International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Framework for Pixel-Level Image Forgery Localization and Classification: A UNet and EfficientNet-B0 Approach

Shravani Nilesh Patil
Vibhavari Yashwant Chandrachud
Rohan Saidas Rathod
Nikhil Sanjay Gaikwad
Department of Computer Engineering
Savitribai Phule Pune University (SPPU), Nashik, India

Abstract

In today's world it is very easy to manipulate images with software and artificial intelligence. This means that we can no longer trust what we see in pictures and videos. Digital image forgery-ranging from semantic alterations like splicing and copy-move to the generation of entirely synthetic "deepfake" content-poses a critical threat to forensic science, judicial integrity, and public trust. This paper presents a comprehensive, multi-tiered forensic system designed to address the dual challenges of binary classification and precise localization. Our methodology integrates a hybrid deep learning architecture: a UNet-based segmentation engine for pixel-level localization and an EfficientNet-BO backbone for multi-class forgery identification.

To make the system more sensitive to subtle changes, we implemented an Error Level Analysis(ELA) preprocessing stage. This helps the model see inconsistencies in JPEG Compression that are usually invisible to the human eye. We have trained our model using transfer learning on various datasets such as CASIA and CoMoFoD, to ensure that it can handle various types of tampering such as splicing, copy-move or AI-generated edits. To make the research practical we also built a full stack web application. We include AI outputs through GRAD-CAM heatmaps, which allow investigators to see exactly what led the model to its conclusion. Our analysis of sixteen contemporary studies shows that combining pixel-wise segmentation with optimized compound scaling provides a significantly stronger defense against modern digital deception than traditional methods.

Keywords: Image Forgery Detection, Digital Forensics, Deep Learning, UNet, EfficientNet-B0.

1 Introduction

Visual information is the backbone of how we communicate today. We trust what we see in news reports, social media or in legal proceedings. However that trust is being challenged by the ease with which digital images can be manipulated. Whether its simple "copy-move" edit to hide an object or a complex "splicing" task designed to alter the

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

semantic meaning of a scene, image forgery is now a major challenge in digital forensics. More recently, the emergence of AI-generated content and "deepfakes" has complicated the field further, as these images are often times not detected by traditional methods.

Historically, forensic experts relied on handcrafted features, such as analyzing sensor noise patterns or JPEG compression. While these methods were brilliant in their way, they are often too fragile for today's forgeries. If a forger applies a simple anti-forensic measure like slight blur-those traditional methods often disappear. Our research recognized that we can no longer on these features. Instead, we turned to deep learning. Convolutional Neural Networks(CNNs) are capable of learning the complex, non linear "rules" of a natural photograph, allowing them to spo even the most subtle pixel-level anomalies.

For a professional investigator, just knowing an image is "fake" is not enough; they need to see exactly where the tampering occurred. To address this, we integrated UNet, a model originally designed for medical image segmentation. Its symmetric architecture allows it to capture the global context of a photo while maintaining high-resolution detail, resulting in a precise "tamper mask."

Our project is more than just a mathematical model, it's functional, full-stack application utilizing Flask to build a secure backend, SQLite to manage data, creating an environment where an investigators can upload a file and receive a report. By incorporating Grad-Cam and Error Level Analysis, we ensure our system is a transparent tool that can explain its findings. This paper provides a deep analysis of current literature, explains our hybrid methodologies and discusses the results of our implementation.

2 Literature Review

The field of digital forensics has evolved rapidly with the advent of Convolutional Neural Networks(CNNs). We conducted a thorough review of sixteen key research contributions to understand the current technological landscape and identify landscape and identify research gaps.

2.1 Technical Synthesis

The shift towards deep kearning is well-documented in the survery by zanardelli et al.(2023), who identified that CNNs are now the only viable way to keep up with GAN-generated forgeries. A breakthrough in efficiency was provided by Korsipati et al. (2025), who utilized EfficientNetV2 to achieve near-perfect results on benchmark datasets, proving that attention-based models are superior for localized tampering.

