

Channel Effects Limit Adversarial Success under Perfect Channel Knowledge

Jaxon Hancock*, Alice Yu†, Tongtong Li*, and Jian Ren *

* Department of Electrical and Computer Engineering, Michigan State University, East Lansing, MI, USA

Email: {hancoc79, tongli, renjian}@msu.edu

† Agentic.AI Corporation, 527 Howard St, San Francisco, CA 94105

Email: aliceyu0721@gmail.com

Abstract—Adversarial Examples (AE) are a constant threat to machine learning-enabled technology, yet their effectiveness in realistic environments is limited by benign perturbations introduced by environmental factors such as sensor variation and channel effects. In this work, we investigate the possibility of incorporating channel knowledge into generative adversarial network-based methods to produce channel-resistant perturbations. We adapt the AdvGAN framework with a channel-in-the-loop fine-tuning procedure and evaluate its performance against a well-trained classifier under additive Gaussian white noise and luminosity shifts. To quantify adversarial effectiveness beyond channel induced performance changes, we introduce Normalized Surplus Adversarial Success Rate (NS-ASR). Our experiments show that while fine-tuning yields modest gains under strong AWGN (up to a 5.44% improvement in ASR at $\sigma = 0.05$), it provides little benefit under weaker noise or luminosity shifts. Moreover, we show that GAN-based adversaries are unable to benefit from classifier-side performance degradation while open-box baselines succeed, showing a tradeoff between efficiency and robustness.

Index Terms—Machine learning, adversarial examples, generative adversarial networks, fine-tuning

I. INTRODUCTION

As machine learning (ML) systems rapidly transform the fabric of daily life, their theoretical vulnerabilities become genuine security risks. Over the past decade, ML security development has been a cat-and-mouse game between model/defense improvements followed quickly by advancements in adversarial capability. In 2014, it was discovered that tiny changes—often imperceptible to human observers—could reliably fool neural networks, sparking an explosion of new research [1]. Since then, adversarial example (AE) generation methods have evolved into a wide variety of niches; circumventing new defenses and exploiting new vulnerabilities with each improvement in ML technology. The development of ML paradigms has enabled the creation of better classifiers than ever before, but their vulnerability to AEs remains cause for serious concern.

In parallel with these advances, ML-enabled classifiers have been integrated into a variety of extremely sensitive domains, often without major consideration for AEs. The widespread deployment of classifiers in the support of highly important modern technologies such as autonomous/assisted driving and facial recognition systems is coincident with the advent of highly sophisticated AE generation methods designed specifically to disrupt them [2], [3].

Existing adversarial example generation methods make specific assumptions about attacker access and environmental conditions. In the open-box setting, optimization-based attacks such as FGSM, PGD, and the Carlini and Wagner attack have achieved high fooling rates in controlled conditions [4]–[6]. Closed-box approaches, including transfer-based and query-efficient attacks have extended this capability to settings without direct gradient access [7]. Generative approaches have been introduced as an alternative to iterative optimization for AE generation. AdvGAN, for example, trains a generative adversarial network (GAN) to produce perturbations directly from clean inputs, bypassing the computational cost associated with many other methods [8]. Meanwhile, physical-world attacks have demonstrated targeted misclassification in a variety of scenarios, exploiting adversarial effects created through real-world manipulation [9].

One critical issue with respect to AE implementation in the real world is a lack of reproducibility; while AEs are designed to be robust, reality often presents a variety of hurdles to successfully implementing an AE-based attack on realistic systems. One such hurdle is channel effects: benign perturbations due to natural phenomena; for computer vision, a channel effect might be image compression, brightness modulation, or visual noise.

While previous work explored robustness to environmental variation primarily through data augmentation strategies, these approaches typically treat the channel as a small contributor to the system as a whole, rather than a key element to be explicitly modeled and leveraged during attack generation. As a result, perturbations generated without explicit channel modeling may fail to survive the channel and have an adversarial effect. Additionally, channel effects in realistic environments often interact with adversarial perturbations in unexpected or nonlinear ways, making their impact difficult to predict using data augmentation. This motivates an integrated, channel-aware adversarial generation framework that learns to produce perturbations resilient to the impairments present in the target environment, but such integration into existing AE generation frameworks may be nontrivial.

