_{i=1}^{N}[y_i \log \sigma(f_\theta(x_i)) + (1-y_i)\log(1-\sigma(f_\theta(x_i)))] \tag{9}$$

where x_i denotes augmented samples, y_i represents ground-truth labels, f_θ is the detector output and σ stands for the sigmoid activation function.

Bias Loss ($\mathcal{L}_{\text{bias}}$): This is the invariant learning regularization term derived from our enhanced IRM framework (see Section 'Invariant Risk Minimization' for derivation details). The hyperparameter γ balances the learning of discriminative features and the maintenance of feature invariance across different augmentation strategies.

Experiments

Experimental Setting

Datasets. Experiments use common deepfake datasets: FaceForensics++ (FF++)(Rossler et al. 2019), Celeb-DF-v1/v2 (CDFv1/v2)(Li et al. 2020b), DeepFake Detection Challenge (DFDC)/its preview (DFDCP)(Dolhansky et al. 2020), and UADFV (Liy and InIctuOculi 2018). FF++ has 1000 pristine and 4000 manipulated videos (4 techniques) with 3 compression levels; we use c23. CDF contains 590 pristine and 5639 fake videos. DFDC has 100,000 clips, DFDCP as its preview. UADFV includes 45 deepfake and 45 pristine videos. Original training/testing splits are adopted.

Implementation Details. For preprocessing and training, we followed the official code and settings from (Yan et al. 2023b) for a fair comparison. We used EfficientNetB4 (Tan and Le 2019) as the detector's backbone. The PolicyNet consists of two convolutional layers and three MLPs with ReLU activation, using softmax to output a probability distribution over seven augmentation strategies. The domain head in the causal loss is identical to the detector's classifier. Reward coefficients (λ) were initialized to 0.6, 0.2, 0.1, and 0.1, and the causal regularization coefficient (γ) was 0.5. The detector was trained with Adam (Kingma and Ba 2014) (lr = 1×10^{-4}, weight decay = 5×10^{-4}, 30 epochs). The PPO (Schulman et al. 2017) algorithm used a learning rate of 3×10^{-5}, a GAE coefficient of 0.8, and a discount factor of 0.95. All experiments were conducted on eight NVIDIA Tesla V100 GPUs.

Overall Performance on Comprehensive Datasets

To validate the effectiveness of our method, we trained it on the FF++ dataset and evaluated its performance on five other cross-domain datasets. In this study, we primarily report the area under the ROC curve (AUC) to facilitate compar-

Method	Venues	FF++	CDFv1	CDFv2	DFDC	DFDCP	UADFV	Avg.
Xception (Chollet 2017)	CVPR'17	0.9637	0.7794	0.7365	0.7077	0.7374	0.9379	0.7798
Meso4 (Afchar et al. 2018)	WIFS'18	0.6077	0.7358	0.6091	0.5560	0.5994	0.7150	0.6431
FWA (Li and Lyu 2018)	CVPRW'18	0.8765	0.7897	0.6680	0.6375	0.6132	0.9049	0.7433
EfficientB4 (Tan and Le 2019)	ICML'19	0.9567	0.7909	0.7487	0.6955	0.7283	0.9472	0.7821
Capsule (Nguyen 2019)	ICASSP'19	0.8421	0.7909	0.7472	0.6465	0.6568	0.9078	0.7498
CNN-Aug (Wang et al. 2020)	CVPR'20	0.8493	0.7420	0.7027	0.6361	0.6170	0.8739	0.7143
X-ray (Li et al. 2020a)	CVPR'20	0.9592	0.7093	0.6786	0.6326	0.6942	0.8989	0.7227
FFD (Dang et al. 2020)	CVPR'20	0.9624	0.7840	0.7435	0.7029	0.7426	0.9450	0.7836
F3Net (Qian et al. 2020)	ECCV'20	0.9635	0.7769	0.7352	0.7021	0.7354	0.9347	0.7769
SPSL (Liu et al. 2021)	CVPR'21	0.9610	0.8150	0.7650	0.7040	0.7408	0.9424	0.7934
SRM (Luo et al. 2021)	CVPR'21	0.9576	0.7926	0.7552	0.6995	0.7408	0.9427	0.7862
CORE (Ni et al. 2022)	CVPRW'22	0.9638	0.7798	0.7428	0.7049	0.7341	0.9412	0.7726
Recce (Cao et al. 2022)	CVPR'22	0.9621	0.7677	0.7319	0.7133	0.7419	0.9446	0.7794
SBI (Shiohara 2022)	CVPR'22	0.8176	0.8311	0.8015	0.7139	0.7794	0.9475	0.8147
UCF (Yan et al. 2023a)	ICCV'23	0.9705	0.7793	0.7527	0.7191	0.7594	0.9528	0.7967
F-G (Lin et al. 2024a)	CVPR'24	0.9739	0.7330	0.7016	0.6027	0.6824	0.8768	0.7193
LSDA (Yan et al. 2024)	CVPR'24	0.9482	0.8467	0.8142	0.7203	0.7894	0.9515	0.8244
Prodet (Cheng et al. 2024)	NIPS'24	0.9591	0.8756	0.8420	0.6973	0.7744	0.9539	0.8282
FreqBlender (Li et al. 2024)	NIPS'24	0.8753	0.8928	0.8387	0.7041	0.7692	0.9487	0.8307
CRDA(ours)	-	0.9374	**0.9010** (↑ 2.90%)	**0.8536** (↑ 1.38%)	**0.7429** (↑ 3.14%)	**0.7973** (↑ 1.00%)	**0.9570** (↑ 0.03%)	**0.8504** (↑ 2.68%)

