International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

Image Forgery Detection Using CNN Transfer Learning

Dr.T.Kavitha[1], Eshwar Veeramalla[2], Manoj Kumar J[3], Jedidya Trividhi[4]
[1]*Associate Professor, MVSR Engineering College (A), Nadergul, Hyderabad, Telangana, India.*
[2,3,4]*B E Students, Dept. of Electronics & Communication Engineering,Nadergul, Hyderabad, Telangana, India.*

Abstract—Digital image manipulation has become increasingly accessible due to the rapid development of editing tools, raising serious concerns about the authenticity of visual content across domains such as media, legal investigations, and online communication. Many existing detection methods are limited to binary classification and focus on specific manipulation types, often overlooking subtle edits like retouching. In this work, we present a deep learning-based approach that can distinguish between multiple types of image forgeries. The proposed model uses a dual-stream architecture that combines EfficientNet-B3 for understanding visual content with a Spatial Rich Model (SRM). The combined feature representation is passed through fully connected layers and a Softmax classifier to categorize images into four classes: real, copy-move, splicing, and retouching. Transfer learning enhances model generalization, while Grad-CAM is employed to provide visual explanations of detected forgery regions. Experimental results demonstrate strong performance, achieving a test accuracy of 87.55% and a macro F1-score of 84.73%, indicating the effectiveness of the proposed approach.

Keywords— Image Forgery Detection, CNN, Transfer Learning, EfficientNet-B3, Copy-Move, Splicing, Retouching, SRM, Grad-CAM, Digital Forensics.

Introduction

In earlier times, producing a realistic manipulated image required specialized software and significant effort. Today, however, image editing can be performed quickly using widely available tools, even on mobile devices. This ease of manipulation has reduced the reliability of visual content, making it difficult to trust images used in critical areas such as journalism, digital media etc. The consequences of manipulated images range from spreading misinformation to influencing public opinion and legal outcomes.

The types of manipulations are mainly divided into three major groups. In copy-move forgery, a certain area from one location in the image is copied and then moved to another location in the same image, normally for the purpose of concealing something or making the scene seem more populated [1]. Splicing involves combining contents of two or more images into one frame, none of which were actually taken [2]. Retouching is the least obvious of the three kinds of manipulations and involves small edits in luminance and/or color values of an image to change its message without drawing too much attention [3].

Earlier approaches depended on manually designed algorithms tailored to specific manipulation types. Block-matching techniques can be used to detect the presence of duplicate blocks often used during copy-and-paste operations [4]. Methods operating in the frequency domain analyze inconsistencies introduced during image composition, but these signals can weaken after further processing. [5]. The compression level verification technique is capable of identifying parts of an image that have been saved with a lower compression level [6]. Although these techniques are still effective, they are all susceptible to the same problem: a simple post-processing procedure like saving the image again, scaling it, or adding noise is sufficient to make them ineffective.

Recent progress in deep learning has significantly changed the way image forgery detection is performed. CNNs learn these characteristics from the labelled data and pick up subtleties that cannot be programmed into the system manually [7]. The use of transfer learning takes this one step further and uses the pre-trained knowledge gained by the network through millions of general

DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

images, thereby ensuring a wealth of visual information for forgery detection with minimal specific data required for excellent performance [8].

The effectiveness of deep learning algorithms in detecting copy-move and splicing forgery has been shown by Jain et al. [9] and Abdalla et al. However, both these systems fail to account for retouching, and instead generate a binary result on whether the image is real or fake.

This work addresses those gaps directly. A single trained model is built that classifies an input image into one of four categories: real, copy-move, splicing, or retouching. The system combines EfficientNet-B3 with an SRM residual branch to capture both high-level visual features and pixel-level inconsistencies, trained on a mixed corpus of synthetic and benchmark data. A Grad-CAM desktop interface provides spatial explanations for every prediction. The proposed system outperforms both base-paper reference models and is, to our knowledge, the first system in this comparison to handle all three forgery types within a unified four-class framework.