In this work, we implement a channel-aware generative adversarial network to test the ability of generative AE methods to simultaneously generalize to both a benign channel and a well-trained classifier. We go on to evaluate the effectiveness of

channel-aware generator fine-tuning in the task of preparing a well-trained adversary to fool a target classifier under realistic channel effects such as luminosity modulation and Gaussian white noise. Our contributions are as follows:

- We adapt the AdvGAN framework to explicitly incorporate channel knowledge during generator fine-tuning, enabling direct evaluation of GAN-based AEs under realistic environmental channels such as AWGN and luminosity shifts, as well as their transferability to channels of varying perturbation strength.
- We introduce Normalized Surplus Adversarial Success Rate, a novel metric that accounts for both classifier degradation and adversarial effectiveness, isolating the adversary’s proportional contribution beyond channel-induced error.
- Through extensive simulation, we show that channel-aware fine-tuning modestly improves adversarial performance under high-noise AWGN, but offers negligible gains under low-noise or linear transformations.
- We provide a strong optimization-based baseline (PGD with Expectation over Transformation) to contrast GAN-based adversarial success under benign channels, reinforcing a fundamental efficiency-robustness tradeoff.

In Section II, we discuss the related work and define the literature niche addressed by this paper. Next, the preliminaries and assumptions are presented in Section III, where we elaborate on the dataset, classifiers, and metrics we use as the basis of our experiments. Section IV comprises a formalization of our channel effect simulation methods, channel-aware GAN-based AE generation method, and performance metrics. We present our findings in Section V, and we further contextualize these results in Section VI.

II. RELATED WORK

A. GAN-based Adversarial Example Generation

Generative Adversarial Networks (GANs) have been widely used to craft adversarial examples (AE) for image classifiers. Xiao et al. introduced AdvGAN [8], which learns a perturbation generator through adversarial training to produce high quality AEs, itself an extension of the foundational work done by Goodfellow et al. in the introduction of the GAN framework [10]. Subsequent extensions such as AdvGAN++ use feature extraction instead of raw pixel inputs, and recent contributions such as PAR-AdvGAN which incorporate auto-regression to significantly accelerate AE generation [11], [12]. These methods aim to exceed traditional attacks in transferability and perceptual quality. However, all operate on clean images and produce perturbations without accounting for potential degradations. None explicitly model a noisy or luminosity-modulating channel and do not consider the generalization of individual adversarial perturbations to channel effects. Nevertheless, GAN-based attacks have been integrated into attacks spanning numerous real-world applications [13].

B. Channel Effect in Adversarial Attack

The adversarial ML community has explored the robustness to channel-like distortions. Athalye et al. introduced the Expectation over Transformation (EoT) algorithm, which generates AEs that remain effective under sampled distortions, and uses realistic image distortion techniques such as rotation and introduction of noise [14]. This approach represents an example wherein a channel effect itself is used as an AE generation method, rather than evaluating the channel’s effect on the generation of AEs through other methods, such as GANs. Additionally, Yang et al. showed that applying random brightness transformations during iterative attacks can improve their transferability [15]. This approach demonstrates the potential of data augmentation for improving gradient-based attacks, but also does not consider adversarial knowledge of the channel-based augmentation itself. In contrast, research in wireless communication routinely incorporates channel knowledge into AE generation in wireless signal classification. Kim et al. showed that channel-awareness is necessary to successfully craft targeted AEs in fading/lossy wireless channels [16]. We claim that visual noise, brightness changes, and visual modulation are analogous to the channel effects considered in wireless communication: as a channel that a digital perturbation must survive, yet existing image classification attacks rarely account for them explicitly.

