CYBER VACCINATOR FOR IMAGE TAMPER RESILIENT AND RECOVERY USING INVERTIBLE NEURAL NETWORK

By

SATHISH A. *

AASHA R. **

ABINAYASRI P. ***

KAVIYA B. ****

*-**** Roever Engineering College, Perambalur, Tamil Nadu, India.

https://doi.org/10.26634/jdf.2.1.21059

Date Received: 06/08/2024 Date Revised: 13/08/2024 Date Accepted: 21/08/2024

ABSTRACT

People frequently interact with their families, friends, and colleagues through Online Social Networks (OSNs). People post and share their photos in online communities and content-sharing sites. The problem addressed in this paper is the susceptibility of digital images to tampering, which compromises security and privacy. Traditional image forgery detection methods face challenges in reproducing original content after manipulation. This paper introduces an advanced Image Immunization System leveraging Invertible Neural Networks. The system, which comprises the cyber vaccinator, vaccine validator, forward pass for tamper detection, and backward pass for image self-recovery, aims to proactively immunize images against various attacks. The run-length encoding in the backward pass transforms hidden perturbations into information, facilitating the recovery of the authentic image. The middleware's expansion to multimodal content analysis, including videos and audio, provides a more comprehensive defense against digital manipulation within OSNs. These advancements reflect a commitment to robust security and holistic content integrity. The Cyber Vaccinator, using Invertible Neural Networks (INNs) for image tamper resilience and recovery, demonstrates significant effectiveness in detecting tampering and restoring images, providing a robust solution for maintaining image integrity. The Cyber Vaccinator uses an Invertible Neural Network (INN) to safeguard image integrity. It detects tampering by analyzing invariant features and responds with precise recovery methods. By continuously monitoring images, it ensures real-time tamper detection and efficient restoration, maintaining image authenticity through advanced neural network resilience and recovery techniques.

Keywords: Backward Pass, Forward Pass, Immunizer, Image Tampering Detection, Image Integrity Verification, Invertible Neural Networks, Cybersecurity, Image Recovery Techniques.

INTRODUCTION

Digital Image Forgery Detection is a binary classification task used to classify an image as either forged or authentic. Forged images cannot always be detected by the naked eye. Passive image forgery detection methods benefit from the features retained by the image during various stages of digital image acquisition and storage (Mehrjardi et al., 2023). These methodologies do not require past information about the image. Multitask learning is a machine learning paradigm that involves training a model to perform multiple tasks simultaneously. Instead of training separate models for each task, multitask learning leverages shared information across tasks to improve overall performance. Invertible Neural Networks (INNs) represent a specialized class of neural networks designed to be reversible, allowing for the

reconstruction of input data from the network's output (Bolourian Haghighi et al., 2020). Used in generative models for realistic data generation, INNs offer a unique ability for reversible data transformations, making them valuable in various applications.

During the immunization phase, the Cyber Vaccinator leverages the power of INNs to embed cryptographic signatures or watermarks directly into the image data. These signatures are imperceptible to the human eye yet robustly encoded within the image. By utilizing INNs, the process of embedding these signatures can be performed in a reversible manner, ensuring that the original image quality remains intact. If an image undergoes tampering or manipulation, the embedded cryptographic signatures serve as digital antibodies. The Cyber Vaccinator employs advanced algorithms to analyze discrepancies between the original image and its altered version. Leveraging the inherent reversibility of INNs, the system can precisely identify the extent and nature of the tampering. Once tampering is detected, the Cyber Vaccinator initiates the recovery process. By leveraging the embedded signatures and the inherent reversibility of INNs, the system can accurately reconstruct the original, untampered image. This recovery process ensures that even in the face of sophisticated tampering techniques, the integrity and authenticity of the image can be restored with a high degree of fidelity.

The Cyber Vaccinator represents a groundbreaking approach to enhancing the resilience of digital images against tampering while also providing an effective mechanism for recovery. In the realm of digital images, authenticity and integrity are paramount across various sectors, including security, media, and forensic analysis. The rise of sophisticated image manipulation techniques has made it increasingly challenging to detect and address tampering effectively. These manipulations can undermine trust in visual evidence, making the development of robust and reliable countermeasures a critical necessity. The paper proposes that an advanced system using Invertible Neural Networks (INNs) could be designed to fortify digital images against tampering and enable reliable recovery of the original content. By leveraging the reversible nature of INNs, this system would encode images in such a way that any unauthorized modifications could be detected and potentially reversed. This approach aims to preserve the integrity and authenticity of digital images, even after they have been altered. Such a system would be particularly beneficial in areas where the protection of digital image data is critical, including digital forensics, legal documentation, and secure information exchange.

