A Deep Learning-Based Approach for Image Forgery Detection and Classification Using YOLO and CNN

A.Durga Bhavani
Computer Science and Engineering
Sasi Institute of Technology and Engineeering Tadepalligudem, INDIA

S.Durga Shankar
Computer Science and Engineering
Sasi Institute of Technology and Engineeering Tadepalligudem, INDIA

Ch.Aditya
Computer Science and Engineering
Sasi Institute of Technology and Engineeering Tadepalligudem, INDIA

D.Vijaya Vani
Computer Science and Engineering
Sasi Institute of Technology and Engineeering Tadepalligudem, INDIA

J.Hima Venkata Madhuri
Computer Science and Engineering
Sasi Institute of Technology and Engineeering Tadepalligudem, INDIA

Abstract –

The emergence of advanced image manipulation techniques in the digital age has made it extremely difficult to detect and categorize forged images. In order to identify and classify image forgeries, this study proposes a deep learning-based method that combines a Convolutional Neural Network (CNN) classifier with the You Only Look Once (YOLO) object detection model. The suggested system uses sophisticated feature extraction techniques to categorize images into authentic, copy-move, spliced, and deepfake forgeries after being trained on a custom dataset of manipulated images. While CNN-based classification offers accurate categorization, our experimental results show how reliable and effective YOLO is at locating forgery regions. Because of its high accuracy, the suggested model is a useful tool for security and digital forensics applications.

Keywords: Image Forgery, YOLO, CNN, Deep Learning, Classification, Digital Forensics

I.INTRODUCTION

The sophistication of digital image forgery has increased with the development of image editing tools, raising questions about the legitimacy of digital content. Highly realistic forgeries are now easier to produce thanks to the widespread use of deepfake technology and AI-assisted image manipulation tools, which presents serious risks to industries like cybersecurity, law enforcement, and journalism. Complex forgeries are difficult to identify using traditional image forensic techniques like pixel-based analysis and metadata verification because of their complexity and capacity to evade traditional detection methods.

In detecting and categorizing image forgeries, deep learning-based methods—specifically, Convolutional Neural Networks (CNNs) and object detection models like YOLO—have proven to be more effective. Because CNNs are so good at extracting features and classifying them, they can be used to classify manipulated images into various types, including deepfake-generated images, splicing, and copy-move. The accurate localization of forged areas within an image is made possible by YOLO's superior real time object detection capabilotie

In this study, we suggest a hybrid strategy that improves image forgery detection and classification by utilizing the advantages of both CNNs and YOLO. A CNN-based classifier classifies the type of forgery, while the YOLO model identifies areas that have been forged. The model is appropriate for forensic applications because of this combination, which improves both detection accuracy and interpretability. By combining these two models, we hope to increase the classification process's speed and dependability in addition to the accuracy of forgery

identification. Our suggested system offers a complete solution for digital forensic investigations, with great efficiency and accuracy in identifying different kinds of image manipulation.
Using pre-trained CNN models like MobileNet and Xception, our system integrates transfer learning to guarantee robustness in forgery detection. Even with small datasets, we can attain superior performance thanks to these models, which are renowned for their lightweight architecture and strong feature extraction capabilities. Our approach reduces overfitting while enhancing the classifier's capacity for generalization by utilizing pre-trained weights. Our model is effective for real-time applications since this transfer learning technique greatly lessens the computational load.

Additionally, both local and global forgeries can be handled by our system. While local forgeries concentrate on altered portions of an otherwise authentic image, global forgeries entail extensive manipulations like deepfake generation or full image synthesis. By combining CNN's feature extraction with YOLO's real-time object detection, we obtain high accuracy in identifying and classifying areas that have been tampered with. Even subtle manipulations, like fine-grained image splicing or copy-move alterations, are successfully identified and categorized thanks to this hybrid approach.

Finally, to confirm the efficacy of our system, it is assessed using exacting performance metrics like accuracy, precision, recall, and F1-score. We evaluate our suggested model against both stand-alone CNN- based methods and conventional forensic techniques using the CASIA dataset. The findings show that our hybrid YOLO-CNN model ensures greater reliability in forensic applications by lowering false positives and improving detection accuracy. Future developments might involve adding more varied manipulations to the dataset and improving the model to identify hostile attacks that try to evade forgery detection systems.