Table 1: Cross-dataset evaluations (AUC) from FF++ (Rossler et al. 2019) (in-dataset) to CDFv1 (Li et al. 2020b), CDFv2 (Li et al. 2020b), DFDC (Dolhansky et al. 2020) , DFDCP (Dolhansky et al. 2020) and UADFV (Liy and InIctuOculi 2018) (cross-dataset). Avg. denotes the average value of cross-dataset results. The best results are highlighted in **bold**. Cross-dataset improvements compared with the previous best one are written in small.

Backbone	**CDFv1**	**CDFv2**	**DFDC**	**DFDCP**	**UADFV**	**Avg.**
ResNet50 + SBI	0.8438	0.8215	0.6862	0.7102	0.8958	0.7915
ResNet50 + FreqBlender	0.8527	0.8372	0.7016	0.7385	0.9124	0.8085
ResNet50 + Ours	0.8384	0.8301	0.6748	0.7159	0.8947	0.7908
Xception + SBI	0.8427	0.8132	0.7069	0.7614	0.9318	0.8112
Xception + FreqBlender	0.8593	0.8261	0.7142	0.7716	0.9387	0.8220
Xception + Ours	0.8768	0.8425	0.7197	0.7859	0.9521	0.8354
EfficientNet-B4 + SBI	0.8311	0.8015	0.7139	0.7794	0.9475	0.8147
EfficientNet-B4 + FreqBlender	0.8928	0.8387	0.7041	0.7692	0.9487	0.8307
EfficientNet-B4 + Ours	0.9010	0.8536	0.7429	0.7973	0.9570	0.8504
Swin-Transformer + SBI	0.8793	0.8427	0.7196	0.7984	0.9347	0.8250
Swin-Transformer + FreqBlender	0.9012	0.8463	0.7248	0.8087	0.9395	0.8441
Swin-Transformer + Ours	0.9114	0.8488	0.7352	0.8264	0.9476	0.8539

Table 2: Performance comparison across different backbone networks (AUC scores). The results show that our method can achieve performance improvements on various backbones except ResNet50 — highlights its adaptability.

isons with prior works. As shown in Table 1, we conducted frame-level comparisons with 18 state-of-the-art methods, all evaluated under the same experimental settings as ours. EfficientB4 serves as the baseline using only deepfake data, while SBI (Shiohara 2022) and FreqBlender (Li et al. 2024) are baselines using a single type of blendfake data. By progressively combining diverse forgery augmentation strategies, our method achieved the best performance across all datasets.

Ablation Study On CRDA's Generalizability

Results in video-level: Table 4 shows the video-level results obtained by averaging the prediction results of each frame in the evaluation video, and evaluates them using the AUC metric. Compared with SOTA video-level methods, CRDA achieves an improvement of above 5% in cross-dataset generalization, demonstrating its superior capacity in temporal generalization while maintaining frame-level performance.

