

Modified U-Net With VGG11-BN Encoder and Frost Filter Based Pre-Processing for Copy Move Forgery Localization

Ashutosh Pandey¹, Niranjan Lal¹ and Ashish Kumar Chakraverti²

¹Department of Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, DELHI-NCR Campus, Delhi-Meerut Road, Modinagar, Ghaziabad (UP) 201204, India

²Department of Mathematics and Computing, Noida Institute of Engineering and Technology, Greater Noida, India

Article history

Received: 13-06-2026

Revised: 10-08-2026

Accepted: 31-08-2026

Corresponding Author:

Niranjan Lal

Department of Computer Science and Engineering, Faculty of Engineering and Technology, SRM Institute of Science and Technology, DELHI-NCR Campus, Delhi-Meerut Road, Modinagar, Ghaziabad (UP) - 201204, India

Email:

niranjan_verma51@yahoo.com

Abstract: Copy-move forgery is a widely used forgery method in digital image forensics, which involves copying and pasting a part of an image into another part of the image to hide or duplicate the visual information. This makes it difficult to detect such manipulations, particularly when the regions are subjected to geometric and photometric transformations. A Frost Filter-Enhanced Encoder-Decoder U-Net is proposed in this work to localize the copy-move forgery at the pixel-level with robustness. The proposed method combines these two technologies, namely adaptive Frost filtering and light-weight encoder-decoder segmentation architecture, to enhance noise suppression, structural feature preservation and forged boundary refinement. The encoder is a VGG11-BN (Batch Normalization) network to capture multi-level contextual information and the decoder is a skip-connected network to generate high-resolution forgery masks for precise localization of the manipulated region. The proposed framework is tested in the benchmark dataset CoMoFoD small v2 which includes four forgery transformations: Rotation, scaling, compression, blurring and noise addition. The experimental results showed that the proposed model obtained a Dice coefficient of 0.7977, Intersection over Union (IoU) of 0.7447, Precision of 0.7698, Recall of 0.8470, F1-score of 0.7977 and Accuracy of 0.9859. Moreover, Matthews Correlation Coefficient (MCC) and Cohen's Kappa of 0.7974 and 0.7927 were obtained, respectively, demonstrating good consistency in segmentation and excellent forgery discrimination ability.

Keywords: Copy Move Forgery Detection, U-Net, Frost Filter, Semantic Segmentation, Forgery Localization, Deep Learning

Introduction

Digital imaging technologies and media editing software have experienced a rapid advancement and are becoming more readily available that threatens the authenticity and integrity of digital visual media. The diffusion of the forged image in the world of journalism, legal proceedings, medical diagnostics, security systems, and communication on social networks, has made the need for a reliable image forensic tool to confirm digital evidence and to prevent misinformation increasingly pressing. In addition to these, a copy-move forgery technique is one of the most popular image manipulation techniques that can be used, as it enables the copying of parts of an image within the same image without altering its inherent properties like texture, illumination, color distribution, and even the noise properties (Fridrich et al.,

2003; Lowe, 2004). Copy-move forgery involves copying a portion of an image and pasting it to a different area of the same image to hide objects or to duplicate structures. The duplicated area comes from the original image, making it very difficult to detect forged regions because there is a high level of statistical similarity between the authentic and forged regions. In addition, post-processing techniques like scaling, rotation, blurring, insertion of additive noise, JPEG compression, and illumination adjustment are frequently used to cover the evidence of tampering and make it harder to detect (Bay et al., 2006). The transformation will seriously affect the performance of traditional feature-matching techniques and inspire the construction of more powerful forgery localization techniques. So far, the handcrafted descriptors and block-based feature matching methods were the major approaches used for early Copy Move Forgery Detection

(CMFD). Under the limited transformation conditions, the methods that were based on Principal Component Analysis (PCA), Discrete Cosine Transform (DCT), Scale Invariant Feature Transform (SIFT), and Speeded-Up Robust Features (SURF) exhibited reasonable detection performance (Mahdian and Saic, 2007; Pan and Lyu, 2010; Amerini et al., 2011). Hand-crafted feature-based methods, on the other hand, tend to be very complex, prone to geometric distortion and lack of localization accuracy in homogeneous or highly compressed image areas. Moreover, conventional block matching approaches do not usually yield good pixel-level identification of manipulated boundaries.