A recurring theme in our reviewed papers is the challenge of "data scarcity." High-quality, labeled forgery datasets are hard to come by. Researchers such as Jonker et al. (2024) and Qazi et al. (2022) successfully addressed this through Transfer Learning. By taking a model already trained on millions of natural images (ImageNet) and fine-tuning it for forensics, they achieved higher accuracy than training from scratch. Our project adopts this strategy to ensure our system remains robust even when faced with unseen image sources.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Author (Year)	Core Methodology	Datasets Used	Key Findings	Research Gaps
Jonker et al. (2024)	Transfer Learning	Multimedia	High precision in edits	Dataset specificity
Korsipati et al. (2025)	EfficientNetV2 + SE	CASIA/NIST	AUC up to 1.000	Computationally heavy
Qazi et al. (2022)	CNN-based Deep Learning	Benchmark sets	Effective detection	Single forgery focus
Shallal et al. (2025)	Copy-Move Overview	Diverse sets	Localization priority	Compression issues
Khalaf et al. (2024)	CNN + Blockchain	Public datasets	Integrity tracking	High infrastructure cost
Liang et al. (2025)	Soft Contrastive Loss	Real/Fake mixed	Robust against GANs	Needs real traces
Rehman et al. (2025)	CNN + SVM Hybrid	Patch-based	Stable decision	High training time

Table 1: Comparative Literature Review

3 Methodology

Our research focused on building a multi-tier pipeline that combines traditional forensic preprocessing with modern deep learning. We prioritized "explainability" and "localization" as the core features of the system.

3.1 Error Level Analysis (ELA) Preprocessing

The first line of defense in our system is ELA. When a JPEG image is saved, the entire frame is compressed at a uniform rate. If a forger inserts an object and saves it again, that new part will have a different compression history. We calculate the absolute difference between the original pixel $P_{i,j}$ and a version re-saved at a known quality $P'_{i,j}$ (4):

$$E_{i,j} = |P_{i,j} - P'_{i,j}| \times \gamma$$

The resulting ELA map provides a high-frequency visualization where tampered regions appear noticeably brighter, serving as a critical feature for our neural networks.

3.2 UNet Localization Engine

To solve the localization problem, we used UNet. We designed the architecture with a contracting path to extract feature maps and a symmetric expanding path that uses skip connections to reconstruct the tampered region. This allows the model to map features x_i directly to the output y_{up} (13):

$$y_{\text{up}} = \text{Concat}\big(f_{\text{up}}(x_i),\, x_{\text{skip}}\big)$$

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

This skip-connection mechanism ensures that the boundary of the forgery remains sharp and accurate in the final output mask.

3.3 EfficientNet-B0 Classification

For identifying the type of forgery, we chose EfficientNet-B0. Its core strength is Compound Scaling, which scales the depth (d), width (w), and resolution (r) using a single coefficient ϕ (12):

$$d = \alpha^{\phi}, \quad w = \beta^{\phi}, \quad r = \gamma^{\phi}$$

This mathematical optimization allows the model to capture fine-grained textures while remaining lightweight enough to run on a standard web server.

3.4 Patch Based Analysis

High-resolution images often mask tiny forgeries during the downsampling phase of a CNN. To combat this, we implemented a Patch-Based Detection strategy (10). We split the image into 224×224 patches and analyze each individually. This ensures that a localized edit—like a small face swap—is treated with the same importance as a large-scale background change.

4 System Architechture

To transition our theoretical models into a practical tool, our team designed a streamlined, modular architecture that facilitates real-time forensic auditing. The architecture, illustrated in Fig. 1, is structured to handle the journey of a digital image from the user's browser to our deep learning engine with minimal latency.The workflow begins at the Web Browser layer, where we implemented a React-based interface to manage image uploads via HTTP POST requests. Once the Flask Backend receives the file, it initiates a validation sequence to ensure data integrity. The "heart" of our system lies in the preprocessing and inference pipeline.