Literature Review

Traditional Detection Methods

Early research focused on manually engineered detection techniques. Techniques such as block matching attempt to identify repeated regions within an image, which can indicate copy-move manipulation, though they often struggle when transformations like scaling or rotation are applied. Frequency-based techniques analyzed compression artifacts to identify splicing boundaries, though these signals could be lost after image processing. Error Level Analysis (ELA) offered a simple way to detect inconsistencies, but its effectiveness was limited in uniformly compressed images.

Deep Learning Approaches

More recent work has explored the use of convolutional neural networks for forgery detection. Some models are designed to learn residual noise patterns rather than image content, which helps in identifying manipulations. Dual-stream architectures have also been proposed, where one branch processes the original image while another focuses on filtered or residual information. These approaches have shown promising results, particularly for copy-move and splicing detection, but many of them are still limited to binary classification.

Drawbacks of Existing System

Narrow set of forgery categories:
Most models focus only on copy-move and splicing, ignoring retouching.

Binary classification results:
Some models only classify images as real or fake without identifying the forgery type.

Vulnerability to post-processing:
Simple operations like resizing or noise addition can remove important forgery traces.

Lack of transfer learning:
Many CNN models are trained from scratch, requiring large datasets.

Failure to achieve interpretability:
Models do not explain their decisions, limiting their use in forensic analysis.

Proposed System and Methodology

In this work, we design a dual-stream convolutional architecture that combines semantic and noise-based analysis for improved forgery detection, which employs EfficientNet-B3 along with an SRM residual stream to obtain both semantic and noise information from input images. In our approach, SRM filters are used to emphasize subtle noise variations that may reveal hidden manipulation traces. In other words, EfficientNet-B3 is used as the primary feature extractor due to its ability to capture meaningful visual patterns efficiently, while the SRM residual stream improves the appearance of input images with the help of applying high pass filters. Next, the outputs from both branches are combined into a single representation. This combined feature vector is passed through a fully connected layer, followed by a Softmax classifier that assigns the image to one of four categories: real, copy-move, splicing, or retouching.,

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

How the System Works

Before training, all images are adjusted to a consistent size and scaled to ensure stable input for the model. With data augmentation techniques like flipping, rotation, and color jittering applied only during training to improve generalization. The processed images are fed into a CNN architecture combining EfficientNet-B3 and an SRM residual branch. EfficientNet-B3 extracts high-level visual features, while SRM captures pixel-level inconsistencies using high-pass filtering. The combined features are passed through a fully connected layer and classified into four classes using a Softmax classifier. Grad-CAM visualization is used to highlight important regions influencing the model's decision.

For complete flow chart, refer to fig.1

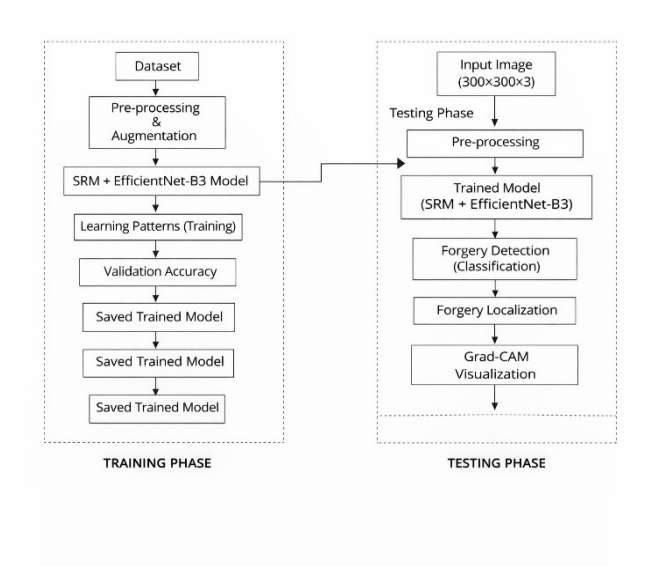

Fig. 1. Overall System Block Diagram

Training Phase

1. Dataset

A balanced dataset was constructed across four forgery categories through careful planning and organization. Authentic images were sourced from well-established forensic benchmark datasets, including MICC-F220, IMD2020, Columbia Uncompressed Image Splicing, and Columbia Color Splicing.