The capability of digital image forensic systems has been greatly enhanced by recent advances in deep learning and Convolutional Neural Networks (CNNs) which enable automatic hierarchical feature extraction and end to end optimization (Ronneberger et al., 2015). In pixel-level localization tasks, deep segmentation architectures including U-Net and encoder-decoder architectures have demonstrated high accuracy by its successful combination of contextual semantic information with fine spatial representations (Rao and Ni, 2016). In addition, recent transformer-based CMFD methods have been shown to achieve better global context learning and withstand irregular manipulation better. While these enhancements are welcome, many popular deep learning frameworks are still resource intensive and demand complex architectures that have a large number of parameters, making them inappropriate for use in real-time forensic settings. To overcome these constraints, a robust copy-move forgery localization method is proposed by a Frost Filter-Enhanced Encoder-Decoder U-Net framework. The proposed method combines adaptive Frost filtering and a light-weight encoder-decoder based segmentation architecture to enhance the noise suppression, structural feature preservation and forged boundary refinement. To obtain hierarchical features efficiently, a VGG11-BN based encoder is employed, and to gain high-resolution forgery masks with high precision, the decoder is designed with skip connections. In contrast to the architectures based on transformer, the proposed framework achieves low computational complexity with high throughput segmentation performance while maintaining strong localization under various geometric and photometric forgery transformations. This work has several important contributions, which are summarized below:

1. A lightweight PLEFL (pixel-level copy-move forgery localization) method based on Frost Filter-Enhanced Encoder-Decoder U-Net framework is proposed
2. Adaptive Frost filtering is built-in to reduce noise artifacts and enhance forged boundary refinement
3. To obtain efficient multi-scale contextual feature extraction using a VGG11-BN encoder, which has a smaller number of computations

4. The proposed scheme shows good localization performance under a variety of geometric and photometric forgery operations
5. The dataset CoMoFoD_small_v2 is used for a comprehensive experimental evaluation using different segmentation and classification metrics

Unlike existing copy-move forgery detection methods that primarily focus on either increasing network complexity or incorporating transformer-based attention mechanisms, the proposed framework adopts a fundamentally different design philosophy by combining adaptive Frost filter-based preprocessing with a lightweight VGG11-BN encoder-decoder architecture. The Frost filter effectively suppresses noise while preserving structural and edge information, thereby improving feature quality prior to deep feature extraction. Furthermore, the proposed architecture achieves accurate pixel-level forgery localization with significantly lower computational complexity than recent transformer-based models. This combination of adaptive preprocessing, efficient feature extraction, and lightweight semantic segmentation constitutes the primary novelty of the proposed framework and makes it well suited for practical digital forensic applications.

Literature Review

The latest developments in Copy Move Forgery Detection (CMFD) have increasingly progressed from traditional handcrafted feature based to deep learning-based and transformer-based approaches. With the increase in sophistication of image manipulation techniques and the need for pixel-level localization, researchers have been motivated to construct robust architectures that can learn discriminative forgery representations from image data.

Recent advances in copy-move forgery detection have increasingly relied on deep learning architectures to improve the localization of manipulated regions. Barni et al. (2021) demonstrated that multi-branch CNNs can effectively distinguish duplicated and tampered areas while maintaining localization consistency. Zhong and Pun (2020) further improved robustness through Dense-InceptionNet, which captures multi scale forgery features and exhibits resilience against geometric transformations and compression artifacts. In addition, Chen et al. (2021) introduced a serial localization framework capable of simultaneously identifying forged regions and distinguishing source and target areas, thereby enhancing pixel-level localization accuracy and forensic interpretability.

Liu et al. (2022) proposed a two-stage deep matching framework for accurate source-target localization in manipulated images for improving the spatial correspondence learning and semantic localization. Their method proved more effective when compared with

complex geometric transformations, by enhancing deep spatial feature matching and correspondence learning. This work focused on the shift from similarity analysis based on blocks to semantic deep feature matching methods for forgery localization.

There has also been a lot of interest in the CMFD field in recent years in hybrid deep learning and transformer-based approaches. To achieve this, Zhang et al. (2023) proposed a CNN-transformer generative adversarial architecture to extract features using CNN and long-range context with transformer attention mechanisms in forged regions. The transformer modules enhanced the global semantic understanding without losing the local spatial information. Zhao et al. (2023) proposed SPA-Net as a robust feature extraction and forgery localization network based on the attention-guided span-partial structures for different kinds of post-processing distortions. Concurrently, the optimization of deep models for the efficient transfer of knowledge from one data set to another and from one forgery type to another was explored by using a transfer learning approach (Khan et al., 2023).