III. PRELIMINARY AND ASSUMPTIONS

A. Dataset and Preprocessing

To simulate a computer vision task, we use CIFAR-10 for image classification. CIFAR-10 comprises 60,000 full color labeled images of size $[32 \times 32 \times 3]$; for model training, we partition the original training dataset into a 90% training/testing subset (45,000 samples), and a 10% validation subset (5,000 samples), with 10,000 samples reserved for GAN training and cross-validation. The validation set is used exclusively for monitoring classifier generalization during training and is not used for model updates. We normalize all samples to the range $[0, 1]$ and apply on-the-fly dataset augmentation by using random horizontal flipping and random rotations of 0-10 degrees. To train the GAN, we selected 7,500 correctly-classified samples at random from the disjoint GAN training dataset. To validate the GAN, we selected 500 correctly-classified samples at random from a further disjoint subset as our normative input examples. This ensures the GAN is trained exclusively on high-quality input samples not seen by the classifier and itself validated using input samples seen by neither the classifier nor the GAN.

B. Classifier Architecture and Training Protocol

Our target classifier is ResNet-56, a CNN consisting of 27 ResNet blocks, downsampling, and filtering techniques that has been used as a benchmark target classifier for many image classification adversarial example techniques [8], [17].

We trained the model over a course of 200 epochs with a batch size of 128, an adaptive learning rate decay schedule starting at $lr = 10^{-1}$, and on-the-fly dataset augmentation using

the stochastic gradient descent optimizer with a momentum of 0.9. Checkpointing was utilized to preserve the best-performing model on the validation set, tested at the end of every epoch. We achieved a maximal training accuracy of 0.9999 and loss of 9.3887e-04; our best checkpoint model’s validation set accuracy was 0.9176 with validation loss of 0.5010.

C. Metrics

For untargeted attacks, we define the adversarial success rate (ASR) as:

$$\text{ASR} = \frac{1}{N} \sum_{i=1}^N \mathcal{I}[f(x_i^{\text{adv}}) \neq y_i], \quad (1)$$

where x_i^{adv} is the adversarial example corresponding to the clean input x_i , y_i is the true class label, $f(\cdot)$ is the classifier, and $\mathcal{I}(\cdot)$ is the indicator function, which returns 1 if the condition is true and 0 otherwise.

D. Baseline Expectation-over-Transform Attack

As a reference point, we implemented a projected gradient descent (PGD) attack with Expectation over Transformation (EoT). This method iteratively optimizes perturbations by averaging gradients over multiple channel instances, producing AEs robust to stochastic noise. We used a maximum perturbation budget ϵ of 8/255, a step size of α of 2/255, 10 iterations, 5 EoT samples per step, and followed the standard practice for CIFAR-10 PGD attacks.

IV. PROPOSED MODEL

A. Channel Effects

In realistic deployment scenarios, images rarely reach the classifier unchanged. Environmental and sensor-based effects can drastically alter the pixel distribution of inputs, degrading both clean and adversarial classification performance. Two relevant impairments for computer vision tasks are luminosity modulation, which can occur due to changes in illumination conditions or camera exposure, and additive white Gaussian noise (AWGN), which models sensor noise and low-SNR transmission channels.

Our model for uniform luminosity modulation is:

$$\tilde{x}_{\text{Luminosity}} = \text{clip}(lx, 0, 1), \quad (2)$$

where $l \in \mathbb{R}$ is the luminosity scaling factor and clip enforces valid pixel values. The resultant perturbed sample \hat{x} is clipped to $[0, 1]$ to remain compliant with visual expectations and the bounds of the CIFAR-10 dataset. Our model for uniform AWGN channel effect is as follows:

$$\tilde{x}_{\text{AWGN}} = \text{clip}(x + \mathcal{N}(0, \sigma^2), 0, 1), \quad (3)$$

where $\mathcal{N}(\cdot)$ is the Gaussian noise function centered at 0, controlled by the standard deviation σ .

