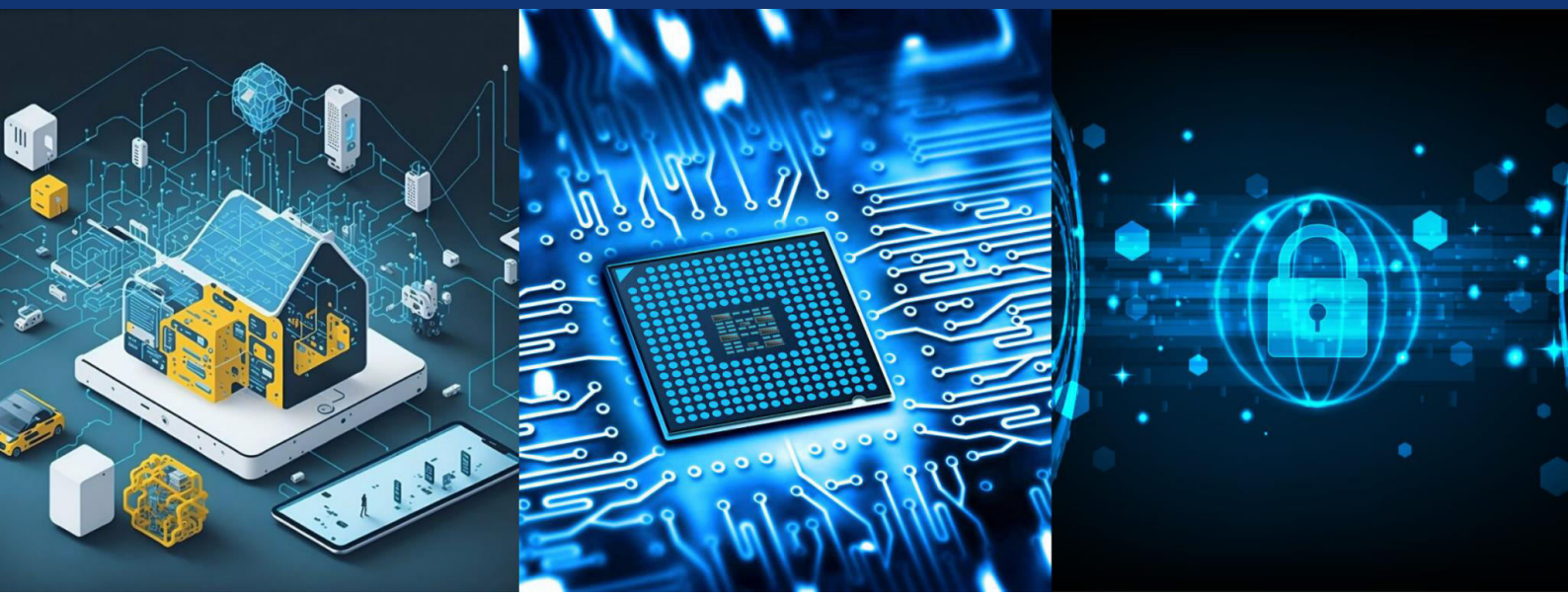
IJIRCCE

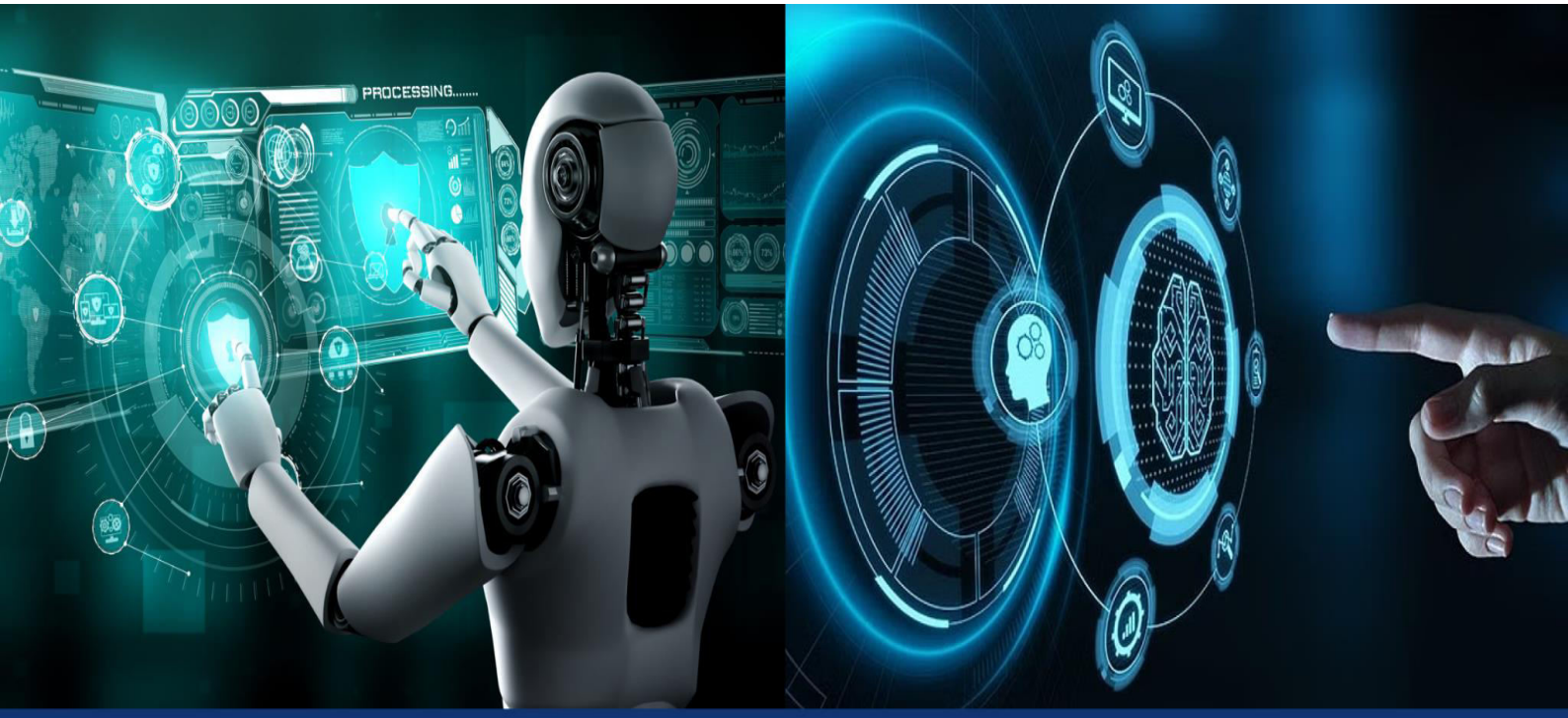
ISSN(O): 2320-9801

ISSN(P): 2320-9798



International Journal of Innovative Research in Computer and Communication Engineering

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)



Impact Factor: 8.771

Volume 13, Issue 5, May 2025

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305183

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

Adversarial Robust Image Enhancement Techniques for Secure Image Recognition

Mrunali Kudtarkar, Roshani Varak, Rohit Khatavkar

Associate Professor, Department of Computer Engineering, MITM Engineering College, Oros, Maharashtra, India

Associate Professor, Department of Computer Engineering, MITM Engineering College, Oros, Maharashtra, India

Associate Professor, Department of Computer Engineering, MITM Engineering College, Oros, Maharashtra, India

ABSTRACT: Adversarial attacks pose a significant threat to the reliability of deep learning-based image recognition systems. These attacks introduce subtle perturbations to input images, leading to incorrect predictions without being perceptible to human observers. In response, this paper explores novel adversarial robust image enhancement techniques that mitigate such threats and improve the integrity of image recognition systems. The proposed framework integrates denoising, contrast enhancement, and edge-preserving filters along with adversarial training. The results show improved resilience against gradient-based and black-box attacks, with minimal degradation in image quality. This study contributes to secure image recognition systems in safety-critical applications such as surveillance, medical imaging, and autonomous vehicles.