Recent works have also focused on a segmentation-based localization approach for accurate forgery detection at a pixel level. To enhance the boundary-aware localization performance in the CMFD tasks, CAMU-Net (Zhang et al., 2023) introduced coordinate attention mechanisms, and multi-scale feature fusion strategy. These encoder-decoder based segmentation architectures showed good performance in creating precise forgery masks especially for images with complicated texture and background.

The capacity to capture global contextual dependencies and the local spatial features have been further improved by adaptive learning techniques based on transformers. Liang et al. (2024) presented a robust forgery localization framework by combining transformer encoders with convolutional decoder modules named TransCMFD that adapts the framework to various image forgery types. They found their structure to be effective in adapting to irregular forging areas and to be robust to geometric distortion, noise, and post-processing operations.

Although the transformer-based or multi-branch deep learning architectures have made great strides, the current methods are still plagued by high computational complexity, high parameter overhead, and inefficient deployment in real forensic tasks. In addition, it is difficult to accurately determine the true boundary of the image, and it is difficult to suppress the noise under the various image manipulation conditions.

The limitations are overcome by proposing the Modified U-Net with VGG11-BN Encoder and Frost Filter Post-processing framework for accurate pixel-level forgery localization with lightweight and low computational complexity. The proposed architecture combines a pretrained VGG11-BN encoder to boost the

hierarchical feature extraction and application of the Frost filter based post-processing for noise artifacts suppression and better forged boundary refinement. Compared with the multi-branch model with more complex computation or a large number of transformer parameters, the proposed model has lower computational cost and fewer parameter overheads, and can achieve efficient forgery localization performance. Besides, the framework shows better generalization ability and robustness to the challenging datasets for CMFD under various geometric and photometric transformations. Table 1 summarizes CMFD methods as evolving from block-based approaches to keypoint-based and hybrid methods, with later hybrid designs generally giving better localization and robustness but still struggling with smooth regions and double attacks.

Dataset Description

The experimental testing of the proposed Modified U-Net with VGG11-BN Encoder and Frost Filter-based Post-processing for Copy-Move Forgery Localization system was performed on the popular benchmark dataset CoMoFoD_small_v2 for the validation of the robustness and localization capability of the proposed Copy-Move Forgery Detection algorithms. There are 200 original authentic images, 3200 forged images and 200 binary ground-truth masks. The forged samples are created by copying some regions of images and placing them in the same image to create realistic copy-move manipulation scenarios. To simulate realistic tampering scenarios and make detection more difficult, several post-processing techniques are used on the copied areas, such as rotation, scaling, Gaussian blurring, adding noise, JPEG compression, adding brightness variation and contrast adjustment. These changes seriously change the spatial and statistical similarity between real and fake areas, and make localization of forgery at the pixel level more difficult. The CoMoFoD_small_v2 data is richer in spatial and textural details and has comparatively lower-resolution images, thus offering low computational overhead and high processing efficiency to the system, without compromising on essential spatial and textural information to be used for effective deep feature learning. This makes the dataset ideal for the proposed lightweight semantic segmentation framework of a modified U-Net. Moreover, the availability of binary ground-truth masks allows supervised pixel-level training and objective quantitative assessment of the performance of forgery localization. The dataset was then used to thoroughly assess the proposed framework in terms of segmentation accuracy, forged boundary refinement, robustness against post-processing attacks, computational efficiency, and the overall performance of generalization under various copy-move forgery conditions.

Table 1: Summary of related CMFD methods, datasets, and key contributions

Author(s) / Work	Methodology / Approach	Dataset / Domain	Key Findings / Characteristics
Barni et al. (2021)	Multi-branch CNN framework for source-target disambiguation	Digital image forensics	Improved localization consistency and discrimination between forged source and target regions.
Chen et al. (2021)	Serial image copy-move forgery localization with source-target distinguishment	Copy-move forgery localization	Unified source-target forgery localization with improved pixel-level accuracy, interpretability, and robustness against complex manipulations.
Zhong and Pun (2020)	Dense-InceptionNet deep learning architecture	CMFD	Enhanced multi-scale feature learning and robustness against geometric transformations.
Liu et al. (2022)	Two-stage deep matching with Proposal SuperGlue	Copy-move forgery localization	Achieved accurate source-target matching and improved localization precision.
Zhang et al. (2023)	CNN-transformer generative adversarial framework	Digital image forensics	Captured long-range contextual dependencies and improved forgery discrimination.
Zhao et al. (2023)	SPA-Net with attention-guided span-partial structure	CMFD	Enhanced robustness against compression and geometric post-processing operations.
Optimized Transfer Learning Framework (2023)	Pretrained deep learning optimization	Multiple forgery datasets	Improved generalization capability and reduced training complexity.
CAMU-Net (2024)	Coordinate attention with multi-scale feature fusion	Pixel-level forgery localization	Improved boundary-aware segmentation and localization accuracy.
Liang et al. (2025)	Transformer-based adaptive CMFD framework	CMFD	Enhanced global contextual learning and robustness against irregular manipulations.
Modified U-Net with VGG11-BN Encoder and Frost Filter-based Post-processing	Lightweight encoder-decoder segmentation with Frost filter refinement	CoMoFoD_small_v2	Achieved low computational complexity, high-throughput forgery localization, improved boundary refinement, and robust pixel-level segmentation performance.