B. Channel-in-the-Loop Fine Tuning

We adapt the AdvGAN framework to explicitly incorporate channel effects during a pre-inference fine-tuning phase of generator-only training. We go on to compare this fine-tuning technique to training without AWGN or luminosity channel-in-loop fine-tuning to evaluate the relative ability of the GAN to fool the target classifier under noisy conditions, as well as the ability of the target classifier itself to classify inputs in noisy conditions.

Our fine tuning method is described fully in Algorithm 1, wherein our generator G trained to produce adversarial examples for a given classifier f is trained for E epochs to fool the same classifier under a channel \mathcal{C} , either luminosity modulation or AWGN as described above with parameter σ or l . In this process, we made use of the Carlini-Wagner loss, defined as follows:

$$\ell_{\text{CW}}(z, y) = \max(0, z_y - \max_{j \neq y} z_j + \kappa) \quad (4)$$

where

- $z \in \mathbb{R}^K$ are the classifier logits for K classes,
- $y \in \{1, \dots, K\}$ is the true class index,
- z_y is the logit of the true class,
- $\max_{j \neq y} z_j$ is the maximum logit over all incorrect classes, and
- $\kappa \geq 0$ is the confidence margin.

C. Performance Measurement

Given that a benign channel affects both the classifier and GAN, a measurement is needed to describe the performance of the GAN under the condition of a classifier impaired by a similar process; for this purpose, we introduce Normalized Surplus ASR (NS-ASR), implemented as follows:

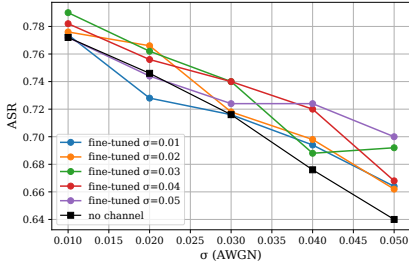
$$\text{NS-ASR} = \frac{\text{ASR}_{\text{adv}} - \text{ASR}_0}{1 - \text{ASR}_0} \quad (5)$$

Where the difference between the ASR of the adversary and benign channel effect $\text{ASR}_{\text{adv}} - \text{ASR}_0$ is normalized with respect to the residual success rate of the classifier under noisy conditions $1 - \text{ASR}_0$. The resultant measurement describes the proportion of the available adversarial headroom wherein the adversary is able to fool the classifier. High NS-ASR indicates the ability of an adversary to fool the classifier beyond the expected success rate disruption effect of the channel, while low NS-ASR indicates an inability of the adversary to do so. Negative NS-ASR is possible in cases wherein the ASR of an adversary is lower than the expected success rate under the benign channel while incorporating said channel.

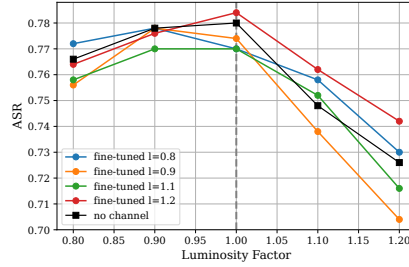
V. SIMULATION RESULTS

A. Channel-Aware Fine-Tuning

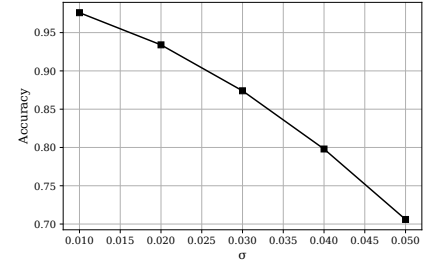
To evaluate the extent to which channel-aware fine-tuning enables GANs to generate channel-resistant perturbations, we tuned a trained AdvGAN generator function over 50 additional epochs for AWGN channels with $\sigma \in 0.01, 0.02, \dots, 0.005$ and luminosity modulation for $l = 0.8, 0.9, \dots, 1.2$ (excluding