The dataset used in the "Cyber Vaccinator" paper for image tamper resilience and recovery using Invertible Neural Networks (INN) is DIV2K (Manjunatha & Patil, 2021). Known for its high-resolution images, DIV2K is frequently utilized in tasks like super-resolution and image restoration, providing a rich and diverse set of visuals. This makes it an ideal choice for testing the robustness and effectiveness of INN-based methods in recovering tampered images. The ultimate contribution of the Cyber Vaccinator lies in its ability to combine detection, localization, and recovery in a unified framework, thereby improving overall image integrity and trustworthiness. The Cyber Vaccinator is an advanced system designed for image tamper resilience and recovery, leveraging Invertible Neural Networks (INNs). By utilizing INNs, which are adept at identifying invariant features in images, the Cyber Vaccinator can detect tampering with high accuracy. It continuously monitors images, detects alterations, and ensures their integrity. When tampering is identified, the system employs sophisticated recovery techniques to restore images to their original state. This approach combines robust tamper detection with effective recovery mechanisms, providing a comprehensive solution for maintaining image authenticity in a digital environment. The Image Immunizer Middleware for Online Social Networks (OSN) using Invertible Neural Networks (INN) is designed to enhance the security and integrity of images shared on social media platforms. The proposed system comprises several key modules and functionalities to achieve this objective.

1. Related Works

Zhang et al. (2022) proposed a facial prior and semantic guidance approach for iterative face inpainting using

GAN inversion techniques and predicted semantic information. This system involves iteratively refining images multiple times, updating semantic maps at each iteration.

Liang et al. (2021) proposed an enhanced tamper detection algorithm using YOLOv5s with CBAM attention and EIOU loss. YOLOv5s with CBAM attention improves feature representation, while the EIOU loss function enhances detection accuracy. YOLOv5s provides real-time processing capabilities, making the proposed algorithm suitable for various applications.

Park et al. (2021) and Parashar et al. (2015) proposed image tampering localization by integrating singular value decomposition (SVD) into the demosaicing process. This method involves decomposing the green channel into four sub-images according to the Bayer pattern and extracting prediction residues without requiring knowledge of the demosaicing interpolation.

Cheng et al. (2021) proposed TransU2-Net, a hybrid transformer architecture for image splicing forgery detection. TransU2-Net integrates self-attention and cross-attention into U2-Net, achieving better performance compared to state-of-the-art methods, with an 8.4% improvement in F-measure on the Casia 2.0 dataset.

2. System Model

Digital images are susceptible to malicious tampering, such as content addition or removal, which can significantly alter their original meaning. Image immunization is a technology that protects images by introducing trivial perturbations. These perturbations make the protected images resistant to tampering, allowing for the automatic recovery of tampered content.

2.1 Disadvantages

- Invertible neural networks can be computationally expensive.
- Storage requirements.
- The system balances image immunization and recovery.
- Striking the right balance is crucial; overly aggressive immunization could hinder content recovery.
- Users have limited control over the immunization process.

3. Proposed System Architecture

Digital images face an ever-evolving threat landscape, including malicious tampering that alters their original content. In this context, the concept of cyber vaccination is introduced as an approach to confer immunity to images against tampering.

3.1 Advantages

- It enhances resistance to unauthorized alterations.
- It ensures recovery without loss of original image information.
- It improves the security of digital images against tampering threats.
- It is applicable across domains such as forensics and digital communication.
- It maintains the authenticity of recovered images.
- It effectively addresses challenges posed by compressed or low-resolution images.

The architecture for a system consisting of one user and an attacker leverages Invertible Neural Networks (INNs). Figure 1 shows the system architecture.

The INN takes the original image as input and generates an adversarial image. The INN architecture resembles a U-shape, allowing for efficient transformation. It ensures that the transformation is reversible, preserving authorized users' ability to revert the protection process. The user's valuable image undergoes the INN-based transformation, resulting in an adversarial image that is non-recognizable and non-trainable. The INN's reversibility ensures that authorized models can still perform effectively. Invertible Neural Network technology provides a formidable defense, securing the authenticity and integrity of images shared on social networking platforms. Through a process involving the Cyber Vaccinator Module, the system adeptly pre-processes, vaccinates, and post-processes images, introducing imperceptible perturbations to fortify them against potential tampering. The middleware's seamless