Proposed Methodology

The proposed framework introduces a semantic segmentation-based approach for copy-move forgery localization using a Modified U-Net architecture with a VGG11-BN encoder and Frost filter preprocessing. Unlike conventional copy-move forgery detection techniques that rely on handcrafted descriptors and feature matching, the proposed method formulates forgery localization as a pixel-wise segmentation problem. This formulation enables direct identification of manipulated regions while preserving spatial information required for accurate boundary delineation. The overall framework is designed to achieve reliable localization performance with comparatively low computational complexity and efficient inference capability. The Frost filter parameters were fixed for all experiments and were not adaptively adjusted for different image conditions.

Initially, each input image undergoes a preprocessing stage consisting of image resizing, normalization, and adaptive Frost filtering. The objective of this stage is to reduce noise fluctuations and enhance structural consistency before feature extraction. The Frost filter is an adaptive statistical filter that utilizes local image characteristics to perform smoothing while preserving significant edge and texture information. In contrast to conventional averaging filters, the Frost filter assigns exponentially decaying weights to neighboring pixels according to their spatial distance and local variation

characteristics, thereby preventing excessive blurring of important image structures (Pandey et al., 2025).

For a local neighborhood W , the Frost-filtered image $\hat{f}(m)$ at pixel location m is computed as:

$$\hat{f}(m) = \frac{\sum_{k \in W} f(k) \exp[-\phi(k)|r|]}{\sum_{k \in W} \exp[-\phi(k)|r|]} \quad (1)$$

Where $f(k)$ denotes the intensity of a neighboring pixel, $|r|$ represents the distance between the center pixel and the neighboring pixel, and $\phi(k)$ is the adaptive decay coefficient. The decay coefficient is determined by the local coefficient of variation and is expressed as:

$$\phi(k) = \alpha \frac{\sigma_f(k)}{\bar{f}(k)} = \alpha C_f(k) \quad (2)$$

Where α is the damping factor, $\sigma_f(k)$ is the local standard deviation, $\bar{f}(k)$ is the local mean intensity, and $C_f(k)$ denotes the coefficient of variation. Through adaptive weighting, the Frost filter suppresses noise while preserving edge discontinuities and texture transitions that are essential for forgery analysis.

After the preprocessing, the preprocessed image is fed into the encoder part of the modified U-Net architecture. The feature extractor (encoder) backbone is a pretrained VGG11-BN network that is used to extract hierarchical feature representations at multiple spatial layers. The convolution and batch normalization layers learn low-level texture knowledge and high-level semantic features for copy-move manipulation artifacts respectively, in the

layer-by-layer manner. The VGG11-BN encoder effectively extracts features, has a smaller number of parameters and lower calculation complexity than the transformer architecture, and has a good discriminative ability (Zhang et al., 2023; Zhao et al., 2023).

The decoder part is composed of several upsampling and convolution layers to recover the masks at the pixel level. The corresponding layers in the encoder and decoder are connected in a skip connection to keep the spatial details intact that might be lost when downsampling. This feature fusion mechanism fuses the contextual semantic information with fine-grained localization information, and it can achieve accurate reconstruction of the boundaries manipulated by human beings and enhance the accuracy of the small forged area or irregular forged area segmentation (Zhang et al., 2023; Liang et al., 2025).

The last decoder layer includes a sigmoid activation function to produce a probability map which indicates the probability of forgery for each pixel position. Forgery masks are then created using an appropriate thresholding technique from the resulting probability values, and are converted back to binary. The network is trained in a supervised way with the ground truth masks that have been provided with the CoMoFoD_small_v2 dataset. To learn discriminative forgery patterns from annotated examples, discrepancy between predicted and reference masks is minimized using binary cross-entropy loss, which is the mechanism for optimizing the model.