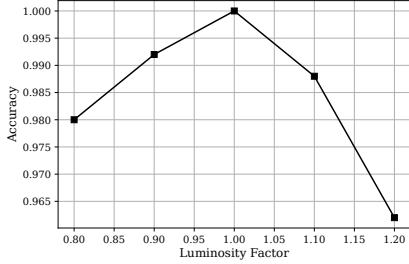
(a) Adversarial Success Rate for fine-tuning and inference $\sigma \in [0.01, 0.05]$



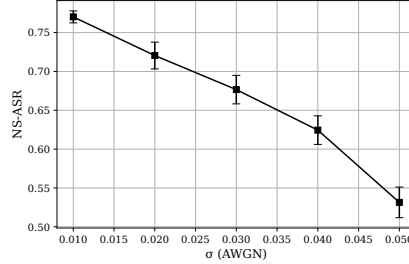
(b) Adversarial Success Rate for fine-tuning and inference $l \in [0.8, 1.2]$



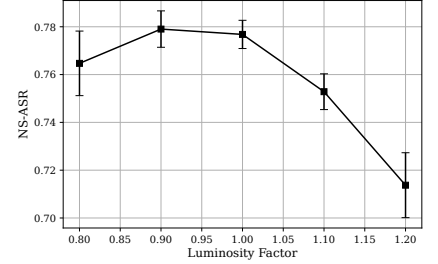
(c) Classifier-only success rate under AWGN channel inference $\sigma \in [0.01, 0.05]$



(d) Classifier-only success rate under luminosity modulation channel inference $l \in [0.8, 1.2]$



(e) NS-ASR under AWGN Channel (all AWGN fine-tuned generators)



(f) NS-ASR under luminosity modulation channel (all luminosity fine-tuned generators)

Fig. 1: Comparative results across AWGN and luminosity modulation channels for fine-tuning, classifier success, and NS-ASR metrics.

Algorithm 1 Channel-Aware Generator Fine-Tuning

Input: Trained generator G , classifier f , optimizer \mathcal{O} , training dataset \mathcal{D} , perturbation budget ϵ , number of epochs E , channel type \mathcal{C} with parameters (σ, l)

```

for  $e = 1$  to  $E$  do
  for each batch  $(x, y)$  in  $\mathcal{D}$  do
    Generate perturbation:  $\delta \leftarrow G(x)$ 
    Clip perturbation:  $\delta \leftarrow \text{clip}(\delta, -\epsilon, \epsilon)$ 
    Form adversarial examples:  $x_{\text{adv}} \leftarrow \text{clip}(x + \delta, 0, 1)$ 
    Apply channel:  $x_{\text{ch}} \leftarrow \text{Channel}(x_{\text{adv}}, \mathcal{C}, \sigma, l)$ 
    Compute classifier logits:  $z \leftarrow f(x_{\text{ch}})$ 
    Compute CW loss:  $\ell \leftarrow \ell_{\text{CW}}(z, y)$ 
    Compute gradients:  $g \leftarrow \nabla_{\theta} \ell$  ( $\theta = \text{params of } G$ )
    Update generator:  $\theta \leftarrow \mathcal{O}(\theta, g)$ 
  end for
end for
return Fine-tuned generator  $G$ 

```

$l = 1.0$, which is identical to the base case). The results of the fine-tuning experiment are shown in Figures 1a and 1b. We observed a general decrease in ASR with respect to channel perturbation strength, noting that channel-aware fine-tuning appears to modestly lessen this effect under AWGN, with channel-aware fine-tuned generators outperforming the baseline by up to 5.44% for inference $\sigma = 0.05$. At lower σ , improvements averaged $< 1\%$, suggesting diminishing benefits except in high-noise regimes. Under luminosity modulation,

both ASR and classifier performance remained comparatively stable, suggesting that linear transforms impact robustness less severely than comparatively slight injection of stochastic noise.