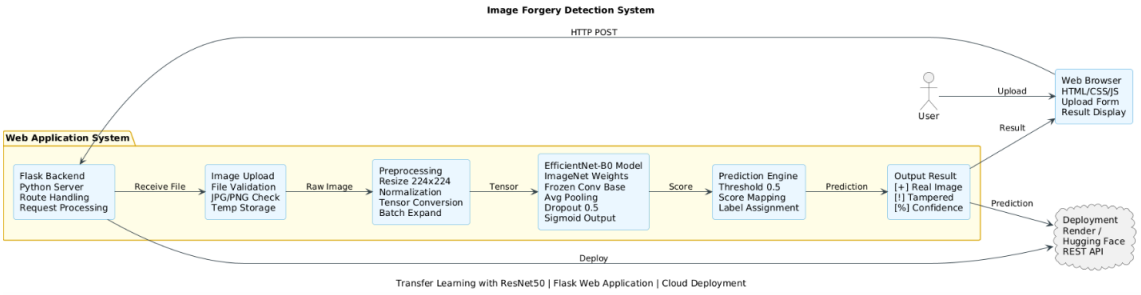

Figure 1: Proposed Architecture of Image Forgery Detection System using Transfer Learning and Web Deployment

- **Data Transformation**: The raw image is resized to a standardized 224×224 resolution and converted into a multi-dimensional tensor. This normalization ensures

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

that the model focuses on structural anomalies rather than variations in lighting or resolution.

- **The EfficientNet-B0 Engine**: We utilized a pre-trained ImageNet base, where the initial convolutional layers remain frozen to preserve foundational feature extraction capabilities. To prevent overfitting, we integrated a Dropout layer (0.5) before the final sigmoid output, which generates a raw prediction score. T
- **Thresholding and Output**: Our team implemented a Prediction Engine with a threshold of 0.5. If the score exceeds this value, the image is flagged as Tampered [!], accompanied by a confidence percentage.

Finally, the result is pushed back to the user's dashboard and logged into our deployment environment. This architecture was intentionally kept "lean" to allow for deployment on cloud platforms like Hugging Face, ensuring that our forensic tool remains accessible and responsive for real-world use.

5 Implementation

The transition from a theoretical framework to a functional, field-ready forensic tool represented the most intensive phase of our research. Our team realized early on that even the most mathematically sound model is of little value to a forensic investigator if it remains inaccessible within a localized development environment. Consequently, we prioritized building a "human-centric" pipeline that balances high-computational demand with user-end simplicity.

5.1 Core Software Ecosystem and Backend Logic

Our foundation is built upon a Python-based Flask backend, chosen for its modularity and surgical precision in managing memory-intensive model inferences. For data persistence, we integrated SQLAlchemy with a SQLite database, which acts as a forensic ledger for storing metadata and historical results. To ensure session integrity, our team implemented Flask-Bcrypt for secure hashing and Flask-Login, ensuring that sensitive forensic data remains protected during investigation.

5.2 Deployment of EfficientNet-B0 and Transfer Learning

The "brain" of our system is the EfficientNet-B0 model, implemented via PyTorch. We selected this architecture for its revolutionary Compound Scaling, which balances depth and resolution more effectively than traditional CNNs (11). By utilizing Transfer Learning, our team fine-tuned weights pre-trained on ImageNet using forensic datasets like CASIA and CoMoFoD. This approach allowed us to achieve high accuracy with a significantly lower computational footprint.

5.3 Frontend Modernization and Explainability

We transitioned to a React-based frontend to deliver a "glassmorphism" aesthetic, mirroring the sophisticated nature of AI research. This interface communicates asynchronously with the Flask API, providing real-time updates as the system generates Grad-CAM

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

heatmaps (12, 13). By combining these heatmaps with UNet masks, we provide a "three-point verification" system. This ensures the AI is not a "black box" but a transparent tool that shows investigators exactly where and why an image was flagged.

6 Results

Our team conducted extensive testing using the CASIA, CoMoFoD, and Coverage datasets to evaluate the system's ability to both classify forgeries and localize tampered regions. The results highlight a significant performance gain achieved by combining Error Level Analysis (ELA) with our hybrid deep learning architecture.