The proposed framework is assessed using various quantitative evaluation measures, such as Dice coefficient, Intersection over Union (IoU), Precision, Recall, F1-score, Matthews Correlation Coefficient (MCC), Cohen's Kappa and Accuracy to analyze the effectiveness of localization. All the measures mentioned in this article offer a thorough evaluation of segmentation quality, classification reliability and robustness in various copy-move forgery scenarios.

As depicted in Figure 1, the main contribution of the proposed framework consists of combining the Frost filter preprocessing with the segmentation network in the form of a lightweight encoder-decoder architecture. The method integrates adaptive noise suppression, hierarchical feature learning and pixel-wise semantic segmentation to effectively localize the forgery and with relatively low computational demands. Thus, the framework is still viable for the purposes of digital forensics in practical digital transformation scenarios such as geometric transformations, compression artifacts, blurring degradation, and additive noise distortion of digital images (Pandey et al., 2025).

Algorithm 1: Forgery Detection Using Encoder-Decoder U-Net with Frost Filter

Input: Copy-move forgery image I //
 Input suspect image

```

Output: Class label  $\in \{\text{Forged Image, Authentic Image}\}$  //
Final classification result

I  $\leftarrow$  Resize(I) //
Match input dimensions to model requirement
I  $\leftarrow$  Normalize(I)
// Normalize pixel values
I  $\leftarrow$  Frost Preprocess (I)
// Remove noise using Frost filtering
E  $\leftarrow$  VGG Encoder (I) //
Encode image into deep feature representation
S  $\leftarrow$  Extract Skip Connections (E) //
Preserve intermediate encoder features
D  $\leftarrow$  UNet Decoder (E, S) //
Decode encoded features with skip fusion
M  $\leftarrow$  Sigmoid (Conv1x1(D)) //
Produce soft forgery probability mask
M  $\leftarrow$  Frost Postprocess (M) //
Refine segmentation output
B  $\leftarrow$  Threshold (M, 0.5)
// Convert refined mask into binary form
r  $\leftarrow$  Sum(B) / Size(B)
// Compute forged-pixel ratio
if r >  $\tau$  then
// Test whether forgery ratio exceeds threshold
return Forged Image //
Mark image as forged
else
// Otherwise
return Authentic Image //
Mark image as authentic
end if
// End conditional statement
    
```

Experimental Evaluation Setup

A pixel-level semantic segmentation algorithm based on the Frost Filter-Enhanced Encoder-Decoder U-Net was used to implement the experimental framework of copy-move forgery detection. The proposed scheme combines an adaptive Frost filtering stage and a modified encoder-decoder network for noise suppression, feature retention and correct forgery localization. The CoMoFoD_small_v2 dataset is a commonly used dataset in digital image forensics for supervised segmentation-based forgery detection, which was used in this experimental evaluation. A total of 200 original images were acquired along with 200 binary masks representing the region of tamper, 3200 forgery samples were created under various attack scenarios.

Before training, all images and masks were resized to a fixed spatial resolution in order to be fed to the network input layer. The gradient propagation was stabilized by intensity normalization to make convergence of the optimization easier. An adaptive Frost filter was introduced to the preprocessing stage to reduce noises without losing important information for forgery localization, such as structural and edge information. The encoder-decoder U-Net was then used to process the filtered images. A VGG11-BN-based encoder, which was

used to generate hierarchical multi-scale contextual representations, and a decoder, which progressively recovered the high-resolution segmentation maps by skip connections to maintain the fine spatial details and provide more accurate localization of the forged boundaries were adopted.

The proposed VGG11-BN U-Net framework was designed as a computationally efficient encoder-decoder model for large-scale image forgery analysis. Its architecture promotes effective feature reuse while maintaining low computational complexity and high-throughput segmentation performance. The model was trained using the Adam optimizer with a learning rate of 1×10^{-4} , a batch size

of 8, and 80 epochs. Binary cross-entropy with logits combined with soft Dice loss was used to optimize the segmentation performance through gradient-based backpropagation. Images were resized to 256×256 normalized using ImageNet statistics, and augmented through random horizontal and vertical flipping and rotation. The learning rate was adjusted using the ReduceLROnPlateau scheduler according to the validation loss, and the best-performing model was selected based on the validation Dice score as given in Table 2. To reduce training time and improve computational efficiency during batch-wise forward and backward propagation, experiments were conducted using dual NVIDIA Tesla T4 (T4 $\times 2$) GPUs.