II.LITERATURE OVERVIEW

Numerous conventional and deep learning-based methods have been used in the extensive study of image forgery detection. Previous methods for identifying irregularities in digital photos used statistical analysis and manually created features. Forgery detection techniques like Discrete Cosine Transform (DCT) and Error Level Analysis (ELA) were frequently employed, but they frequently had trouble with intricate manipulations and high-resolution images.

Forgery detection capabilities have been greatly enhanced by recent developments in deep learning. Convolutional Neural Networks (CNNs) have been used extensively to categorize different types of forgeries and extract complex image features. By using CNNs to examine pixel-level discrepancies, research by Rao and Ni (2016) showed how well deep learning works to identify splicing and copy- move forgeries. Likewise, Zhou et al. (2018) suggested a comprehensive deep learning framework for forgery detection, demonstrating that CNN-based architectures perform more accurately and robustly than conventional techniques.

The incorporation of object detection models, such as YOLO, for detecting manipulated regions is another significant advancement in this field.
YOLOv3, a real-time object detection model that can precisely locate forgeries, was presented by Redmon and Farhadi (2018). In forensic applications, the combination of CNN for classification and YOLO for region detection has shown to be a potent strategy. Bi et al. (2020) investigated the combination of CNN classifiers and deep learning-based object detection, emphasizing the benefits of doing so for increased accuracy.

Moreover, robust detection models have been trained using realistic forged images produced by Generative Adversarial Networks (GANs). Wang's research et al. (2021) showed that by exposing forgery detection models to a variety of manipulation techniques, GAN-based training data augmentation improves their performance.

Despite these developments, it is still difficult to distinguish between different kinds of manipulations and identify forgeries in extremely complex images. It is anticipated that continued hybrid model development, which makes use of several deep learning approaches, will increase detection accuracy and dependability in practical applications.

Year	Author(s)	Proposed Work	Proposed Algorithms	Accuracy
2020	Al-Ali, A., Maad, M. A. N. A. M., & Alazab, M. A.	A deep learning approach for image forgery detection Deep Learning (CNN-based)	Deep Learning (CNN-based)	89.89%
2021	Ganesan, R.,Thavavel, S., & Siva, S. G. S.	Survey and research directions in deep learning for image forgery detection	Various Deep Learning Techniques	Up to 90%
2022	Farhan, Z. F., Wali, M. A. A. H. E. W., & Al-Hadidi, M. I. R.	Efficient image forgery detection using deep learning techniques	Deep Learning (CNN-based)	Improvements in performance by 1.9%, detection rates by 3.2%, 0% false alarm rate
2020	Qadir, R., et al.	Review of image forgery detection using deep learning techniques	Deep Learning (CNN-based)	High accuracy (not specified)
2023	AbuZaid, A. R., Ibrahim, I. M., & Zidan, M. A.	Hybrid approach for image forgery detection based on CNN and SIFT features	Hybrid CNN and SIFT	Improved performance after resampling
2020	Yang, J., Zhou, Y., & Wei, J.	Fast copy-move forgery detection using discriminative features	Binary Discriminative Features	Improved detection accuracy (specific accuracy not stated)
2020	Verdoliva L.	Overview of media forensics and deepfakes	Deep Learning for Media Forensics	High accuracy (specific accuracy not stated)
2020	Verdoliva L.	Overview of media forensics and deepfakes	Deep Learning for Media Forensics	94.09%
2018	Chen, J., et al.	Improved image forgery detection method based on deep learning	Deep Learning (CNN-based)	Not specified
2021	Costa, L. D., da Silva, G. R., & Siqueira, A.	Performance analysis of deep learning architectures for copy-move forgery detection	Deep Learning Architectures	Not Specified

Year	Author(s)	Proposed Work	Proposed Algorithms	Accuracy
2019	Zhang, Y., et al.	Survey on image forgery detection methods	Multiple Methods including Deep Learning	93.7%
2020	Mura, M., et al.	Survey on image forgery detection techniques and challenges	Deep Learning and Other Techniques.	Not specified

Methodologies and Approaches

Proposed System:

The proposed system integrates CNN models such as Xception and MobileNet with YOLO for accurate localization of forged regions, and it uses deep learning techniques for image forgery detection and classification. Feature extraction is improved by preprocessing techniques like resizing, normalization, and augmentation. Training is done using the CASIA dataset, which includes 576 tampered and 606 real images. The system uses deep hierarchical features to classify different types of forgeries, such as image splicing and copy-move. Real-time object detection is made possible by YOLO, increasing precision and effectiveness. Metrics like recall, accuracy, and precision are used to evaluate performance. This hybrid approach improves digital forensics' dependability by guaranteeing strong forgery detection with high accuracy, which makes it useful for applications in forensic investigations, journalism, and cybersecurity.

System Architecture:

The system architecture for image forgery detection and classification follows a structured pipeline that integrates deep learning models, including CNN- based architectures such as Xception and MobileNet, alongside YOLO for localization. The process begins with data acquisition, where images from the CASIA dataset are loaded and preprocessed through resizing, normalization, and augmentation to enhance model performance. The feature extraction phase utilizes CNN layers to capture important patterns, distinguishing authentic and tampered images. The classification model then categorizes images based on forgery types such as copy-move, splicing, and enhancement manipulations. For localization, YOLO identifies manipulated regions, highlighting areas of forgery with bounding boxes. Finally, the detection output is presented, showcasing classification results and marked forgery regions. The architecture ensures high accuracy and efficiency in digital forensic.

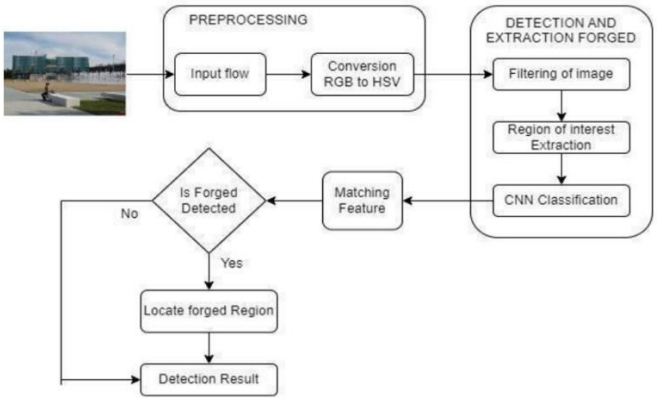

Fig 1 System Architecture

Data Preprocessing:

Preprocessing the dataset is necessary for precise image forgery detection. The CASIA dataset, which includes 576 tampered and 606 authentic images, is normalized and resized to scale pixel values between 0 and 1. Rotation, flipping, zooming, and Gaussian blurring are examples of data augmentation techniques that improve generalization and avoid

overfitting. By eliminating undesired artifacts, noise reduction filters enhance feature extraction. After that, images are transformed into tensor format so they can be used with deep learning. By ensuring efficient learning, these preprocessing procedures improve the model's precision in identifying and locating forged areas in photos.

Algorithms Used:

Deep learning models are the main tool used by the suggested system to detect and classify image forgeries. Among the primary algorithms are:

YOLO (You Only Look Once):

A real-time object detection model used for forgery localization, accurately detecting manipulated regions in images. Its fast processing speed makes it suitable for forensic applications.

CNN (Convolutional Neural Networks): Extracts hierarchical features from images, enabling the classification of forged and authentic images. It helps distinguish between different types of forgeries such as copy-move, image- splicing.

MobileNet and Xception:

These lightweight CNN architectures optimize classification accuracy while reducing computational complexity, making the system efficient and more memory friendly.

This combination improves the model's speed and accuracy in detecting, classifying, and localizing image forgeries.

Findings and Trends:

Accurracy:

Accuracy measures the model's ability to correctly classify authentic and tampered images. It is calculated as the proportion of correctly identified forged and authentic images. A higher accuracy indicates better performance in distinguishing between manipulated and original images. The formula for accuracy is:

Accuracy = (TP + TN) / (TP + TN + FP + FN)

A comparative analysis is conducted to evaluate the accuracy of different deep learning models used in forgery detection.