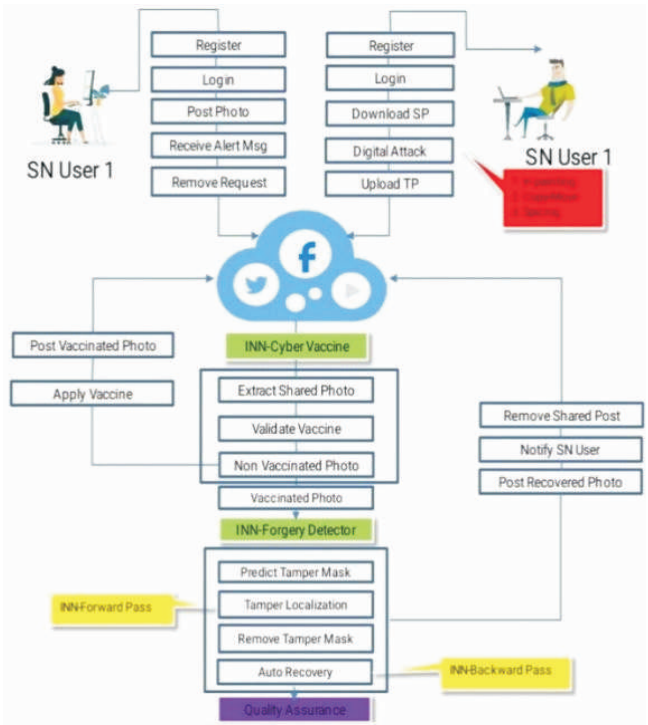

Figure 1. System Architecture

integration with existing OSN architectures not only ensures compatibility but also facilitates widespread adoption across popular social media platforms. Figure 2 shows the INN.

4. Proposed Work

4.1 Cyber Vaccine

The core module involves pre-processing, mid-processing, and post-processing steps. Landmark detection algorithms are utilized to create binary masks that distinguish object contours in images shared on OSNs. The mid-processing step generates a raw output by combining the image and mask, while the post-processing step replaces the object region in the raw output with that of the original image. Imperceptible perturbations are introduced to the non-object regions, ensuring visual consistency while embedding crucial information.

4.2 Vaccine Validator

The system includes a vaccine validator module specific to OSN. It distinguishes between vaccinated (secured) and unvaccinated (potentially tampered)

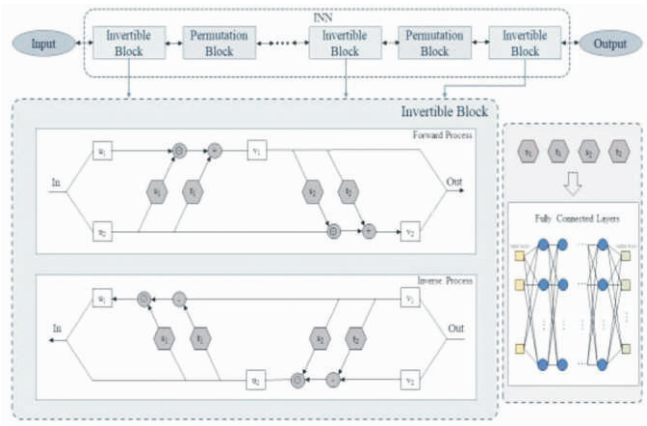

Figure 2. INN

media shared on the platform. This component ensures the validation of image integrity, preventing the dissemination of potentially manipulated content. An adversary is integrated to simulate potential threats, including deepfake attempts, within the social network context.

4.3 Forward Pass

The forward pass involves transforming the original image and its associated metadata into an immunized version using an INN. In the case of an attacked image, a localizer is employed to determine the tampered areas by predicting the tamper mask and the type of attack. This step is crucial for identifying and localizing potential manipulations within the social network environment (Chennamma & Madhushree, 2023).

4.4 Backward Pass - Image Self-Recovery

In the backward pass of the INN, the hidden perturbation is transformed into information, facilitating the recovery of the original image and its associated metadata. Image self-recovery is encouraged to ensure that the recovered image closely resembles the original, maintaining visual and contextual consistency within the OSN context.

4.5 Adversarial Simulation for OSN

The system incorporates an adversarial simulation strategy during training, tailored for OSN scenarios. This approach exposes the network to potential threats specific to social media, including image-based attacks such as deepfakes and contextually relevant manipulations.

5. Discussion

The rapid integration of Online Social Networks (OSNs) into daily life has heightened concerns about digital image tampering, which undermines security and privacy. This paper introduces an advanced Image Immunization System utilizing Invertible Neural Networks (INNs) to address these issues. Traditional forgery detection methods struggle with the reconstruction of original content after manipulation, but the Cyber Vaccinator overcomes this by employing a multifaceted approach: Cyber Vaccinator, Vaccine Validator, Forward Pass for Tamper Detection, and Backward Pass for Image Self-Recovery (Xiu-Jian & Sun, 2022). By embedding cryptographic signatures imperceptibly within images, the Cyber Vaccinator creates a robust defense against tampering (Bondi et al., 2017). The system's Forward Pass identifies and localizes tampering, while the Backward Pass, leveraging run-length encoding transforms hidden perturbations into recoverable information. This innovative approach enhances the resilience of digital images against tampering and offers a reliable mechanism for restoring original content. The expansion of this middleware to encompass multimodal content analysis further strengthens its applicability in diverse contexts, such as videos and audio, thus providing a comprehensive solution for safeguarding digital integrity across OSNs. By combining detection, localization, and recovery within a unified framework, the Cyber Vaccinator represents a significant advancement in image integrity and cybersecurity.