Spatial-Domain	Freq-Domain	CDFv1	CDFv2	DFDC	DFDCP	UADFV	Avg.
base_spatial	FSBI	0.8573	0.8451	0.7237	0.7428	0.9349	0.8209
base_spatial + SBI	FSBI	0.8729	0.8493	0.7458	0.7846	0.9374	0.8380
base_spatial + SBI	FreqBlender	0.8842	0.8425	0.7393	0.8016	0.9443	0.8424
standard (both domains)	-	0.9010	0.8536	0.7429	0.7973	0.9570	0.8504
base_spatial+SBI+X-ray+PCL+I2G	FreqBlender+FSBI	0.8937	0.8562	0.7258	0.8073	0.9615	0.8489

Table 3: Ablation study on selection of augmentation strategy. More augmentations do not necessarily lead to better performance; future research should focus on improving the quality of augmentation schemes.

Method	CDFv2	DFDC
Two-branch (ECCV'2020)	76.7	–
Face X-ray (CVPR'2020)	79.5	65.5
LipForensics (CVPR'2021)	82.4	73.5
PCL+I2G (ICCV'2021)	90.0	67.5
FTCN (ICCV'2021)	86.9	74.0
HCIL (ECCV'2022)	79.0	69.2
AUNet (CVPR'2023)	92.8	73.8
AltFreezing (CVPR'2023)	89.5	–
EffcientNetB4 + Ours	**95.22**	**82.64**

Table 4: Performance comparison on video-level. CRDA surpasses specialized approaches by a margin of over 5%.

ID	Components			AUC(%)	
	RL-DA	CIL	CS	CelebDF-v2	DFDC
1	×	×	×	74.87	69.55
2	×	×	✓	76.34	70.22
3	✓	×	✓	81.37	72.81
4	×	✓	✓	80.95	73.16
5	✓	✓	✓	85.36	74.29

Table 5: Ablation study. 'RL-DA' refers to our RL-based data augmentations. 'CIL' refers to Multi-environment Causal Invariance Learning. 'CS' refers to Curriculum Scheduling.

Effect on different backbones: We evaluate CRDA on various backbones (ResNet-50, Xception, EfficientNet-B4, Swin-Transformer) against SBI and FreqBlender. The results in Table 2 confirm its model-agnostic nature, with consistent gains on Xception (+1.34%), EfficientNet-B4 (+1.97%), and Swin-Transformer (+1.22%). The minor performance loss on ResNet-50 is likely because its weaker feature extraction limits the effectiveness of our method's operations.

Selection of augmentation strategy: We analyze various spatial and frequency-domain augmentations as RL actions (Table 3). Results show that increasing the quantity of augmentations beyond a standard set yields negligible performance gains. This highlights that our method's performance is tied to the quality of the augmentation scheme, not its quantity. Consequently, our method is scalable and will benefit from future advancements in higher-quality augmentations.

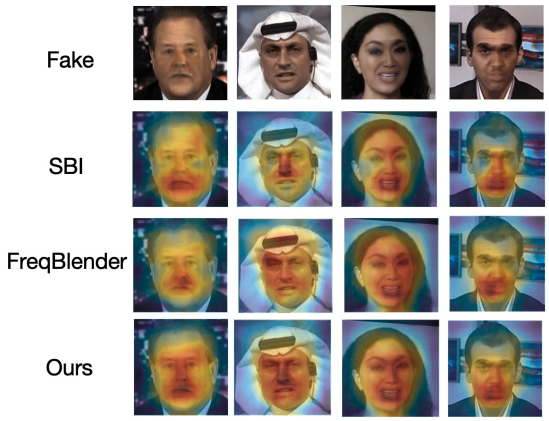

Figure 3: Grad-CAM visualization of SBI, FreqBlender and CRDA. Compared with baselines, CRDA focuses more on the manipulated structural boundaries.

Saliency Visualization: We employ Grad-CAM (Chattopadhay et al. 2018) to visualize the attention of our method compared to SBI and FreqBlender on four manipulations in the FF++ dataset. CRDA pays less attention to irrelevant regions, with enhanced focus on facial regions.