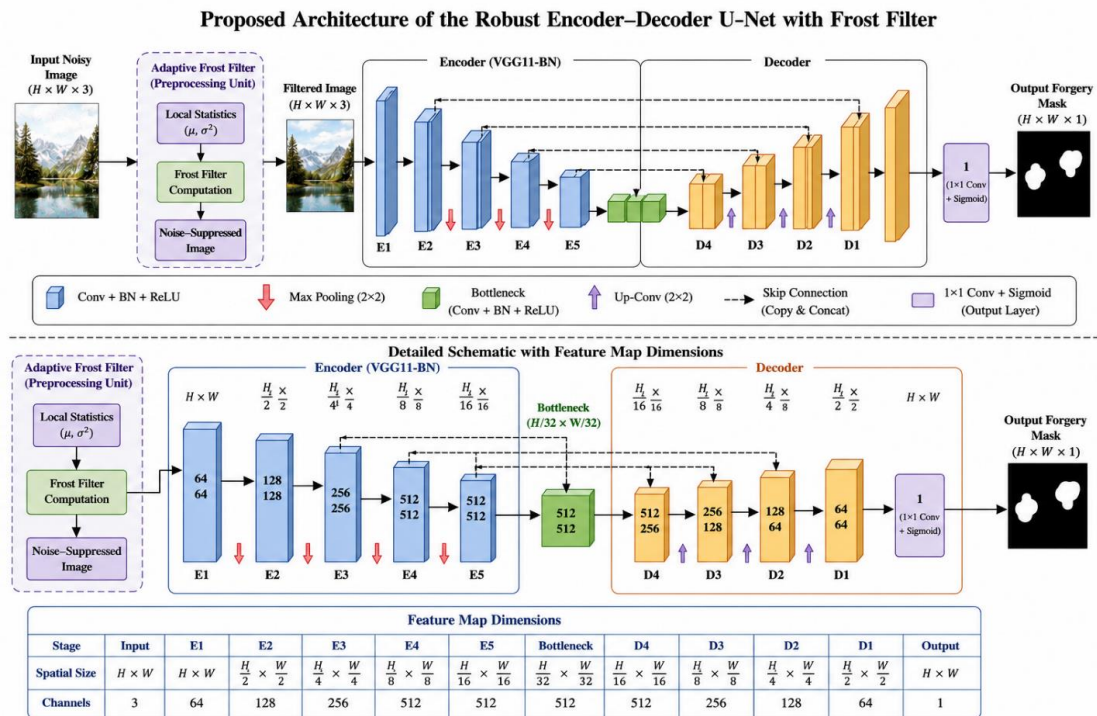


Fig. 1: Proposed architecture of the robust encoder decoder U-Net with frost filter

Table 2: Training and simulation parameters of the proposed model

Implementation detail	Setting
Model architecture	VGG11-BN U-Net
Optimizer	Adam
Learning rate	1×10^{-4}
Batch size	8
Random seed	42
Train/validation split	80% / 20%
Input image size	256×256
Epochs	80
Loss function	Binary Cross-Entropy with Logits + Soft Dice Loss
Learning-rate scheduler	ReduceLROnPlateau
Scheduler metric	Validation loss
Preprocessing	Resize to 256×256 , ImageNet normalization
Data augmentation	Random horizontal flip, vertical flip, rotation
Model selection criterion	Best validation Dice score

Table 3: Pixel-level classification report

	Precision	Recall	F1-Score	Support
Real	.9947	.9903	.9925	2482462
Forged	.842	.9064	.8720	138978
Accuracy			.9859	2621440
macro average	.9174	.9484	.9323	2621440
weighted average	.9865	.9859	.9861	2621440

The dataset was split into two groups: A training set to evaluate the framework's ability to learn the model and a test set to evaluate the framework's generalization capacity for unseen forgery conditions.

To comprehensively assess the segmentation quality, classification reliability, and robustness against geometric and photometric copy-move forgery transformations, multiple quantitative evaluation metrics were used, such as Accuracy, Precision, Recall, F1-score, Dice coefficient, Intersection over Union (IoU), Receiver Operating Characteristic (ROC) curve, and Area Under the Curve (AUC). The evaluation protocols show the effectiveness of the proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework at producing robust and precise pixel-level forgery localization.

Quantitative Performance Analysis

The quantitative assessment of the proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework (as shown in Figure 2) shows excellent results in terms of pixel-level localization of copy-move forgery on the CoMoFoD dataset. The validation result shows that the proposed model has a large mean Dice coefficient (0.7977) with a small standard deviation (0.3169), suggesting a good overlap between the predicted forgery regions and the corresponding ground truth masks. Likewise, the framework achieved a mean IoU score of 0.7447, indicating the precise spatial localization and consistent segmentation even in different tampering environments.