2. Pre-processing & Data Augmentation

Pre-processing involves resizing images to a fixed dimension (e.g., 300×300) to match the input requirements of the model. Data augmentation techniques such as rotation, flipping, cropping, and color jittering are used to artificially increase dataset size. These techniques help in improving model robustness and preventing overfitting.

3. SRM + EfficientNet-B3 Model

In this block, both spatial and noise-based features are extracted using a dual-stream approach. SRM (Spatial Rich Model) filters are applied to capture pixel-level noise and manipulation artifacts. EfficientNet-B3 processes the RGB image to learn high-level semantic features. These two feature types complement each other for better forgery detection.

4. Learning Patterns (Training)

During this phase, the model learns to map input images to their corresponding labels. During training, the model iteratively updates its parameters based on prediction errors to improve accuracy over time.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

The process is repeated over multiple epochs to improve accuracy. Gradually, the model learns discriminative features for different forgery types.

5. Validation Accuracy

Validation is performed using a separate dataset not seen during training. It helps evaluate how well the model generalizes to new data. After each epoch, validation accuracy is calculated to monitor performance. This step is crucial to detect overfitting or underfitting. A higher validation accuracy indicates better generalization.

6. Saved Trained Model

After training, the best-performing model is saved for future use. The model is typically saved when validation accuracy improves. This ensures that the most optimal version is preserved.

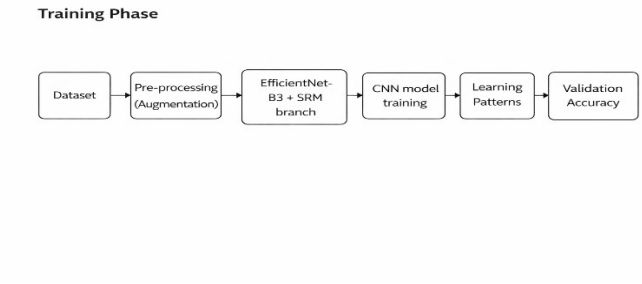

Fig. 2. Training Phase

Transfer Learning based CNN Architecture

The structure of the CNN, depicted in Fig. 3, is aimed at integrating semantic and forensic information. The backbone of the model is EfficientNet B3, which has been optimized for detecting subtle manipulations. The SRM residual stream uses high-pass filters to filter out all information related to the content of the scene, leaving only image inconsistent information.

The feature vectors are combined and fed to the Fully Connected layer, which contains 512 neurons. Dropout is used to prevent overfitting with a rate of 0.4. After that, the Softmax activation function generates four outputs, each corresponding to one class, namely real, copy move, splicing, and retouching.

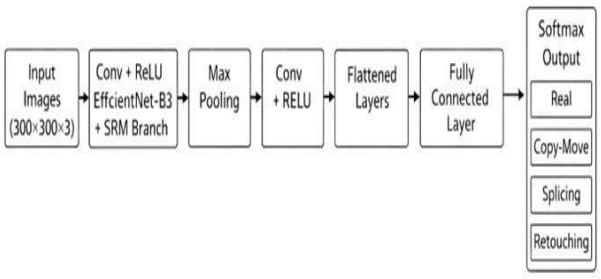

Fig. 3. CNN Architecture

Testing Phase (fig 4)

1. Input Image (300×300×3)

The input image is provided to the system for forgery analysis. It can be any real-world image whose authenticity needs to be verified. The image is resized to match the model's required input dimensions.

