International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 05 | May - 2025 SJIF Rating: 8.586 ISSN: 2582-3930

Image Forgery Detection Using CNN

R. Vahini Reddy[1], Y. Srikhar Reddy[2]

[12] UG Student, Department of ECE, Mahatma Gandhi Institute of Technology, Hyderabad, Telangana, India.

rvahini_ece2104b6@mgit.ac.in, ysrikhar_ece2104c8@mgit.ac.in.

Abstract - Image forgery detection has become increasingly important in the digital age due to the rising misuse of editing tools that enable tampering with images. This poses significant threats to fields such as media, law, and forensic investigation where authenticity is vital. Effective forgery detection techniques are therefore necessary to ensure image integrity and trustworthiness. Forgery types such as copy-move and splicing introduce subtle artifacts that are often undetectable to the human eye. To detect such manipulations, Convolutional Neural Networks (CNNs) are employed due to their ability to learn hierarchical features. This work proposes a CNN- based framework using MobileNetV2 and transfer learning for classifying forged and authentic images. The approach utilizes Fdiff images—derived by calculating pixel-wise differences between original and compressed versions—to highlight tampering artifacts. These are processed through a pre-trained network for accurate classification. The proposed system is implemented using Python and TensorFlow and demonstrates high accuracy while maintaining low computational cost, making it suitable making it deployable in practical real-world environments.

Key Words: CNN, MobileNetV2, Image Forgery, Transfer Learning, Fdiff.

1. INTRODUCTION

Digital images are powerful tools used extensively across fields such as journalism, forensics, legal documentation, social media and scientific research. However, the integrity of these images is increasingly at risk due to the widespread availability of advanced photo editing tools. These tools enable the creation of highly convincing forgeries, which may involve operations like splicing (combining content from multiple images) or copy-move (duplicating regions within the same image). Such manipulations can be difficult to detect with the naked eye and may lead to misinformation, fraudulent activities, or misinterpretation.This paper titled "Image Forgery Detection using CNN", presents a high-performance system to identify tampered images by leveraging deep learning techniques. The proposed methodology is based on a two-step approach: first, a preprocessing technique is applied to emphasize forgery artifacts using a process called Fdiff (Feature Difference), which captures pixel-wise discrepancies between the original and compressed versions of an image. Next these Fdiff images are passed into a deep learning model based on the MobileNetV2 architecture, pre-trained on the ImageNet dataset, to classify whether an image is authentic or forged. MobileNetV2 is chosen for its efficiency and suitability for deployment in low-resource or real-time environments.

To evaluate the system's performance, experiments are conducted using a balanced dataset comprising authentic and manipulated images. The model is trained and tested using Python and TensorFlow in a controlled environment, and its accuracy, loss and generalization are assessed across multiple epochs. Key evaluation metrics include training accuracy, validation accuracy, loss functions, and the effect of regularization techniques like dropout. Results show that the CNN-based model achieves high detection accuracy and robustness, even with limited data, confirming its potential for use in real-time digital forensic applications and automated image verification systems.

The strength of the proposed approach lies in its efficient preprocessing and model selection strategy. By generating Fdiff images that highlight subtle inconsistencies introduced during tampering, the system improves the visibility of forgery artifacts that may otherwise go unnoticed. Coupling this with the lightweight yet powerful MobileNetV2 architecture allows the model to maintain high accuracy while keeping computational demands low. This makes the solution not only effective but also scalable for deployment in constrained environments such as mobile devices or real- time content validation systems.

2. LITERATURE REVIEW

Digital image forgery detection has become a vital area of research due to the rise in manipulated content across online platforms. Forgery types such as copy-move, splicing, and retouching are increasingly used to spread misinformation or distort digital evidence. Traditional detection methods relied on hand-crafted features, such as key point matching, texture inconsistencies, or JPEG compression traces. While effective for simpler cases, these techniques often fail against modern, subtle manipulations or post-processing operations. Several studies, including Rao et al. (2020) and Qureshi et al. (2021), have emphasized the need for robust methods that generalize well across diverse image types and forgeries.