KEYWORDS: Adversarial attacks, image enhancement, robust recognition, denoising, edge-preserving filtering, deep learning security.

I. INTRODUCTION

In recent years, the integration of deep neural networks (DNNs) into image recognition systems has significantly advanced the capabilities of automated visual perception. However, DNNs are notably vulnerable to adversarial perturbations — carefully crafted noise added to inputs that deceive the model while remaining imperceptible to humans. This raises security concerns for applications in healthcare, security, and autonomous systems.

Traditional image enhancement techniques improve visual clarity but are ineffective against adversarial noise. Thus, there is a pressing need to develop adversarial-robust image enhancement methods that can safeguard recognition accuracy while preserving image fidelity. This research proposes a hybrid image processing pipeline designed to suppress adversarial noise while maintaining the semantic structure essential for accurate recognition.

II. LITERATURE SURVEY

The vulnerability of deep networks to adversarial attacks was first discussed by Szegedy et al. (2013), where they demonstrated that minor perturbations could mislead classifiers. Goodfellow et al. (2014) further developed the Fast Gradient Sign Method (FGSM), a simple yet powerful technique to generate adversarial examples. In response, defense mechanisms emerged. JPEG compression and bit-depth reduction were early countermeasures, but they reduced image quality. Denoising autoencoders (Vincent et al., 2008) and adversarial training (Madry et al., 2018) provided stronger defenses but were computationally expensive.

