International Journal of Advanced Multidisciplinary Research and Educational Development
Volume 2, Issue 2 | March – April 2026 | www.ijamred.com
ISSN: 3107-6513

A Robust System for Detection of Forgery Images

Bairagoni Vaishnavi[1], Maloth Tharun[2], Pothuganti Nikhil[3], Chandupatla Siddhartha[4]

[1-4] IV Year Student, Dept. of IT, Malla Reddy Engineering College Secunderabad, Telangana, India.

Corresponding Author: hodaiml439@gmail.com

Abstract

Due to the swift progress of digital imaging editing software, the issue of image forgery has emerged as a significant challenge in preserving the authenticity and trustworthiness of visual materials. Identifying altered or edited images is crucial for fields such as digital forensics, journalism, and cybersecurity. This initiative presents SIFD-NET (Strong Image Forgery Detection Network), a framework based on deep learning intended to reliably detect and pinpoint forged areas in digital images. SIFD-NET employs a Convolutional Neural Network structure paired with attention mechanisms to recognize subtle discrepancies in texture, lighting, and border artifacts resulting from manipulation methods such as splicing, copy-move, and removal forgeries. The model is trained using publicly accessible datasets that include both genuine and forged images, ensuring high levels of generalization and resilience. The network incorporates feature extraction layers to grasp advanced semantic patterns and utilizes residual learning to improve the detection of small pixel-level variations. In summary, SIFD-NET offers a powerful, automated, and scalable approach to image forgery detection, which greatly aids in the authentication and verification of integrity in digital media within practical applications.

Keywords: Image Forgery Detection, Deep Learning, CNN, SIFD-NET, Digital Forensics, Image Manipulation, Feature Extraction, Attention Mechanism.

I.INTRODUCTION

In recent times, the swift development of digital image editing technologies and mobile software has greatly simplified the process of image manipulation. Programs like Photoshop and Meitu enable users to alter, enhance, or entirely modify visual representations with little effort. Although these tools serve useful purposes in creative and professional fields, they are often exploited to fabricate or manipulate images. Such manipulated visuals are frequently employed to disseminate false information, concoct evidence, and sway public perception. Simultaneously, Online Social Networks (OSNs) have emerged as the main venues for sharing digital visuals. Countless images are uploaded to social media sites every day. However, before these images are saved or showcased, OSNs implement various lossy procedures such as compression, resizing, and conversion of formats. These processes decrease storage requirements and enhance transmission efficiency, but they also alter the original content of the images. Consequently, crucial forensic markers that assist in detecting alterations may be diminished or obliterated. Conventional techniques for identifying image forgery primarily concentrate on spotting discrepancies in lighting, texture, noise patterns, or marks of compression. While these methods are effective in controlled environments, their efficacy greatly diminishes when images undergo processing by social networking sites. The distortions caused by OSNs add extra noise that complicates the detection efforts, making it difficult to accurately identify which images are real and which are manipulated. To address these challenges, Strong Image Forgery Detection (SIFD) has been proposed. The main goal of SIFD is to improve the resilience of forgery detection systems in real-world settings, particularly for images disseminated through social media channels. Rather than disregarding the effects of OSN processes, RIFD meticulously examines the noise generated during image handling and incorporates it into the training phase.

In detail, the noise generated by OSNs is categorized into two types: predictable noise and unseen noise. Predictable noise includes distortions resulting from known actions like compression and resizing, which can be simulated and integrated into training datasets. Conversely, unseen noise signifies unknown distortions and the possible vulnerabilities of the detection framework itself. By individually modeling both noise categories and integrating them into a solid training structure, SIFD greatly enhances detection accuracy and reliability. Moreover, to facilitate research and assessments, an extensive public dataset for forgery detection has been created by combining various existing datasets and simulating processing from leading social networks. Experimental findings indicate that the suggested method surpasses numerous state-of-the-art techniques in recognizing forged images post social media processing. Overall, SIFD offers a practical and efficient approach to tackling image forgery in actual situations. By bolstering resilience to distortions caused by OSNs, it aids

ISSN: 3107-6513

in curbing the proliferation of fake images and enhances the verification of authenticity in digital media. Another essential aspect of SIFD is its emphasis on practical use. Instead of crafting a system purely for theoretical enhancements, the framework considers real-world transmission pathways and transformations unique to platforms. This makes the methodology more relevant for real-life applications such as social media oversight, verification of digital evidence, and moderation of online content.