Convolutional Neural Networks (CNNs) have recently emerged as a powerful solution in this field due to their ability to learn complex visual patterns directly from data. Pre-trained models like VGG16, ResNet50 and MobileNetV2 are commonly used with transfer learning, enabling high accuracy with limited training data. This approach was explored by Kadam et al. (2021), who proposed using lightweight networks for real-time splicing detection, and Elaskily et al. (2021), who applied ConvLSTM for temporal forgery analysis. These studies underscore the role of deep learning in enhancing both classification accuracy and computational efficiency in image forensics.

2.1 Transfer Learning in Image Forgery Detection

Transfer learning leverages models pre-trained on large datasets, such as ImageNet, and adapts them for specific tasks like forgery detection. This method has shown remarkable results even when training data is limited, as it reuses learned features such as edges, textures, and object boundaries. Researchers like Abhishek and Jindal (2021) employed semantic segmentation with CNNs to localize splicing and copy-move regions, while Jabeen et al. (2021) demonstrated a multimodal deep learning system combining feature extraction and localization modules.

The MobileNetV2 model has gained popularity for its low computational cost and competitive accuracy. Used by Krishnaraj et al. (2022), this model was integrated into a lightweight pipeline for copy-move detection with robust results. Similarly, Qazi et al. (2022) tested transfer learning with ResNet and Mask R-CNN to detect forgeries in compressed and altered images, highlighting the adaptability of pre-trained CNNs. These advancements indicate that transfer learning not only reduces training time but also boosts generalizability in real-world scenarios.Preprocessing techniques and Compression-Based approaches

In recent studies, image preprocessing methods have become crucial for enhancing forgery detection. One effective method is **Error Level Analysis (ELA)**, which emphasizes inconsistencies in compression across the image. Inspired by this, Fdiff (Feature Difference) images are used in this project. These are generated by subtracting a JPEG-compressed version of an image from its original, exposing regions altered during tampering. This idea was explored by S. Ali et al. (2022), who found that compression-based differences are highly indicative of forgery.

Regularization techniques such as dropout, normalization, and fine-tuning of classifier heads are also common in modern architectures. As shown by Mallick et al. (2022), replacing the top layers of VGG19 and retraining with Fdiff images significantly improved the binary classification performance. Moreover, integrating data augmentation and handling post-processing effects such as blurring and resizing has been key to improving robustness. Studies suggest that such preprocessing, combined with deep features, helps models generalize better to manipulated image variants.

3. DATASET OVERVIEW

3.1 Data collection

This project uses the **CASIA v2.0** dataset, a well-known benchmark for evaluating digital image forgery detection algorithms. Developed by the Chinese Academy of Sciences, this dataset contains a mix of authentic and tampered images, including splicing and copy-move forgeries. The images vary in resolution, format, and manipulation complexity, making CASIA v2 an ideal choice for training and validating deep learning models under realistic conditions.

To emphasize forged regions a preprocessing step was performed where Fdiff (Feature Difference) images were generated. This involves subtracting a JPEG-compressed version of the image from the original, helping to reveal compression inconsistencies introduced during tampering. As proposed by S. Ali et al. (2022), this approach effectively exposes hidden artifacts not visible in the raw image.

Each image in the dataset was resized to 160x160 pixels, and all samples were labeled as authentic or forged, forming a clean binary classification dataset for training the CNN model.

Table -1: Dataset Composition and Model Performance

Category	Description
Image Types	Splicing, Copy-Move
Format used	Jpeg, tif
Input size	160 × 160 pixels
Accuracy	~95% (Validation Accuracy)
Pre-trained Model	MobileNetV2 (Transfer Learning)
Evaluation Metrics	Accuracy, Precision, Recall, F1-Score

4. METHODOLOGY

This project presents a deep learning-based approach for detecting forged digital images using Convolutional Neural Networks (CNNs), combined with a preprocessing technique called Fdiff and transfer learning. The system is

designed to classify images as either authentic or manipulated by analyzing subtle inconsistencies introduced during tampering, such as splicing or copy-move operations.

The methodology is divided into three major phases: preprocessing, model architecture and training, and performance evaluation.