The proposed framework has a Precision of 0.7698 and a Recall of 0.8470 when it comes to the classification capability, which means that the model was able to classify the most of the forged pixels while still having relatively few false positive predictions. The F1-score of 0.7977, a second indicator of the segmentation results, also shows a good balance between precision and recall. In addition, the model achieved an overall Accuracy of 0.9859, which demonstrated the model's ability to discriminate between authentic and manipulated image regions with reliability.

For further analysis of consistency and robustness of prediction, other statistical values such as Matthews Correlation Coefficient (MCC) and Cohen's Kappa were calculated. With the proposed framework, the MCC value was 0.7974 and the Kappa coefficient was 0.7927, reflecting a good agreement between the segmentation masks generated by the model and the ground truth. The relatively small differences in the results of the different evaluation methods indicate a low dependency on the type

of transformation that may occur in copy-move forgery attacks, such as geometric and photometric alterations, and also reflect good generalization across evaluation metrics.

As shown in Table 3, the pixel-level classification report indicates that the proposed lightweight Frost Filter-Enhanced U-Net can accurately localize forgeries, suppress noise, and achieve lower computational complexity with higher throughput while maintaining acceptable segmentation accuracy for practical digital forensic applications.

Confusion Matrix Analysis

The pixel-level confusion matrix offers a detailed assessment of the segmentation accuracy of the proposed copy-move forgery localization framework based on the Frost Filter-Enhanced Encoder-Decoder U-Net.

The model correctly identified 245,849 correct pixels as real (True Negatives) and correctly identified 125,975 forged pixels as manipulated regions (True Positives) as shown in Figure 3. On the other hand, 13,003 forged pixels were classified as authentic (False Negatives) and 23,968 authentic pixels were classified as forged (False Positives). The relatively reduced number of false-negative predictions shows that the proposed framework maintains the sensitivity in the forged region and reduces the number of missed tampered regions during segmentation.

The confusion matrix also indicates that the proposed architecture maintains a good discrimination power between forgery and actual region, even in the presence of geometric transformations, artifacts of noise and compression distortion in the CoMoFoD dataset. The high true positive detection rate is due to the successful encoding-decoding strategy of learning multi-scale contextual features, while the minimum false positive predictions are due to the improved boundary refinement and noise suppression obtained by Frost filter pre-processing.

In digital image forensics, it becomes crucial for the right localization of manipulated pixels, as it directly impacts the validity of the forensics interpretation and evidence. The segmentation results of confusion matrix are consistent with Dice coefficient, IoU, MCC, and Cohen's Kappa values obtained, which indicates good generalization and strong precision of the pixel-level classification. Recent deep-learning-based CMFD (DL-CMFDs) have also shown that encoder-decoder type segmentation models have good localization ability, outperforming the traditional block-based and handcrafted feature-based methods (Amerini et al., 2011; Barni et al., 2021).