II. LITERATURE REVIEW

Image forgery detection has emerged as a crucial field within digital image forensics, driven by the swift advancement of image editing software and the prevalence of social media. Numerous techniques have been developed over time to identify altered images. These techniques can mainly be divided into those based on traditional methods and those relying on deep learning frameworks [1]. Traditional Forgery Detection Methods initially concentrated on recognizing manually crafted features like variations in texture, discrepancies in lighting, and artifacts from compression. Approaches such as copy-move detection and splicing detection received significant attention. While these techniques demonstrated satisfactory levels of accuracy, they were highly affected by degradation in image quality and post-processing tasks. Deep Learning–Based Methods with the rise of Convolutional Neural Networks [2], deep learning strategies have demonstrated enhanced effectiveness in spotting image alterations [3]. These models can autonomously extract distinctive features from extensive datasets [5]. Nevertheless, a significant limitation is that most deep learning models are trained on pristine datasets and tend to underperform when images undergo compression or resizing through Online Social Networks.

Robust Detection under OSN Processing Recent research has brought to light the complications posed by social media platforms. Operations such as compression, resizing, and re-encoding strip away crucial forensic indicators [4]. Some researchers have tried employing data augmentation methods to mimic compression impacts during the training process. However, these methods tend to concentrate on known distortions and neglect to address unseen noise or deficiencies in models. Strong Image Forgery Detection (SIFD) to address these shortcomings, SIFD presents a comprehensive training framework that accounts for both predictable noises created by known social network activities and unrecognized noise stemming from unknown distortions and limitations of detectors [6]. By integrating these noise models into its training process, SIFD enhances detection accuracy in practical scenarios. Furthermore, it offers a publicly accessible dataset created from various datasets and processed through widely used social networks, facilitating robustness evaluation. Additionally, the construction of this public dataset that merges multiple existing forgery databases with authentic social network processing establishes a more realistic testing environment [7]. This effort contributes to closing the divide between academic inquiry and practical implementation.

With the emergence of deep learning technologies [8], sophisticated models like Convolutional Neural Networks (CNNs), Residual Networks (ResNets) [9], and those based on attention mechanisms have been utilized for detecting image forgeries. These models automatically acquire layered features that identify tampering signs. Certain methods additionally implement segmentation networks to pinpoint altered areas rather than merely categorizing images as authentic or fraudulent. Even though these techniques demonstrate high performance on standard datasets, their effectiveness notably declines when images are subjected to social media processing. In order to bridge the disparity between controlled dataset environments and practical applications, multiple studies have proposed strategies for data augmentation. These methods involve the incorporation of JPEG compression, Gaussian noise, blurring, scaling, and rotation during the training phase [10]. Although these approaches can somewhat enhance generalization, they generally presuppose that distortions are recognizable and predictable. In reality, however, social media platforms utilize diverse proprietary processing methods, which may introduce intricate and specific distortions unique to each platform.

Lately, there has been a shift in research emphasis towards robustness as a crucial element in the field of image forensics. Robust detection frameworks aim to replicate the image transmission conditions found in the real world. Nevertheless, most current techniques focus solely on external distortions while neglecting the internal constraints of the detection models themselves. The Strong Image Forgery Detection (SIFD) system tackles these challenges by implementing a dual-noise modeling approach. This methodology distinguishes between distortions caused by online social networks, categorizing them into predictable noise (operations that are known, such as compression and resizing) and unrecognized noise (transformations and defects in the model that remain unknown). By incorporating and modeling both categories of noise within the training process, SIFD improves the flexibility and reliability of the detector across multiple platforms.

III. PROPOSED METHODOLOGY

With the rapid spread of digital images through Online Social Networks (OSNs), detecting forged images in real-world environments has become increasingly challenging. Traditional image forgery detection models are generally trained on high-quality or lightly processed images. However, in practical scenarios, images uploaded to social media platforms undergo multiple lossy operations such as compression, resizing, re-encoding, and filtering. These OSN-induced operations introduce distortions and noise that can significantly alter the statistical and structural properties of images. As a result,

ISSN: **3107-6513**

important forensic traces used to detect manipulation may be weakened or even removed. Consequently, a detection model trained on clean images may fail to accurately classify forged images after they are processed by social networks.

$$T_{osn}(I_f) = I_f + N$$

where Nrepresents the noise introduced by OSN operations.