Ablation Study On Key Componments

To better understand the contribution of each module, we conduct an ablation study on CRDA by decomposing it into three components: RL-DA (Reinforcement Learning-based Data Augmentation), CIL (Multi-environment Causal Invariance Learning), and CS (Curriculum Scheduling). The experimental setup is detailed in Table 5. While the CS module contributes the least individually, its combination with other modules significantly boosts performance. The synergistic effect of all three components yields the best results.

Conclusion

In this work, we introduce CRDA: an RL-driven data augmentation strategy that significantly improves detector performance by generating a diversified training environment. A key trade-off of CRDA is the increased computational cost, and a limitation is the lack of a quantitative method to guide the selection of base augmentation strategies. Our future work will address these issues by optimizing efficiency.

Acknowledgements

This work is supported in part by the National Natural Science Foundation of China under Grants Nos. 62302309 and 62571298, and by the Shenzhen Science and Technology Program under Grant No. JCYJ20220818101014030. This work is also partially supported by the Tsinghua University Shenzhen International Graduate School (Tsinghua SIGS) KA Cooperation Fund.

References

Afchar, D.; Nozick, V.; Yamagishi, J.; and Echizen, I. 2018. Mesonet: a compact facial video forgery detection network. 1–7.

Arjovsky, M.; Bottou, L.; Gulrajani, I.; and Lopez-Paz, D. 2019. Invariant risk minimization. *arXiv preprint arXiv:1907.02893*.

Ba, Z.; Liu, Q.; Liu, Z.; Wu, S.; Lin, F.; Lu, L.; and Ren, K. 2024. Exposing the deception: Uncovering more forgery clues for deepfake detection. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 38, 719–728.

Bai, W.; Liu, Y.; Zhang, Z.; Li, B.; and Hu, W. 2023. AUNet: Learning Relations Between Action Units for Face Forgery Detection. In *CVPR*, 24709–24719.

Cao, J.; Ma, C.; Yao, T.; Chen, S.; Ding, S.; and Yang, X. 2022. End-to-end reconstruction-classification learning for face forgery detection. In *CVPR*, 4113–4122.

Chattopadhay, A.; Sarkar, A.; Howlader, P.; and Balasubramanian, V. N. 2018. Grad-cam++: Generalized gradient-based visual explanations for deep convolutional networks. In *WACV*, 839–847. IEEE.

Chen, L.; Zhang, Y.; Song, Y.; Liu, L.; and Wang, J. 2022. Self-supervised learning of adversarial example: Towards good generalizations for deepfake detection. In *CVPR*, 18710–18719.

Cheng, J.; Yan, Z.; Zhang, Y.; Luo, Y.; Wang, Z.; and Li, C. 2024. Can We Leave Deepfake Data Behind in Training Deepfake Detector? *arXiv preprint arXiv:2408.17052*.

Chollet, F. 2017. Xception: Deep learning with depthwise separable convolutions. In *CVPR*, 1251–1258.

Cubuk, E. D.; Zoph, B.; Mane, D.; Vasudevan, V.; and Le, Q. V. 2019. Autoaugment: Learning augmentation strategies from data. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 113–123.

Dang, H.; Liu, F.; Stehouwer, J.; Liu, X.; and Jain, A. K. 2020. On the detection of digital face manipulation. In *CVPR*.

Deepfakes. 2019. Faceswap. https://github.com/deepfakes/faceswap. Accessed 2025-04-24.

Dolhansky, B.; Bitton, J.; Pflaum, B.; Lu, J.; Howes, R.; Wang, M.; and Ferrer, C. C. 2020. The deepfake detection challenge dataset. *arXiv preprint arXiv:2006.07397*.

Haliassos, A.; Vougioukas, K.; Petridis, S.; and Pantic, M. 2021. Lips Don't Lie: A Generalisable and Robust Approach To Face Forgery Detection. In *CVPR*.

Hasanaath, A. A.; Luqman, H.; Katib, R.; and Anwar, S. 2025. FSBI: Deepfake detection with frequency enhanced self-blended images. *Image and Vision Computing*, 105418.

Heidari, A.; Jafari Navimipour, N.; Dag, H.; and Unal, M. 2024. Deepfake detection using deep learning methods: A systematic and comprehensive review. *Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery*, 14(2): e1520.