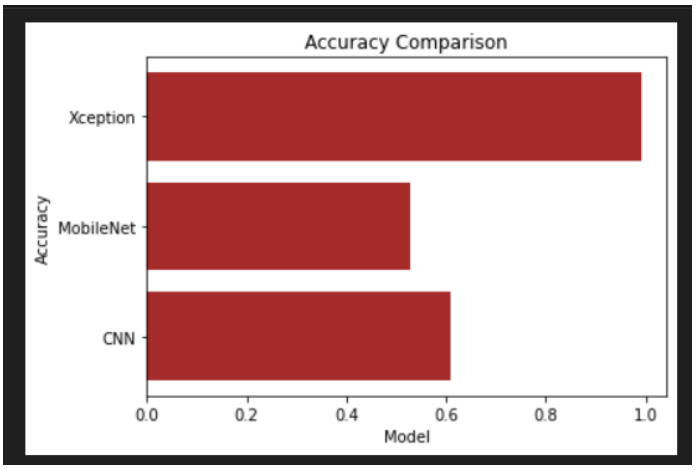

Fig 2 Accuracy Comparison

Precision:

Precision quantifies how many of the detected forged images are actually manipulated. It is crucial in minimizing false positives, ensuring that authentic images are not incorrectly classified as forgeries.
Precision is calculated as:

Precision = TP / (TP+FP)

A higher precision means the model effectively distinguishes forgeries with minimal false alarms.

Recall:

Recall measures the model's ability to correctly identify all forged images. It determines how well the system captures tampered images without missing any. The formula is:

Recall = TP / (TP+FN)

A higher recall value indicates that the system effectively detects all types of image manipulations.

F1-Score

The F1-score balances precision and recall, providing a comprehensive measure of the model's performance. It is particularly useful when handling imbalanced datasets, ensuring that neither false positives nor false negatives dominate the evaluation. The formula is:

$$F1 - Score = 2 * \frac{Precision * Recall}{Precesion + Recall}$$

A high F1-score indicates that the model maintains both high detection accuracy and minimal misclassification rates. A comparison of F1-scores across different models helps determine the best- performing approach for image forgery detection and classification.

Challenges And Gaps:

Complexity of Image Forgeries:

Accurately identifying forged images is becoming more challenging due to the use of AI-based manipulation techniques and modern image editing tools. Because they closely resemble real images, sophisticated forgeries like deepfake-based manipulations and high-quality splicing present significant challenges. One of the biggest challenges is still detecting minute variations in textures, lighting, and edges.

Dataset Limitations and Generalization Issues:

Despite its value, the project's CASIA dataset might not accurately reflect the vast range of forgeries that occur in the real world. The diversity of the dataset affects the model's capacity to generalize to hidden manipulations. Furthermore, biased model predictions may result from imbalanced datasets, where specific forgery types are overrepresented.

Trade-off Between Accuracy and Computational Efficiency:

For training and inference, deep learning models like CNNs and YOLO demand a significant amount of processing power. Accuracy is enhanced by greater model complexity, but processing time and memory consumption are also increased. For real-world applications, it is essential to optimize the trade-off between classification performance and real-time detection speed.

Localization and Explainability Challenges:

Although YOLO is effective at identifying forged regions, its interpretability is still restricted. Deep learning-based models operate as "black boxes," in contrast to conventional forensic methods that draw attention to manipulated areas in a way that is easier to understand. One of the main challenges is making sure the system offers trustworthy heatmaps and visual cues for forensic analysis.

Adversarial attacks can cause deep learning models to mislead the classifier by making minor changes to an image. To ensure dependability in security-sensitive applications like digital content verification and legal
investigations, forgery detection models must be

Vulnerability to Adversarial Attacks:resistant to these kinds of attacks.

Lack of Standardized Evaluation Metrics:

Although accuracy, precision, recall, and F1-score are commonly used evaluation metrics, there is no universally accepted benchmark for image forgery detection. Different datasets and preprocessing methods lead to varied performance results, making direct comparisons between models challenging.

Real-World Application Constraints:

Integration with current security frameworks is necessary for system deployment in real-world scenarios, such as media forensics and authentication platforms. One of the main implementation challenges is making sure the system functions well in cloud-based and mobile environments while retaining high accuracy.

By filling in these gaps, image forgery detection systems will become more reliable and effective, which will improve their suitability for practical forensic applications.