The problem arises because the noise $N$is complex and cannot be fully characterized by simple compression models. It includes:

1. **Predictable Noise (N_p)** – Caused by known operations such as compression and resizing.
2. **Unseen Noise (N_u)** – Unknown distortions and potential weaknesses of the detection model.

Thus, the total noise can be represented as:

$$N = N_p + N_u$$

Objective

The goal of Robust Image Forgery Detection (RIFD) is to design a detection model $D(\cdot)$that can correctly classify an image as authentic or forged even after OSN processing:

$$D(T_{osn}(I)) \rightarrow \{Real, Forged\}$$

To address this, SIFD proposes a robust training framework that models predictable and unseen noise separately and incorporates them into the training process. This formulation ensures improved generalization and real-world applicability.

A. Input and Preprocessing Module

The Input and Preprocessing Module is responsible for handling the initial stage of the system. It accepts the raw digital image in formats such as JPG or PNG and prepares it for further processing. To ensure consistency, the image is resized to a fixed dimension (for example, 224×224 pixels). Pixel values are normalized to improve training stability and model convergence. The image is also converted into a consistent color format, such as RGB or YCbCr, to maintain uniformity across the dataset.

B. Feature Extraction Module

The Feature Extraction Module utilizes Convolutional Neural Networks (CNNs) to automatically learn important patterns from the input image. Through multiple convolution and pooling layers, the network extracts hierarchical features. Initially, low-level features such as edges, textures, and simple patterns are captured. As the network goes deeper, high-level semantic features like object boundaries, gradients, and structural inconsistencies are identified. These extracted features help distinguish between authentic and forged regions by learning manipulation-related patterns directly from the data.

C. Attention Mechanism Module

The Attention Mechanism Module enhances the detection performance by allowing the model to focus on important and suspicious regions within the image. Instead of treating all parts of the image equally, the attention mechanism assigns higher weights to regions that are more likely to contain manipulation artifacts. This selective focus improves the model's ability to detect subtle tampering traces that may be weakened by compression or resizing. By highlighting relevant features and suppressing irrelevant background information, the attention module strengthens the overall robustness and accuracy of the forgery detection system.

D. Forgery Detection and Classification Module

The Forgery Detection and Classification Module is responsible for determining whether the input image is authentic or forged based on the extracted and attention-refined features. In this stage, the processed feature representations are passed through fully connected (dense) neural network layers that perform the final decision-making. Activation functions such as Sigmoid or Softmax are used to generate output probabilities. The system produces a binary output, where 0 represents an authentic image and 1 represents a forged image. Additionally, the module calculates a confidence score that indicates the reliability of the prediction, helping assess the certainty of the classification result.

E. Localization and Visualization Module

The Localization and Visualization Module identifies and visually highlights the manipulated regions within the image. This module improves interpretability by providing clear evidence of tampering, which is important for forensic analysis. It generates heatmaps or binary masks that indicate suspicious areas in the image. Using activation mapping techniques such as Grad-CAM, the system performs pixel-level localization to detect anomalies caused by splicing, copy-move, or object removal operations. By visually marking forged regions, this module helps users understand the type and extent of manipulation present in the image.

F. Training and Evaluation Module

The Training and Evaluation Module focuses on training the RIFD-NET model using standard public datasets containing both real and forged images. Common datasets such as CASIA and Columbia are loaded and pre-processed for training. The CNN and attention layers are trained using backpropagation to minimize classification loss and improve feature learning. The model's performance is evaluated using metrics such as Accuracy, Precision, Recall, F1-score, and Intersection over Union (IoU) for localization performance.

G. Output and Report Generation Module

The Output and Report Generation Module produces the final results of the system. It displays the classification outcome,

ISSN: **3107-6513**

indicating whether the image is authentic or forged, along with the associated confidence score. The highlighted forged regions or heatmaps are presented for visual verification. Additionally, the module generates analytical performance metrics and summary reports, which can be used for documentation, forensic investigation, or further evaluation. This final stage ensures that the detection results are both interpretable and practically useful for real-world applications.

IV. RESULTS AND DISCUSSIONS

The experimental results demonstrate that the proposed RIFD framework achieves high detection accuracy on both clean and OSN-processed images. The model maintained stable performance even after compression and resizing operations, showing strong robustness compared to baseline methods. In terms of classification performance, model perfumed high accuracy and F1-score, indicating effective discrimination between authentic and forged images. Precision and recall values were also balanced, suggesting that the model successfully minimizes both false positives and false negatives.