2. Pre-processing

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

The input image is resized and normalized to maintain consistency with the training phase. Pixel values are scaled to improve model performance and stability. This step ensures uniform input conditions for accurate predictions. No augmentation is applied during testing.

3. Trained Model (SRM + EfficientNet-B3)

The preprocessed image is passed through the trained dual-stream model. EfficientNet-B3 extracts high-level visual features from the RGB image. The SRM branch captures noise and manipulation artifacts using high-pass filtering. Both features are combined for robust prediction.

4. Forgery Detection

The model predicts the class label of the input image. It classifies the image into one of four categories: real, copy-move, splicing, or retouching. A Softmax layer is used to generate probability scores for each class. The class with the highest probability is selected as the final output.

5. Forgery Localization

This block identifies the specific regions in the image that are likely manipulated. It helps in understanding where the forgery has occurred. Localization is important for visual interpretation and validation.

6. Grad-CAM Visualization

Grad-CAM generates a heatmap highlighting important regions influencing the model's decision. It overlays the heatmap on the original image for better visualization. This improves model interpretability and trust

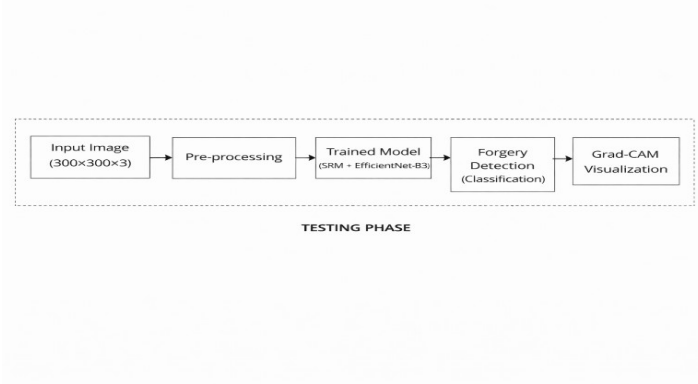

Fig. 4. Testing phase

Dataset and Experimental Setup

Where the Data Came From

Building a balanced dataset across four forgery categories requires careful planning. Authentic images were drawn from four established forensic benchmarks: MICC-F220,IMD2020,Columbia Uncompressed Image Splicing, and Columbia Color Splicing. Critically, the split into training, validation, and test partitions was done before generating any forged examples.

Forged samples were then generated from the training split only. Copy-move examples were produced by extracting a region from a training image and pasting it back into the same image at a different position, with random rotation and scaling applied to increase variety. Spliced examples were made by compositing a patch from one image into a different image. Retouching examples were created by applying randomised localised photometric edits — brightness, contrast, saturation, and sharpness — to sub-regions of authentic images. Each of the four classes ended up with a balanced number of samples.

TABLE I. DATASET CLASS SOURCES

Class	Main Source Datasets	How Samples Were Made
Real	Columbia Uncomp.,	Original unaltered

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

Class	Main Source Datasets	How Samples Were Made
	IMD2020, MICC-F220	images
Copy-Move	MICC-F220 + Synthetic	Region copied within same image
Splicing	Columbia Color Splicing + Synthetic	Region pasted from another image
Retouching	IMD2020 + Synthetic	Local pixel edits applied

Training Settings

The experimental setup and training configuration, including optimizer settings, learning rate schedule, batch size, and regularization techniques, are summarized in detail in the table II. All implementation details such as AMP usage, label smoothing, and class-balanced sampling are also included.