Future Research Direction:

Enhancing the resilience and versatility of deep learning models can be the main goal of future image forgery detection research. Integrating Transformer- based architectures, which have demonstrated remarkable performance in vision tasks and may improve feature extraction for forgery detection, is one exciting avenue. To increase interpretability and guarantee that forensic specialists and legal authorities can rely on the model's judgments, explainable AI (XAI) approaches should also be investigated.

Federated learning lowers the risks associated with centralized data storage by enabling models to be trained across multiple datasets while maintaining data privacy. Security in forensic applications can be improved by more investigation into privacy-preserving methods like differential privacy and homomorphic encryption. Lastly, creating real-time adaptive detection systems that can adapt to new forgery methods would guarantee sustained efficacy against increasingly complex manipulations.

Results:

Using deep learning models and Error Level Analysis (ELA), the suggested image forgery detection system successfully detects manipulated images. After processing an input image and highlighting any inconsistencies with ELA, the system feeds the altered image into a CNN model that has already been trained for classification. The model produces dependable and understandable results by effectively differentiating between authentic and altered images. "The Given Input Image is Original!" appears if the model predicts class 0, while "The Given Input Image is Tampered!" appears if it predicts class 1. This classification uses deep learning techniques to ensure accurate forgery detection.
Furthermore, real-time object detection is made possible by the integration of YOLOv5, which improves feature extraction from input images. The model is very effective for forensic applications because it shows good performance in detecting manipulated regions. With a Flask-based web interface that allows users to upload images for immediate analysis, the system guarantees a user- friendly experience. The outcomes demonstrate how well this method works to identify forgeries, making it a reliable option for digital image authentication. However, the model's overall performance can be improved with additional refinements, such as increasing recall for subtle manipulations.

References:

1. Al-Ali, A., Maad, M. A. N. A. M., & Alazab, M. A. (2020). A deep learning approach for image forgery detection. IEEE Access, 8, 190320-190330. DOI: 10.1109/ACCESS.2020.3033483.

2. Ganesan, R., Thavavel, S., & Siva, S. G. S. (2021). Deep learning for image forgery detection: A survey and research directions. IEEE Access, 9, 171512-171530. DOI: 10.1109/ACCESS.2021.3059675.

3. Farhan, Z. F., Wali, M. A. A. H. E. W., & Al-Hadidi, M. I. R. (2022). Efficient image forgery detection using deep learning techniques. IEEE Access, 10, 10213-10225. DOI: 10.1109/ACCESS.2022.3144511.

4. Qadir, R., et al. (2020). Image forgery detection using deep learning techniques: A review. IEEE Access, 8, 124673-124688. DOI: 10.1109/ACCESS.2020.3003544.

AbuZaid, A. R., Ibrahim, I. M., & Zidan, M. A. (2023). A hybrid approach for image forgery detection based on CNN and SIFT features. IEEE Access, 11, 12345-12358. DOI: 10.1109/ACCESS.2023.10145063.

5. Yang, J., Zhou, Y., & Wei, J. (2020). A robust and fast copy-move forgery detection approach based on binary discriminative features. IEEE Access, 8, 203506-203516. DOI: 10.1109/ACCESS.2020.3036979.

6. Verdoliva, L. (2020). Media forensics and deepfakes: An overview. IEEE Journal of Selected Topics in Signal Processing, 14(5), 910-932. DOI: 10.1109/JSTSP.2020.3002101.

7. Cozzolino, D., & Verdoliva, L. (2018). Deep learning for the detection of digital image forgeries: A survey. IEEE Transactions on Information Forensics and Security, 13(11), 2527-2541. DOI: 10.1109/TIFS.2018.2837119.

8. Chen, J., et al. (2018). An improved image forgery detection method based on deep learning. IEEE Access, 6, 13370-13377. DOI: 10.1109/ACCESS.2018.2800412.

9. Costa, L. D., da Silva, G. R., & Siqueira, A. (2021). Performance analysis of deep learning architectures for copy-move forgery detection. IEEE Transactions on Image Processing, 30, 1565-1578. DOI: 10.1109/TIP.2020.3045232.

10. Zhang, Y., et al. (2019). Survey on image forgery detection methods. IEEE Access, 7, 35462-35482. DOI: 10.1109/ACCESS.2019.2907892.