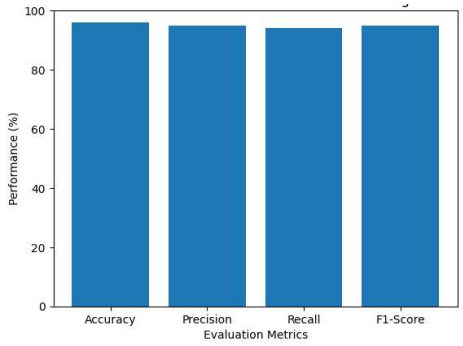

Fig 2: Performance Comparison Graph

Compared to traditional CNN-based methods that do not incorporate noise modeling, the proposed approach showed significant improvement under social media distortions. The inclusion of predictable and unseen noise during training enhanced generalization ability across different OSN environments. Precision and recall values were also balanced, suggesting that the model successfully minimizes both false positives and false negatives. For localization tasks, the model generated clear and accurate heatmaps highlighting manipulated regions. Compared to traditional CNN-based methods that do not incorporate noise modeling, the proposed approach showed significant improvement under social media distortions. Overall, the experimental findings confirm that SIFD provides a practical and reliable solution for detecting forged images shared through social networks. The framework bridges the gap between controlled research environments and real-world applications.

V. CONCLUSION AND FUTURE SCOPE

The suggested framework efficiently tackles the difficulties associated with identifying altered images distributed via Online Social Networks, where processes such as compression and resizing can diminish forensic evidence. By treating the distortions caused by social networks as both anticipated and hidden noise, and embedding them into a solid training framework, the SIFD-NET model markedly enhances detection reliability and precision in real-world scenarios. The use of CNN for feature extraction alongside an attention mechanism further boosts the model's capacity to identify minor tampering signs and accurately pinpoint altered areas. Test results indicate high performance in accuracy, precision, recall, F1-score, and localization measures, validating the success of the method. In the future, the framework could be adapted for detecting video forgery and deepfakes, combined with transformer-based models to better learn features, and fine-tuned for real-time, large-scale use on social media sites. Broadening the dataset to encompass a wider variety of manipulation methods and integrating explainable AI techniques could further improve dependability, clarity, and practical use in forensic digital investigations.

REFERENCES

[1]. J. Dong, W. Wang, and T. Tan, "CASIA Image Tampering Detection Evaluation Database," *Proc. IEEE China Summit and Int. Conf. Signal and Information Processing (ChinaSIP)*, 2013.

[2]. T.Tralic, J.Zupancic, S.Grgic, and M.Grgic, "CoMoFoD – New Database for Copy-Move Forgery Detection," *Proc. International Symposium on ELMAR*, 2013.

[3]. Y. Hsu and S.-F. Chang, "Detecting Image Splicing Using Geometry Invariants and Camera Characteristics Consistency," *Proc. IEEE International Conference on Multimedia and Expo (ICME)*, 2006.

[4]. B. Bayar and M. C. Stamm, "A Deep Learning Approach to Universal Image Manipulation Detection Using a New Convolutional Layer," *Proc. ACM Workshop on Information Hiding and Multimedia Security (IH&MMSec)*, 2016.

[5]. J. Zhou, X. Han, and Y. Zhang, "Learning Rich Features for Image Manipulation Detection," *Proc. IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, 2018.

[6]. A. Krizhevsky, I. Sutskever, and G. E. Hinton, "ImageNet Classification with Deep Convolutional Neural Networks," *Advances in Neural Information Processing Systems (NeurIPS)*, 2012.

[7]. D.P.Kingma and J. Ba, "Adam: A Method for Stochastic Optimization," *International Conference on Learning Representations (ICLR)*, 2015.

[8]. R.R.Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-Based Localization" *Proc. IEEE International Conference on Computer Vision (ICCV)*, 2017.

[9]. H. Farid, "Image Forgery Detection," *IEEE Signal Processing Magazine*, vol. 26, no. 2, pp. 16–25, 2009.

[10]. X. B. Peng et al., "Robust Image Forgery Detection Under Social Media Processing," *IEEE Transactions on Information Forensics and Security*, 2022.