TABLE II. MODEL TRAINING CONFIGURATION

Setting	Value
Framework	PyTorch + torchvision
Backbone	EfficientNet-B3 (ImageNet pre-trained)
Input size	300 × 300 × 3
Optimiser	Adam
Learning rate	0.0001 with cosine annealing
Label smoothing	0.1
Batch size	32 (balanced per class)
Fully connected	512 units, ReLU, Dropout 0.4
Output	4-class Softmax
Best epoch	11 of 14
Training platform	Modal cloud GPU

Results and Discussion

Overall Results

The checkpoint from epoch 11 was used to evaluate the model's performance on the hold-out test set, which it had never been exposed to during training or tuning. This resulted in an accuracy rate of 87.55% and a macro F1 score of 0.8473 for the four classes combined. The validation metrics at this checkpoint were accuracy of 88.29% and macro F1 of 0.8570. The slight difference shows that the model is not overfitting.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

TABLE III. FINAL MODEL PERFORMANCE

Metric	Validation	Test
Accuracy (%)	88.29	87.55
Macro F1-Score	0.8570	0.8473
Best Epoch	11	11
Total Epochs	14	14

Per-Class Results
The per – class (real, cpoymove, splicing, retouching) f1 scores are displayed in the below table IV

TABLE IV. PER-CLASS TEST F1-SCORES

Class	F1-Score	Notes
Real	0.9125	Easiest; consistent natural noise structure
Splicing	0.8637	Cross-image noise mismatch aids detection
Retouching	0.8337	SRM branch captures localised noise changes
Copy-Move	0.7794	Hardest; duplicated patch shares same noise as source
Macro Average	0.8473	Even performance across all four classes

User Interface and Output Demonstration

The project includes a user-friendly graphical interface as shown in fig.5 designed for easy interaction and analysis. The UI provides options such as selecting a model, choosing an input image, and initiating the analysis through an “Analyze Image” button. Once an image is uploaded, the system displays prediction results along with confidence scores for each forgery class. Additionally, the interface shows the selected image and generates a Grad-CAM visualization to highlight important regions. This intuitive design allows users to efficiently perform forgery detection and interpret the model’s decisions.

We tested the images of each class and analysed them. Below are 4 cases representing each class (real, copymove, splicing, retouching).

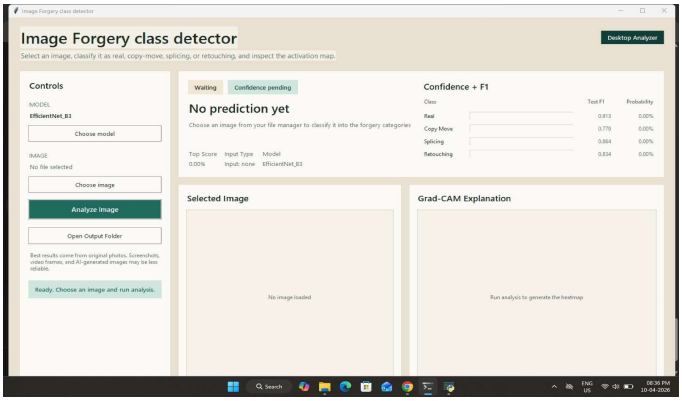

Fig. 5. User Interface

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

Case 1: Real

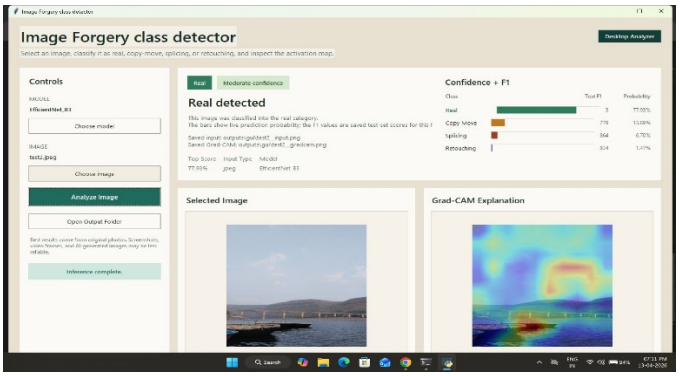

Fig. 6 Output showing - Real Image Detected

The system classified the photo(fig.6) as real, indicated high probability of around 80% and generated a GRAD CAM visualisation