Additionally, we found that for the domain tested, both changes in benign luminosity and AWGN noise strength corresponded with changes in classifier performance, as seen in Figures 1c and 1d. Finally, Figure 1e and 1f show a general decline in normalized surplus ASR associated with stronger perturbations, indicating that all GAN methods tested appear to decline in performance with respect to stronger channel conditions. Under the domain tested, NS-ASR remained non-negative, meaning adversaries consistently managed to exploit residual vulnerabilities, but failed to leverage channel degradations to expand the overall attack success rate.

Overall, a luminosity factor in the range of ± 0.2 had a much smaller effect on both ASR and classifier performance than AWGN with $\sigma \in [0.03, 0.05]$; similar slight reductions in ASR were noted in AWGN for $\sigma \in [0.01, 0.02]$, indicating that the greatest effect of channel-aware fine tuning may be seen in noisy environments, rather than linear transformations such as luminosity modulation. However, such noisy environments may be correlated with lower classifier performance and the normalized surplus ASR indicates that the adversary is unable to fool a greater proportion of noise-robust samples than under noiseless or less noisy conditions.

These results indicate that even with channel knowledge integrated explicitly into the training loop, GAN-based adversaries may be inherently unable to generalize to AE tasks over a channel, despite the presence of the channel conferring decreased

classifier performance. Our findings indicate a decoupling of accuracy and robustness, although the classifier became less accurate under channel perturbations, it increased robustness to AEs from a well-trained adversary. While channel-aware fine-tuning offsets some channel interference, it does not allow GAN-based adversaries to capitalize on the concurrent degradation of classifier accuracy.

The implications of these findings are twofold: firstly, they illustrate a key limitation of current paradigms associated with adversarial generator training, that adversarial generators cannot exploit channel-induced classifier performance degradation. Secondly, mirroring the work in wireless communication, accounting for benign channel effects is required to preserve the robustness of AEs; though there remains a gap in how to prepare an adversary for this task.

B. Comparison to Baseline

To further contextualize our results, we implemented a projected gradient descent (PGD) attack with Expectation over Transformation (EoT) as a strong optimization-based baseline. PGD iteratively optimizes AEs with direct gradient access while averaging over multiple noisy realizations of the channel. When evaluated on a small subset of 50 CIFAR-10 validation samples, PGD/EoT achieved an ASR of 1.0 under both luminosity modulation ($l = 0.8, 1.2$) and AWGN ($\sigma = 0.05$) channels. This near-perfect fooling rate, even under a strong benign channel, highlights the efficiency/robustness tradeoff: while iterative methods can adapt to noisy environments with direct gradient access, generative adversaries like AdvGAN remain non-robust despite fine-tuning with perfect channel knowledge. These results support our observed limitation specific to generative adversaries rather than a fundamental barrier to AE performance under channels.

VI. CONCLUSION

In this work, we evaluated the role of channel effects in shaping the robustness of AEs generated by GANs. By explicitly incorporating AWGN and luminosity modulation channels into a fine-tuning process, we tested whether channel-aware adversarial training could provide AEs with resilience to realistic impairments. Our results showed that while fine-tuning modestly improved adversarial success rates over baseline in high σ stochastic noise, the overall effect was limited. We found that classifier accuracy degraded under channel conditions, but adversaries were unable to convert this lack of accuracy into increased adversarial success, suggesting a decoupling between classifier accuracy and adversarial success. These results suggest that current GAN-based approaches may be inherently constrained in their ability to exploit benign channel degradations for adversarial gain without significant methodological enhancement.

These findings highlight a gap between adversarial generation methods and realistic deployment environments. While luminosity changes had little effect on either classifier or adversarial performance, stochastic noise severely limited both adversarial capacity and transferability despite lowering clean

accuracy. While channel-aware fine-tuning offsets some performance loss, it does not enable adversaries to capitalize on the worsened classifier performance induced by noise.

A promising direction for future work is an information-theoretic analysis of these limitations. Modeling the classifier-adversary interaction as a communication channel itself may clarify why adversarial examples may lose efficacy as channel uncertainty increases, and could provide formal bounds on adversarial capacity under noisy transformations. In parallel, further exploration of adaptive mechanisms to exploit a mismatch in adversarial and classifier knowledge of a benign channel may help close the gap noted in this work.