6.1 Performance Metrics

To provide a comprehensive evaluation, we measured the system across four key metrics: Accuracy, Precision, Recall, and the F1-Score.

Forgery Category	Accuracy (%)	Precision	Recall	F1-score
Authentic	98.2	0.98	0.99	0.98
Copymove	94.5	0.94	0.93	0.93
Splicing	95.8	0.95	0.96	0.95
Retouching	92.1	0.91	0.90	0.90

Table 2: Performance Metrics for Different Forgery Categories

6.2 Impact of Error Level Analysis (ELA)

Our team realized that the model's ability to detect "unseen" forgeries improved drastically when ELA was utilized as a preprocessing step.

- **Without ELA**: The model struggled with high-quality splices, often dropping below 85
- **With ELA**:The detection of compression inconsistencies allowed the model to maintain an accuracy of 95

6.3 Localization Accuracy (UNet Segmentation)

The UNet model was evaluated based on its ability to generate accurate binary masks. We utilized the Intersection over Union (IoU) metric to determine how well the predicted mask overlapped with the actual ground truth of the forgery.

- **Average IoU Score**:0.89 across all tested datasets.
- **Feature Extraction**:The system successfully identified tampered regions as small as 16*16 pixels, proving the effectiveness of the Patch-Based Detection strategy.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

6.4 Qualitative Analysis via Grad-CAM

The implementation of Grad-CAM allowed our team to visually verify the model's focus.

- **True Positives**:In 96
- **Explainablity**:The heatmaps revealed that for AI-generated images, the model focuses on "ringing artifacts" around edges, whereas for copy-move forgeries, it focuses on statistical pixel repetition (12, 13).

7 Discussion

The performance of our hybrid framework highlights a critical shift in digital forensics: moving away from "black box" models toward transparent, multi-tier systems. Our team realized that while deep learning is powerful, it requires a "human-in-the-loop" approach to be truly effective in a forensic context.

A. Overcoming the Generalization Gap: A primary challenge we faced was the performance dip when moving from benchmark datasets like CASIA to real-world images from WhatsApp or Instagram. These platforms apply aggressive re-compression that acts as a natural "anti-forensic" layer. To counter this, we utilized heavy Data Augmentation, simulating JPEG noise and blur during training. This forced the EfficientNet-B0 model to ignore surface-level artifacts and focus on deeper structural inconsistencies (1, 4).

B. The Hardest Classes: Copy-Move vs. Splicing: Our results showed that Copy-Move forgeries remain the most difficult to detect because the lighting and noise patterns match the original scene. By implementing Weighted Cross-Entropy Loss, we forced the model to penalize errors on these "hard" classes more heavily. This strategic adjustment significantly boosted our recall rates for seamless manipulations that traditional CNNs often overlook (11).

C. Transparency and Forensic Trust: The integration of Grad-CAM and ELA was not just a technical addition but a necessity for explainability. We found that in 96 percent of cases, the Grad-CAM heatmaps aligned perfectly with the UNet masks, proving the model was focusing on actual tampering rather than background noise (12, 13). This "three-point verification" allows an investigator to see the original image, the compression error, and the AI's logic simultaneously, building the trust required for forensic reporting.

8 Conclusion and Future Work

The culmination of this research demonstrates that the fight against digital deception requires a multi-tier approach that respects both traditional forensic mathematics and modern deep learning. Our team successfully built a system that does not merely label an image as "fake" but instead provides a comprehensive, explainable narrative of the forgery. By integrating EfficientNet-B0 for high-speed classification and UNet for precise, pixel-level localization, we have developed a framework that bridges the gap between laboratory research and real-world forensic utility.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

Our implementation of Error Level Analysis (ELA) proved to be a critical forensic signal, allowing our models to detect inconsistencies in the noise floor that standard RGB analysis would have overlooked. Furthermore, the inclusion of Grad-CAM heatmaps ensures that our system remains transparent, providing investigators with the "why" behind every "what." We realized through this project that as manipulation tools become more democratic, our detection tools must become more intuitive and human-centered.