Case 2: Copy- Move

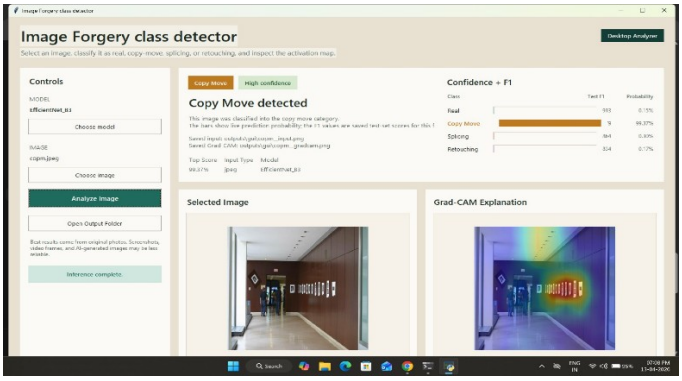

Fig. 7. Output showing – Copy Move Forgery Detected

A duplicated region within the same image is detected(fig.7), and the heatmap highlights the pasted area with high confidence.

Case 3: Splicing

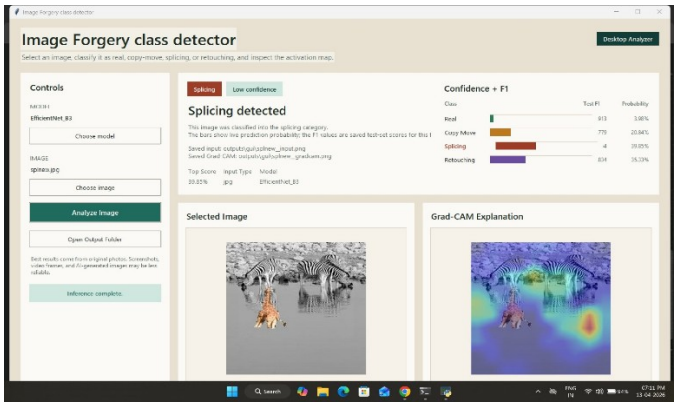

Fig. 8 Output showing – Splicing Forgery Detected

The model identifies splicing(fig.8), with the Grad- CAM focusing on the boundary where content from different sources was joined.

International Journal of Engineering Technology and Management Sciences
Website: ijetms.in Issue:2 Volume No.10 March - April – 2026
DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

Case 4:Retouching

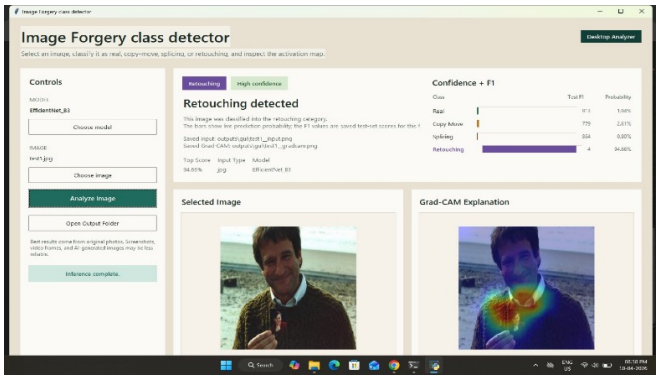

Fig. 9 Output showing – Retouching Detected

Local pixel edits are detected(fig.9), and the heatmap concentrates on the altered sub- region, confirming the retouching manipulation.

Comparison with Base-Paper Models

In Table V, the comparison between the three models is displayed. In case a model cannot perform detection of a particular type of forgery, a dash is put in place of a cell, which shows the inability of the model rather than its score of zero. The paper of Jain et al. [1] provides the highest accuracy of 92.3%; however, this result was obtained when solving a simpler two-class problem where copy-move and splicing are differentiated as fake vs real. For Abdalla et al. [2], the F1 score on the same copy-move problem equals 0.8835. The model we propose, working with four classes, gives 87.55% accuracy and F1 = 0.8473 macro-average; moreover, it is the only model of the three able to distinguish retouched images from real or copy-moved ones.