Jones, C.; Castro, D. C.; Ribeiro, F. D. S.; Oktay, O.; McCradden, M.; and Glocker, B. 2023. No fair lunch: a causal perspective on dataset bias in machine learning for medical imaging. *arXiv preprint arXiv:2307.16526*.

Kaelbling, L. P.; Littman, M. L.; and Moore, A. W. 1996. Reinforcement learning: A survey. *Journal of artificial intelligence research*, 4: 237–285.

Kaur, A.; Noori Hoshyar, A.; Saikrishna, V.; Firmin, S.; and Xia, F. 2024. Deepfake video detection: challenges and opportunities. *Artificial Intelligence Review*, 57(6): 159.

Kingma, D. P.; and Ba, J. 2014. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*.

Kowalski, M. 2018. FaceSwap. https://github.com/MarekKowalski/FaceSwap. Accessed: 2025-04-24.

Li, H.; Zhou, J.; Li, Y.; Wu, B.; Li, B.; and Dong, J. 2024. FreqBlender: Enhancing DeepFake detection by blending frequency knowledge. *arXiv preprint arXiv:2404.13872*.

Li, L.; Bao, J.; Zhang, T.; Yang, H.; Chen, D.; Wen, F.; and Guo, B. 2020a. Face x-ray for more general face forgery detection. In *CVPR*.

Li, Y.; and Lyu, S. 2018. Exposing deepfake videos by detecting face warping artifacts. *arXiv preprint arXiv:1811.00656*.

Li, Y.; Yang, X.; Sun, P.; Qi, H.; and Lyu, S. 2020b. Celeb-df: A new dataset for deepfake forensics. In *CVPR*.

Lin, L.; He, X.; Ju, Y.; Wang, X.; Ding, F.; and Hu, S. 2024a. Preserving fairness generalization in deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 16815–16825.

Lin, Y.; Song, W.; Li, B.; Li, Y.; Ni, J.; Chen, H.; and Li, Q. 2024b. Fake it till you make it: Curricular dynamic forgery augmentations towards general deepfake detection. In *European Conference on Computer Vision*, 104–122. Springer.

Lin, Y.; Zhu, S.; Tan, L.; and Cui, P. 2022. ZIN: When and how to learn invariance without environment partition? *Advances in Neural Information Processing Systems*, 35: 24529–24542.

Liu, H.; Li, X.; Zhou, W.; Chen, Y.; He, Y.; Xue, H.; Zhang, W.; and Yu, N. 2021. Spatial-phase shallow learning: rethinking face forgery detection in frequency domain. In *CVPR*, 772–781.

Liu, Z.; Lv, Q.; Li, Y.; Yang, Z.; and Shen, L. 2023. Medaugment: Universal automatic data augmentation plug-in for medical image analysis. *arXiv preprint arXiv:2306.17466*.

Liy, C. M.; and InIctuOculi, L. 2018. Exposingaicreated fakevideosbydetectingeyeblinking. In *2018IEEEInterG national Workshop on Information Forensics and Security (WIFS). IEEE*.

Luo, Y.; Zhang, Y.; Yan, J.; and Liu, W. 2021. Generalizing Face Forgery Detection with High-frequency Features. In *CVPR*.

Lv, F.; Liang, J.; Li, S.; Zang, B.; Liu, C. H.; Wang, Z.; and Liu, D. 2022. Causality inspired representation learning for domain generalization. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 8046–8056.

Masi, I.; Killekar, A.; Mascarenhas, R. M.; Gurudatt, S. P.; and AbdAlmageed, W. 2020. Two-branch recurrent network for isolating deepfakes in videos. In *ECCV*.

Nguyen, D.; Mejri, N.; Singh, I. P.; Kuleshova, P.; Astrid, M.; Kacem, A.; Ghorbel, E.; and Aouada, D. 2024. Laanet: Localized artifact attention network for quality-agnostic and generalizable deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 17395–17405.

Nguyen, H. H. 2019. Capsule-forensics: Using capsule networks to detect forged images and videos. In *ICASSP*, 2307–2311. IEEE.

Ni, Y.; Meng, D.; Yu, C.; Quan, C.; Ren, D.; and Zhao, Y. 2022. CORE: Consistent Representation Learning for Face Forgery Detection. In *CVPR Workshop*, 12–21.

Pearl, J. 2010. Causal inference. *Causality: objectives and assessment*, 39–58.

