Article

# title

**Jun Li** [1,†,*]

[1] School of Management Science and Information Engineering, Jilin University of Finance and Economics, Changchun 130117, China;
* Correspondence: lijun@jlufe.edu.cn

**Abstract:** As deep learning models are increasingly e

**Keywords:** adversarial examples; adversarial attacks; deep learning; computer vision; complex systems; systems assurance

## 1. Introduction

With the rapid advancement of computer technology and artificial intelligence write lijun write

## 2. Related Works

Adversarial attacks are systematically classified into distinct categories based on various criteria. A primary distinction is established based on the adversary's knowledge of the model's internal architecture, resulting in two principal paradigms: white-box attacks and black-box attacks [1]. A white-box attack is predicated on the assumption that the adversary possesses comprehensive knowledge of the target model's internal structure and gradient information to synthesize adversarial samples. Extensive research has demonstrated that white-box attacks can craft adversarial examples with a high success rate [2]. In contrast, a black-box attack is conducted without access to the internal structure or gradient derivatives of the targeted model. However, adversarial examples generated in white-box settings often exhibit limited transferability when applied to black-box models protected by defensive mechanisms [2]. Furthermore, attacks can be categorized by the intended outcome: targeted attacks are engineered to force the model to misclassify input data into a specific, predetermined class under defined constraints, whereas non-targeted attacks aim solely to induce misclassification without a specific target label constraint. ylh ylh asdasd

The continuous evolution of adversarial defense mechanisms has necessitated the development of increasingly sophisticated attack algorithms. Ensuring the robustness and security of deep learning models requires the deployment of more potent adversarial attack methodologies. Although deep neural networks (DNNs) have achieved remarkable performance, distinct vulnerabilities have been uncovered in multiple state-of-the-art architectures. For instance, Convolutional Neural Networks (CNNs), despite being trained meticulously for image classification, have been shown to perform disastrously when subjected to adversarial attacks [3,4]. Similar fragility has been observed in other domains; neural retrieval models are brittle when faced with distribution shifts or malicious attacks [5], and plant disease classification models remain susceptible to robustness issues [6]. Adversaries exploit these intrinsic blind spots to generate adversarial examples capable of misleading machine learning models through imperceptible perturbations to the input data distribution [2].

## 3. Conclusion

In this work, we introduced the................

**Citation:** Li, J. title. *Systems* **2024**, *1*, 0. https://doi.org/

Received:
Revised:
Accepted:
Published:

**Institutional Review Board Statement:** Not applicable.

**Informed Consent Statement:** Not applicable.

**Data Availability Statement:** The original contributions presented in the study are included in the article, further inquiries can be directed to the corresponding author.

**Conflicts of Interest:** The authors declare no conflicts of interest.

## References

1. Chakraborty, A.; Alam, M.; Dey, V.; Chattopadhyay, A.; Mukhopadhyay, D. A survey on adversarial attacks and defences. *CAAI Transactions on Intelligence Technology* **2021**, *6*, 25–45. [CrossRef]
2. Wang, Y.; Liu, J.; Chang, X.; Wang, J.; Rodriguez, R.J. AB-FGSM: AdaBelief optimizer and FGSM-based approach to generate adversarial examples. *Journal of Information Security and Applications* **2022**, *68*, 103227.
3. Sen, J.; Dasgupta, S. Adversarial attacks on image classification models: FGSM and patch attacks and their impact. *arXiv preprint arXiv:2307.02055* **2023**.
4. Sen, J. The FGSM Attack on Image Classification Models and Distillation as Its Defense. In Proceedings of the Advances in Distributed Computing and Machine Learning; Nanda, U.; Tripathy, A.K.; Sahoo, J.P.; Sarkar, M.; Li, K.C., Eds., Singapore, 2024; pp. 347–360.
5. Lupart, S.; Clinchant, S. A study on FGSM adversarial training for neural retrieval. In Proceedings of the European Conference on Information Retrieval. Springer, 2023, pp. 484–492.
6. You, H.; Lu, Y.; Tang, H. Plant Disease Classification and Adversarial Attack Using SimAM-EfficientNet and GP-MI-FGSM. *Sustainability* **2023**, *15*. https://doi.org/10.3390/su15021233.