TABLE V. COMPARISON WITH BASE-PAPER MODELS

Metric	Jain et al. [1]	Abdalla et al. [2]	Proposed Model
Task	Binary (real vs fake)	Binary (real vs fake)	4-class
Accuracy (%)	92.3	88.26–90.1 (val)	87.55 (test)
Macro F1	0.9110	0.8835 (copy-move)	0.8473
Copy-Move F1	—	0.8835	0.7794
Splicing F1	—	—	0.8637
Retouching F1	—	—	0.8337
Real F1	—	—	0.9125
Transfer Learning	No	No	Yes
Explanation	Binary Mask	Binary Mask	Grad-CAM

DOI:10.46647/ijetms.2026.v10i02.035 ISSN: 2581-4621

Metric	Jain et al. [1]	Abdalla et al. [2]	Proposed Model
UI	No	No	Yes
Forgery Types	Copy-Move, Splicing	Copy-Move only	All 3 types
Confidence score	Binary	Binary	4-class probability

Conclusion

The proposed approach introduced a two-stream architecture utilizing the EfficientNet-B3 and SRM residual branches for detection of image forgeries. With the use of the ImageNet transfer learning capabilities of the EfficientNet-B3 backbone network and forensics traces obtained from the SRM branch, the model achieved balance in its performance on four different forgery classes: real, copy-move, splicing, and retouching. Finally, adding the use of Grad-CAM visualization improved the interpretability by providing explanations behind predictions.

Compared to baseline papers' solutions, the proposed system is unique in unifying the detection of all three main forgery types in one architecture and providing the use of a user-friendly Python-based interface for demonstrating the approach.

References

[1] S. Jain, P. Rajpoot, and T. Yadav, "Deep Learning-Based Image Forgery Detection Using CNN and UNet for Precise Tampered Region Identification," Int. J. Advanced Research in Computer and Communication Engineering, vol. 14, no. 3, pp. 131-141, Mar. 2025, doi: 10.17148/IJARCCE.2025.14316.

[2] Y. Abdalla, M. T. Iqbal, and M. Shehata, "Convolutional Neural Network for Copy-Move Forgery Detection," Symmetry, vol. 11, no. 10, p. 1280, Oct. 2019, doi: 10.3390/sym11101280.

[3] D. Coccomini, N. Messina, C. Gennaro, and F. Falchi, "Combining EfficientNet and Vision Transformers for Video Deepfake Detection," Proc. ICIAP, 2022, arXiv:2107.02612.

[4] R. Caldelli, L. Galteri, I. Amerini, and A. Del Bimbo, "Optical Flow Based CNN for Detection of Unlearnt Deepfake Manipulations," Pattern Recognition Letters, vol. 146, pp. 31-37, 2021.

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

[6] 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, 2016.

[7] M. Tan and Q. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," Proc. ICML, vol. 97, pp. 6105-6114, 2019.

[8] J. Fridrich, D. Soukal, and J. Lukas, "Detection of Copy-Move Forgery in Digital Images," Proc. Digital Forensic Research Workshop, Aug. 2003.

[9] Y. Q. Shi, C. Chen, and W. Chen, "A Natural Image Model Approach to Splicing Detection," Proc. ACM Workshop on Multimedia and Security, 2007.

[10] I. Amerini et al., "A SIFT-Based Forensic Method for Copy-Move Attack Detection," IEEE Trans. Information Forensics and Security, vol. 6, no. 3, pp. 1099-1110, 2011.

[11] G. Zampoglou, S. Papadopoulos, and Y. Kompatsiaris, "Detecting Image Splicing in the Wild," Proc. IEEE ICMEW, 2015