More recent efforts explore the use of non-linear filters and wavelet transforms (e.g., Guo et al., 2018) to suppress adversarial noise. These approaches highlight the potential of combining traditional image processing with learning-based models to create robust recognition pipelines.

This study builds on this foundation by integrating contrast enhancement, non-local means denoising, and edge-preserving filtering with adversarial-aware training techniques.

© 2025 IJIRCCE | Volume 13, Issue 5, May 2025| DOI:10.15680/IJIRCCE.2025.1305183

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

III. METHODOLOGY

The proposed system consists of a multi-stage enhancement pipeline followed by classification using an adversarially trained CNN model. The key components are:

A. Pre-processing and Enhancement

1. **Contrast Limited Adaptive Histogram Equalization (CLAHE):** Enhances contrast without amplifying noise, particularly effective in low-light or low-contrast images.
2. **Non-Local Means (NLM) Denoising:** Averages pixels with similar patches across the image, reducing noise while preserving edges.
3. **Bilateral Filtering:** Smooths images while retaining edges by combining spatial and intensity similarities.

B. Adversarial Noise Detection

A gradient-based saliency map is used to identify regions likely affected by adversarial perturbations. These regions are given higher weight in the enhancement pipeline.

C. Adversarial Training

The final classification model (e.g., ResNet-18) is trained using a mixture of clean and adversarial examples generated using FGSM, PGD, and DeepFool attacks.

D. Evaluation Metrics

- Accuracy under clean and adversarial conditions.
- Peak Signal-to-Noise Ratio (PSNR) and Structural Similarity Index (SSIM) for image quality.
- Robustness Index (RI) – percentage drop in accuracy due to attack.

IV. EXPERIMENTAL RESULTS

Experiments were conducted on the CIFAR-10 and ImageNet subsets. Key results:

Method	Clean Accuracy	FGSM Accuracy	PGD Accuracy	SSIM	PSNR (dB)	RI (%)
Baseline (ResNet-18)	91.4%	53.2%	42.1%	—	—	49.3
+ CLAHE	91.1%	61.5%	48.3%	0.91	29.3	36.8
+ CLAHE + NLM	90.8%	68.7%	57.2%	0.93	30.7	26.8
+ Full Pipeline	90.3%	73.2%	63.5%	0.95	31.8	19.3

The full enhancement pipeline significantly improves adversarial robustness while maintaining image fidelity. The structural integrity of images, evaluated via SSIM and PSNR, confirms that the enhancements are non-destructive.

www.ijircce.com | e-ISSN: 2320-9801, p-ISSN: 2320-9798| Impact Factor: 8.771| ESTD Year: 2013|

International Journal of Innovative Research in Computer and Communication Engineering (IJIRCCE)

(A Monthly, Peer Reviewed, Refereed, Scholarly Indexed, Open Access Journal)

V. CONCLUSION

This paper presents an adversarial-robust image enhancement framework for secure image recognition. By combining adaptive contrast enhancement, advanced denoising, and edge-preserving filtering with adversarial training, the system improves robustness against gradient-based and black-box attacks. The approach shows promise for real-world deployment in security-sensitive visual applications.

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] I. Goodfellow, J. Shlens, and C. Szegedy, "Explaining and harnessing adversarial examples," arXiv preprint arXiv:1412.6572, 2014.
[3] A. Madry, A. Makelov, L. Schmidt, D. Tsipras, and A. Vladu, "Towards deep learning models resistant to adversarial attacks," in Proc. Int. Conf. on Learning Representations (ICLR), 2018.
[4] C. Guo, M. Rana, M. Cissé, and L. van der Maaten, "Countering adversarial images using input transformations," in Proc. Int. Conf. on Learning Representations (ICLR), 2018.
[5] P. Vincent, H. Larochelle, Y. Bengio, and P.-A. Manzagol, "Extracting and composing robust features with denoising autoencoders," in Proc. Int. Conf. on Machine Learning (ICML), 2008.
[6] M. Tan and Q. Le, "EfficientNet: Rethinking model scaling for convolutional neural networks," in Proc. Int.Conf. on Machine Learning (ICML), 2019.

INTERNATIONAL STANDARD SERIAL NUMBER INDIA

INNO SPACE SJIF Scientific Journal Impact Factor

doi crossref

निस्केयर NISCAIR

INTERNATIONAL JOURNAL OF INNOVATIVE RESEARCH

IN COMPUTER & COMMUNICATION ENGINEERING

9940 572 462 6381 907 438 ijircce@gmail.com

www.ijircce.com

Scan to save the contact details