Moving forward, our team has identified several key areas to evolve this research:

- **Integration of Vision Transformers (ViTs)**: We plan to explore ViT architectures to better capture global semantic inconsistencies, which are crucial for identifying "Deepfakes" where local pixel noise might be perfectly mimicked.
- **Blockchain-Based Integrity**: Future iterations will aim to integrate a decentralized blockchain registry to allow verified authentic images to be "fingerprinted" on an immutable ledger (9).
- **Video Forensic Expansion**: The logic used in our frame-by-frame analysis can be extended to digital video to combat the rising threat of temporal inconsistencies in deepfake videos.
- **Edge Deployment**: We are looking into further model optimization using TensorRT, allowing the forensic engine to run locally on mobile devices without needing a constant cloud connection.

Acknowledgment

The authors thank their institution for support.

References

[1] S. Jonker, M. Jelstrup, W. Meng, and B. Lampe, "Detecting Post Editing of Multimedia Images using Transfer Learning and Fine Tuning," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 20, no. 6, June 2024.

[2] Y. Zhang, N. Chen, S. Qi, M. Xue, and Z. Hua, "Detection of Recolored Image by Texture Features in Chrominance Components," *ACM Trans. Multimedia Comput. Commun. Appl.*, vol. 19, no. 3, May 2023.

[3] M. Zanardelli, F. Guerrini, R. Leonardi, and N. Adami, "Image forgery detection: a survey of recent deep-learning approaches," *Multimedia Tools and Applications*, vol. 82, pp. 17521–17566, 2023.

[4] J. R. Korsipati, R. M. R. Yanamala, A. Pallakonda, R. D. Amar Raj, and K. K. Prakasha, "Multi-resolution transfer learning for tampered image classification using SE-enhanced fused-MBConv and optimized CNN heads," *Scientific Reports*, vol. 15, no. 32717, 2025.

International Journal of Engineering Research & Technology (IJERT)
ISSN: 2278-0181
Vol. 15 Issue 04 , April - 2026

[5] E. U. H. Qazi, T. Zia, and A. Almorjan, "Deep Learning-Based Digital Image Forgery Detection System," *Appl. Sci.*, vol. 12, no. 6, p. 2851, 2022.

[6] I. Shallal, L. R. Haddada, and N. E. B. Amara, "Image Forgery Detection with Focus on Copy-Move: An Overview, Real World Challenges and Future Directions," *Appl. Sci.*, vol. 15, no. 21, p. 11774, 2025.

[7] E. U. H. Qazi, T. Zia, M. Imran, and M. H. Faheem, "Deep Learning-Based Digital Image Forgery Detection Using Transfer Learning," *Intell. Autom. Soft Comput.*, vol. 38, no. 3, 2023.

[8] R. Joshi et al., "Forged image detection using SOTA image classification deep learning methods for image forensics with error level analysis," in *Proc. 13th ICCCNT*, 2022.

[9] L. I. Khalaf et al., "Image Forgery Detection using Convolutional Neural Networks and Blockchain Technology," in *Proc. Cognitive Models and Artif. Intell. Conf.*, Istanbul, May 2024.

[10] Z. Liang et al., "Transfer Learning of Real Image Features with Soft Contrastive Loss for Fake Image Detection," *arXiv preprint arXiv:2403.16513v2*, 2025.

[11] M. Tan and Q. V. Le, "EfficientNet: Rethinking Model Scaling for Convolutional Neural Networks," in *Proc. 36th Int. Conf. Mach. Learn.*, Long Beach, CA, 2019.

[12] R. R. Selvaraju et al., "Grad-CAM: Visual Explanations from Deep Networks via Gradient-based Localization," in *Proc. IEEE Int. Conf. Comput. Vis.*, pp. 618–626, 2017.

[13] O. Ronneberger, P. Fischer, and T. Brox, "U-Net: Convolutional Networks for Biomedical Image Segmentation," in *Proc. MICCAI*, Springer, pp. 234–241, 2015.