REFERENCES

- [1] C. Szegedy, W. Zaremba, I. Sutskever, J. Bruna, D. Erhan, I. Goodfellow, and R. Fergus, "Intriguing properties of neural networks," *arXiv preprint arXiv:1312.6199*, 2013.
- [2] Z. Kong, J. Guo, A. Li, and C. Liu, "Physgan: Generating physical-world-resilient adversarial examples for autonomous driving," in *Proceedings of the IEEE/CVF conference on computer vision and pattern recognition*, pp. 14254–14263, 2020.
- [3] I. Singh, T. Araki, and K. Kakizaki, "Powerful physical adversarial examples against practical face recognition systems," in *Proceedings of the IEEE/CVF Winter Conference on Applications of Computer Vision*, pp. 301–310, 2022.
- [4] I. J. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," *arXiv preprint arXiv:1412.6572*, 2014.
- [5] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," *arXiv preprint arXiv:1706.06083*, 2017.
- [6] N. Carlini and D. Wagner, "Towards evaluating the robustness of neural networks," in *2017 IEEE Symposium on Security and Privacy (SP)*, pp. 39–57, IEEE, 2017.
- [7] A. Ilyas, L. Engstrom, A. Athalye, and J. Lin, "Black-box adversarial attacks with limited queries and information," in *International conference on machine learning*, pp. 2137–2146, PMLR, 2018.
- [8] C. Xiao, B. Li, J.-Y. Zhu, W. He, M. Liu, and D. Song, "Generating adversarial examples with adversarial networks," *arXiv preprint arXiv:1801.02610*, 2018.
- [9] H. Wei, H. Tang, X. Jia, Z. Wang, H. Yu, Z. Li, S. Satoh, L. Van Gool, and Z. Wang, "Physical adversarial attack meets computer vision: A decade survey," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 46, no. 12, pp. 9797–9817, 2024.
- [10] I. J. Goodfellow, J. Pouget-Abadie, M. Mirza, B. Xu, D. Warde-Farley, S. Ozair, A. Courville, and Y. Bengio, "Generative adversarial nets," *Advances in neural information processing systems*, vol. 27, 2014.
- [11] J. Zhang, Z. Zhu, X. Wang, S. Liao, Z. Jin, F. D. Salim, and H. Chen, "Paradvgan: Improving adversarial attack capability with progressive auto-regression advgan," *arXiv preprint arXiv:2502.12207*, 2025.
- [12] S. Jandial, P. Mangla, S. Varshney, and V. Balasubramanian, "AdvGAN++: Harnessing latent layers for adversary generation," in *Proceedings of the IEEE/CVF International Conference on Computer Vision Workshops*, pp. 0–0, 2019.
- [13] C. Zhang, S. Yu, Z. Tian, and J. J. Yu, "Generative adversarial networks: A survey on attack and defense perspective," *ACM Computing Surveys*, vol. 56, no. 4, pp. 1–35, 2023.
- [14] A. Athalye, L. Engstrom, A. Ilyas, and K. Kwok, "Synthesizing robust adversarial examples," in *International conference on machine learning*, pp. 284–293, PMLR, 2018.
- [15] B. Yang, K. Xu, H. Wang, and H. Zhang, "Random transformation of image brightness for adversarial attack," *Journal of Intelligent & Fuzzy Systems*, vol. 42, no. 3, pp. 1693–1704, 2022.
- [16] B. Kim, Y. E. Sagduyu, K. Davaslioglu, T. Erpek, and S. Ulukus, "Channel-aware adversarial attacks against deep learning-based wireless signal classifiers," *IEEE Transactions on Wireless Communications*, vol. 21, no. 6, pp. 3868–3880, 2021.
- [17] K. He, X. Zhang, S. Ren, and J. Sun, "Deep residual learning for image recognition," in *Proceedings of the IEEE conference on computer vision and pattern recognition*, pp. 770–778, 2016.