Conclusion

Invertible Neural Network technology provides a formidable defense, securing the authenticity and integrity of images shared on social networking platforms. Through a process involving the Cyber Vaccinator Module, the system adeptly pre-processes, vaccinates, and post-processes images, introducing imperceptible perturbations to fortify them against potential tampering. The middleware's seamless integration with existing OSN architectures not only ensures compatibility but also facilitates widespread adoption across popular social media platforms.

References

[1]. Bondi, L., Lameri, S., Güera, D., Bestagini, P., Delp, E. J., & Tubaro, S. (2017, July). Tampering detection and localization through clustering of camera-based CNN features. In *2017 IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)* (pp. 1855-1864). IEEE.

https://doi.org/10.1109/CVPRW.2017.232

[2]. Xiu-Jian, L., & Sun, H. (2022). Deep learning based image forgery detection methods. *Journal of Cybersecurity,* 4(2), 119.

https://doi.org/10.32604/jcs.2022.032915

[3]. Mehrjardi, F. Z., Latif, A. M., Zarchi, M. S., & Sheikhpour, R. (2023). A survey on deep learning-based image forgery detection. *Pattern Recognition,* 109778.

https://doi.org/10.1016/j.patcog.2023.109778

[4]. Chennamma, H. R., & Madhushree, B. (2023). A comprehensive survey on image authentication for tamper detection with localization. *Multimedia Tools and Applications,* 82(2), 1873-1904.

https://doi.org/10.1007/s11042-022-13312-1

[5]. Parashar, N., Tiwari, N., & Dubey, D. (2015). A survey of digital image tampering techniques. *International Journal of Signal Processing, Image Processing and Pattern Recognition,* 8(10), 91-96.

[6]. Singh, K. N., & Singh, A. K. (2022). Towards integrating image encryption with compression: A survey. *ACM Transactions on Multimedia Computing, Communications, and Applications (TOMM),* 18(3), 1-21.

https://doi.org/10.1145/3498342

[7]. Bolourian Haghighi, B., Taherinia, A. H., & Monsefi, R. (2020). An effective semi-fragile watermarking method for image authentication based on lifting wavelet transform and feed-forward neural network. *Cognitive Computation*, 12, 863-890.

https://doi.org/10.1007/s12559-019-09700-9

[8]. Chen, Y., Liu, L., Phonevilay, V., Gu, K., Xia, R., Xie, J., & Yang, K. (2021). Image super-resolution reconstruction based on feature map attention mechanism. *Applied Intelligence,* 51, 4367-4380.

https://doi.org/10.1007/s10489-020-02116-1

[9]. Liang, X., Tang, Z., Huang, Z., Zhang, X., & Zhang, S. (2021). Efficient hashing method using 2D-2D PCA for image copy detection. *IEEE Transactions on Knowledge and Data Engineering,* 35(4), 3765-3778.

https://doi.org/10.1109/TKDE.2021.3131188

[10]. Park, C. W., Moon, Y. H., & Eom, I. K. (2021). Image tampering localization using demosaicing patterns and singular value based prediction residue. *IEEE Access,* 9, 91921-91933.

https://doi.org/10.1109/ACCESS.2021.3091161

[11]. Zhang, X. Y., Xie, K., Li, M. R., Wen, C., & He, J. B. (2022). Generative facial prior and semantic guidance for iterative face inpainting. *IEEE Access,* 10, 66757-66769.

https://doi.org/10.1109/ACCESS.2022.3185210

[12]. Manjunatha, S., & Patil, M. M. (2021, February). Deep learning-based technique for image tamper detection. In *2021 Third International Conference on Intelligent Communication Technologies and Virtual Mobile Networks (ICICV)* (pp. 1278-1285). IEEE.

https://doi.org/10.1109/ICICV50876.2021.9388471

ABOUT THE AUTHORS

Sathish A., Roever Engineering College, Perambalur, Tamil Nadu, India.

Aasha R., Roever Engineering College, Perambalur, Tamil Nadu, India.

Abinayasri P., Roever Engineering College, Perambalur, Tamil Nadu, India.

Kaviya B., Roever Engineering College, Perambalur, Tamil Nadu, India.