Qian, Y.; Yin, G.; Sheng, L.; Chen, Z.; and Shao, J. 2020. Thinking in frequency: Face forgery detection by mining frequency-aware clues. In *ECCV*, 86–103. Springer.

Reed, C. J.; Metzger, S.; Srinivas, A.; Darrell, T.; and Keutzer, K. 2021. Selfaugment: Automatic augmentation policies for self-supervised learning. In *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, 2674–2683.

Rossler, A.; Cozzolino, D.; Verdoliva, L.; Riess, C.; Thies, J.; and Nießner, M. 2019. Faceforensics++: Learning to detect manipulated facial images. In *ICCV*, 1–11.

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

Shiohara, K. 2022. Detecting deepfakes with self-blended images. In *CVPR*, 18720–18729.

Tan, M.; and Le, Q. 2019. Efficientnet: Rethinking model scaling for convolutional neural networks. In *ICML*, 6105–6114. PMLR.

Thies, J.; Zollhöfer, M.; and Nießner, M. 2019. Deferred neural rendering: Image synthesis using neural textures. *TOG*, 38(4): 1–12.

Thies, J.; Zollhofer, M.; Stamminger, M.; Theobalt, C.; and Nießner, M. 2016. Face2face: Real-time face capture and reenactment of rgb videos. In *CVPR*, 2387–2395.

Wang, S.-Y.; Wang, O.; Zhang, R.; Owens, A.; and Efros, A. A. 2020. CNN-generated images are surprisingly easy to spot... for now. In *CVPR*, 8695–8704.

Wang, Y.; Chou, Y.; Zhou, Z.; Zhang, H.; Wan, W.; Hu, S.; and Li, M. 2025. Breaking barriers in physical-world adversarial examples: Improving robustness and transferability via robust feature. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 39, 8069–8077.

Yan, Z.; Luo, Y.; Lyu, S.; Liu, Q.; and Wu, B. 2024. Transcending forgery specificity with latent space augmentation for generalizable deepfake detection. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, 8984–8994.

Yan, Z.; Zhang, Y.; Fan, Y.; and Wu, B. 2023a. UCF: Uncovering Common Features for Generalizable Deepfake Detection. *ICCV*.

Yan, Z.; Zhang, Y.; Yuan, X.; Lyu, S.; and Wu, B. 2023b. Deepfakebench: A comprehensive benchmark of deepfake detection. *arXiv preprint arXiv:2307.01426*.

Zha, Y.; Dai, T.; Guo, H.; Wang, Y.; Chen, B.; Chen, K.; and Xia, S.-T. 2024a. Point Cloud Mixture-of-Domain-Experts Model for 3D Self-supervised Learning. *arXiv preprint arXiv:2410.09886*.

Zha, Y.; Li, N.; Wang, Y.; Dai, T.; Guo, H.; Chen, B.; Wang, Z.; Ouyang, Z.; and Xia, S.-T. 2024b. Lcm: Locally constrained compact point cloud model for masked point modeling. *Advances in Neural Information Processing Systems*, 37: 104816–104842.

Zha, Y.; Wang, Y.; Guo, H.; Wang, J.; Dai, T.; Chen, B.; Ouyang, Z.; Yuerong, X.; Chen, K.; and Xia, S.-T. 2025. PMA: Towards Parameter-Efficient Point Cloud Understanding via Point Mamba Adapter. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, 16976–16986.

Zhang, X.; Wang, Q.; Zhang, J.; and Zhong, Z. 2019. Adversarial autoaugment. *arXiv preprint arXiv:1912.11188*.

Zhang, Y.; Zhou, Y.; Li, T.; Li, M.; Hu, S.; Luo, W.; and Zhang, L. Y. 2025. Secure transfer learning: Training clean models against backdoor in (both) pre-trained encoders and downstream datasets. *arXiv preprint arXiv:2504.11990*.

Zhao, T.; Xu, X.; Xu, M.; Ding, H.; Xiong, Y.; and Xia, W. 2021. Learning self-consistency for deepfake detection. In *ICCV*, 15023–15033.

Zhou, Y.; and Lim, S.-N. 2021. Joint audio-visual deepfake detection. In *Proceedings of the IEEE/CVF international conference on computer vision*, 14800–14809.