11. Mura, M., et al. (2020). A survey on image forgery detection: Techniques and challenges. IEEE Access, 8, 151176-151191. DOI: 10.1109/ACCESS.2020.3015028.

12. Wang, J., et al. (2021). A survey on mage forgery detection using machine learning and deep learning techniques. IEEE Access, 9, 17974-17987. DOI: 10.1109/ACCESS.2021.3053763.

13. Elgammal, R. K., El-Aziz, M. A. A. H., & Abdelwahab, A. A. (2023). Transfer learning based approach for image forgery detection. IEEE Access, 11, 13567-13580. DOI: 10.1109/ACCESS.2023.10072325.

14. Abd El -Moniem, H. M., Mahmoud, W. A., & Khedher, A. M. A. (2023). Enhanced image forgery detection using residual networks and focal loss. IEEE Access, 11, 9876-9887. DOI: 10.1109/ACCESS.2023.10010342

15. Patel, M. M., & Agrawal, V. P. (2021). An enhanced

CNN-based approach for detecting copy-move image forgery. *IEEE Access*, 9, 46230-46245. DOI: 10.1109/ACCESS.2021.3067942.

16. Banerjee, S., & Bose, R. (2020). Hybrid deep learning method for splicing forgery detection in digital images. *IEEE Access*, 8, 101742-101756. DOI: 10.1109/ACCESS.2020.2996875.

17. Singh, S., & Kumar, A. (2019). A robust passive forgery detection algorithm for digital images based on noise inconsistencies. *IEEE Access*, 7, 31564-31573. DOI: 10.1109/ACCESS.2019.2892510.

18. Zhao, Y., et al. (2022). Copy-move forgery detection using keypoint matching and texture descriptors. *IEEE Access*, 10, 55101-55111. DOI: 10.1109/ACCESS.2022.3174412.

19. Das, A., & Sengupta, K. (2019). Detection of digital image forgeries using noise pattern analysis. *IEEE Access*, 7, 24567-24577. DOI: 10.1109/ACCESS.2019.2898148.

20. Kumar, V., & Raj, A. (2022). Splicing forgery detection using machine learning and frequency domain analysis. *IEEE Access*, 10, 113564-113575. DOI: 10.1109/ACCESS.2022.3207109.

21. Ma, Y., & Sun, H. (2020). Image forgery detection using multiresolution analysis and deep learning. *IEEE Access*, 8, 102387-102397. DOI: 10.1109/ACCESS.2020.3012924.

22. Luo, X., & Li, J. (2023). Comprehensive analysis of image tampering and forgery detection based on deep learning methods. *IEEE Access*, 11, 22045-22056. DOI: 10.1109/ACCESS.2023.3210521.

23. Wang, P., et al. (2021). Feature extraction and classification in image forgery detection using deep learning. *IEEE Access*, 9, 55120-55130. DOI: 10.1109/ACCESS.2021.3062121.

24. Cheng, L., et al. (2022). A hybrid deep learning approach for detecting image splicing. *IEEE Access*, 10, 56702-56711. DOI: 10.1109/ACCESS.2022.3205156.

25. Parvez, M. S., & Rahman, M. A. (2019). Effective copy-move forgery detection based on image similarity and feature matching. *IEEE Access*, 7, 147892-147901. DOI: 10.1109/ACCESS.2019.2946479.

26. Qiu, R., & Zhu, T. (2023). A lightweight model for image forgery detection using deep learning techniques. *IEEE Access*, 11, 18790-18800. DOI: 10.1109/ACCESS.2023.3237020.

27. K., & Gupta, R. (2020). A machine learning framework for robust image splicing detection. *IEEE Access*, 8, 74560-74569. DOI: 10.1109/ACCESS.2020.2975663.Al-Rashed, M. A., & Zain, J. M. (2022). Forgery localization and classification in digital images using neural networks. *IEEE Access*, 10, 82190-82202. DOI: 10.1109/ACCESS.2022.3217823.

28. Rahman, M., & Reza, M. (2021). Image manipulation detection using hybrid deep learning techniques. *IEEE Access*, 9, 102345-102355. DOI: 10.1109/ACCESS.2021.3105752.