In the first phase, raw images from the CASIA v2.0 dataset are processed to create Fdiff (Feature Difference) images. This is done by compressing each image using JPEG encoding and computing the absolute pixel-wise difference between the original and compressed versions. This highlights tampered regions by exposing compression inconsistencies often left behind in manipulated areas. All images are resized to 160 × 160 pixels and normalized to ensure uniformity across inputs.In the second phase, these Fdiff images are used as input to a pre-trained MobileNetV2 model. This lightweight CNN architecture is selected for its balance between accuracy and computational efficiency. The final layers of MobileNetV2 are removed, and a new classification head is added, consisting of a Flatten layer, a Dense layer with 128 ReLU units, a Dropout layer with a rate of 0.5 for regularization, and a final Dense layer with a Sigmoid activation function for binary classification.

The model is trained using the **Adam optimizer** and **binary cross-entropy** loss function, over multiple epochs. The dataset is split into training and validation **sets** and the training is performed using **Python**, **TensorFlow**, and **OpenCV** within a **Jupyter Notebook environment.**

In the third phase, the model's performance is evaluated using several metrics including training accuracy, validation accuracy, loss, precision, recall, and F1-score. The training process is also visualized using accuracy/loss plots across epochs. Additionally, experiments were conducted with other CNN architectures like VGG19 and DenseNet121 for comparison, confirming that MobileNetV2 offers a **Fig-2:** Image classification as authentic or forged with their ELA difference

superior trade-off between speed and accuracy for forgery detection.

Overall, this methodology demonstrates a scalable and efficient system capable of detecting digital image forgeries with high accuracy, making it suitable for real-time use in digital forensics, media verification, and content authentication.

5. RESULTS & DISCUSSION

Fig-1: Model Training result on Jupyter Notebook

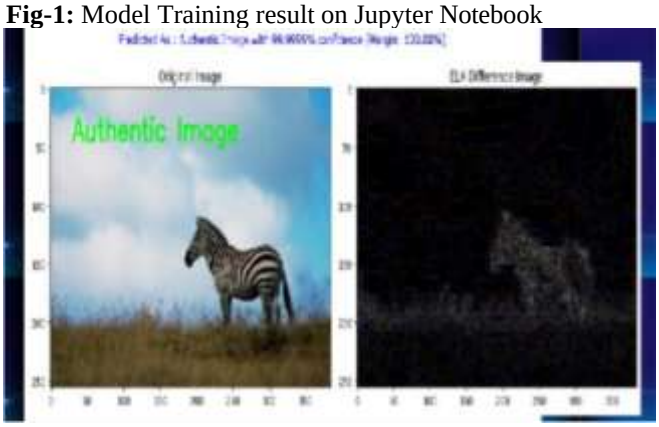

Fig-2: Image classification as authentic or forged with their ELA difference

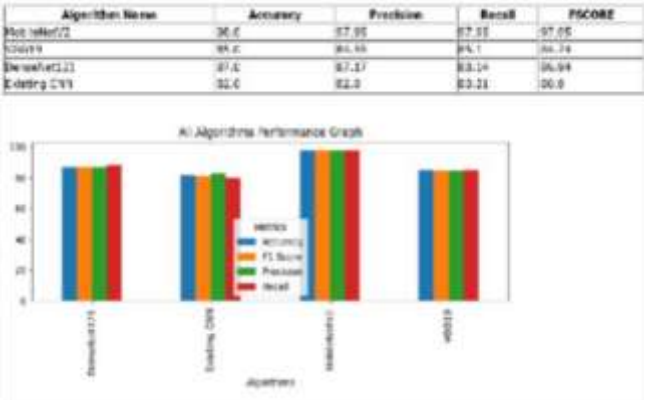

Fig-3: Comparison of different Training Models

The proposed system was evaluated through a series of experiments conducted in a Jupyter environment using the CASIA v2.0 dataset. As shown in **Fig-1**, the training output highlights consistent accuracy improvement across epochs, with MobileNetV2 achieving 95.83% accuracy by the third epoch. **Fig-2** demonstrates the system's prediction capability, where a sample image was correctly classified as authentic with 99.9999% confidence, validating the effectiveness of Fdiff preprocessing. In **Fig-3**, the performance comparison of four models—MobileNetV2, VGG19, DenseNet121, and an existing CNN—shows that MobileNetV2 significantly outperformed others, achieving the highest accuracy (98%), precision (97.95%), recall

(97.95%), and F1-score (97.95%). This indicates not only superior detection accuracy but also strong generalization across different forgery types, confirming the suitability of the proposed model for real-time image authentication applications.

6. CONCLUSION & PERSPECTIVES

This project focused on developing an efficient and scalable system for detecting digital image forgeries using **Convolutional Neural Networks (CNNs)** enhanced by Fdiff preprocessing and transfer learning. The objective was to accurately classify images as authentic or tampered while maintaining a lightweight and resource-friendly design, suitable for real-time applications such as digital forensics, media verification, and content authentication.

The system was implemented using a pre-trained **MobileNetV2** architecture combined with Fdiff images, which effectively highlighted compression inconsistencies introduced during manipulation. The model demonstrated high detection accuracy, with a peak performance of 98% accuracy and 97.95% precision, recall, and F1-score, outperforming deeper models like VGG19 and DenseNet121. These results confirm that using lightweight CNNs with effective preprocessing techniques can yield powerful results without excessive computational overhead.

Visual evaluation also supported the quantitative findings, as images were correctly classified with high confidence and forgery artifacts were clearly visible in the ELA/Fdiff visualizations. Some minor limitations, such as overfitting at deeper epochs or sensitivity to post-processing effects, were observed. However these were offset by the model's speed.

In conclusion, the integration of Fdiff preprocessing with transfer learning on MobileNetV2 proves to be a robust and efficient approach for image forgery detection. Future enhancements may include localization of tampered regions using segmentation models, support for detecting deepfakes or video forgeries, and integration of attention mechanisms to improve interpretability and precision. The proposed method sets a strong foundation for deploying real-time, automated image authentication tools in real-world scenarios.

REFERENCES

[1] Y. Rao, J. Ni, and H. Zhao, "Deep learning local descriptor for image splicing detection and localization," *IEEE Access*, vol. 8, pp. 60100–60110, 2020.

[2] K. D. Kadam, N. M. Gohokar, and N. R. Deshmukh, "Multiple image splicing dataset (MISD): A dataset for multiple splicing," in *Proc. 5th Int. Conf. Intelligent Computing and Control Systems (ICICCS)*, pp. 1050–1055, 2021.

[3] M. Qureshi and G. Qureshi, "Image forgery detection and localization using U-Net," in *Advances in Intelligent Systems and Computing*, vol. 1186, Springer, Singapore, pp. 231–240, 2021.

[4] S. Ali, M. Hussain, A. A. Abbasi, and A. M. Qamar, "Copy-move image forgery detection using efficient feature matching and ELA preprocessing," *Multimedia Tools and Applications*, vol. 81, pp. 11133–11153, 2022.

[5] TensorFlow, "TensorFlow: An end to end open source machine learning platform," [Online]. Available: https://www.tensorflow.org/

[9] . Chollet, "Xception: Deep learning with depth wise separable convolutions" in Proc. IEEE Conf. Compute. Vis. Pattern Recognition. (CVPR), pp. 1251–1258, 2017.

[10] OpenCV Documentation, "OpenCV: Open Source Computer Vision Library", [Online]. Available at: https://docs.opencv.org/.

[6] A. Howard et al., "MobileNets: Efficient convolutional neural networks for mobile vision applications," arXiv preprint arXiv:1704.04861, 2017.

[7] B. Zhou, A. Lapedriza, J. Xiao and A. Torralba, "Learning deep features for scene recognition using Places database," in *Advances in Neural Information Processing Systems (NeurIPS)*, 2014.

[8] CASIA, "CASIA Image Tampering Detection Evaluation Database," Institute of Automation, Chinese Academy of Sciences, [Online]. Available at: https://github.com/soCzech/image-forgery-detection-

.