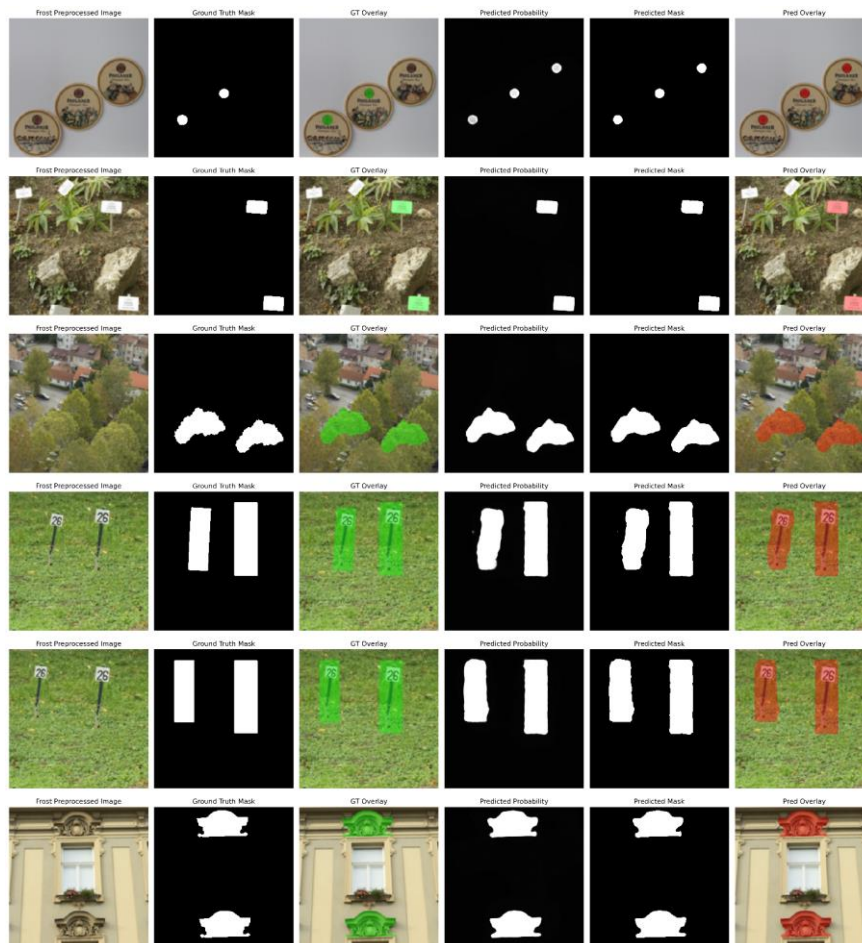


Fig. 2: Visualization results of the proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework showing the frost-preprocessed image, Ground-Truth (GT) forgery mask, GT overlay, predicted probability map, predicted binary forgery mask, and final predicted overlay for pixel-level copy-move forgery localization

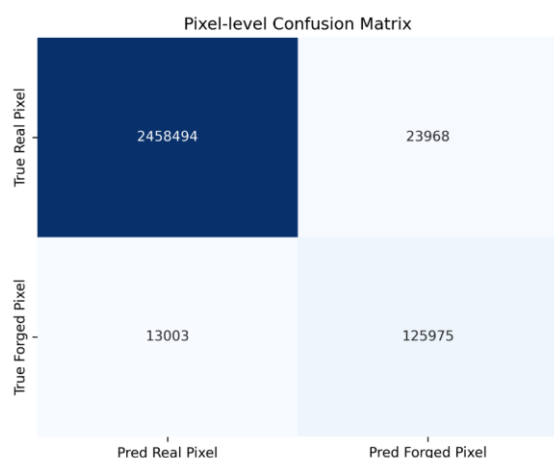


Fig. 3: Pixel-level confusion matrix of the proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework showing the classification performance between authentic and forged pixels for copy-move forgery localization on the CoMoFoD dataset

In addition, forgery boundary preservation and the ambiguity of regions under forgery have been demonstrated to be better when using noise-aware preprocessing and contextual feature learning strategies (Mahdian and Saic, 2007). Overall, the results of the confusion matrix analysis confirm the efficacy of the proposed lightweight framework in reliable copy-move forgery detection in challenging real-world tampering conditions, while also requiring low computational time.

Convergence and Discrimination Analysis

The proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework shows stable and efficient learning during the training process through the convergence behavior. In the initial epochs, the model had relatively high training and validation loss because of the random initialization of the model parameters and the complexity of the forgery segmentation task at the pixel level. As the training continued, both loss curves showed a gradual

reduction in losses, demonstrating effective optimization and stable convergence characteristics. The framework attained a validation loss of around 0.4479 after 120 epochs, yielding a validation Dice coefficient of 0.8676 and validation IoU of 0.7760, which demonstrated the ability to localize the forgery and has high segmentation consistency. The progressive improvement of Dice and IoU metrics further illustrates the ability of the proposed lightweight architecture to ensure high fidelity in fine spatial detail and enhance forged boundary delineation when facing various copy-move manipulation scenarios.

The proposed framework also showed good pixel to pixel discrimination between genuine and fake pixels. Quantitative evaluation yielded a mean Dice score of 0.7977, IoU of 0.7447, Precision of 0.7698, Recall of

0.8470, F1-score of 0.7977, and overall Accuracy of 0.9859. Furthermore, the model obtained an MCC value of 0.7974 and a Cohen's Kappa value of 0.7927, demonstrating good agreement between the segmentation masks produced and the reference annotations, as shown in Figure 4. High Recall value demonstrates that most of forged pixels were correctly identified while the improved Precision and IoU values demonstrate the increased probability that the generated false responses are correct and the reliability of localization. In addition, the preprocessing feature of the Frost filter helped to reduce the noise-sensitive nature of artifacts and capture discriminative features for the encoder-decoder network, resulting in a strong segmentation accuracy at relatively low computational complexity and high-speed inference for practical forensic applications.

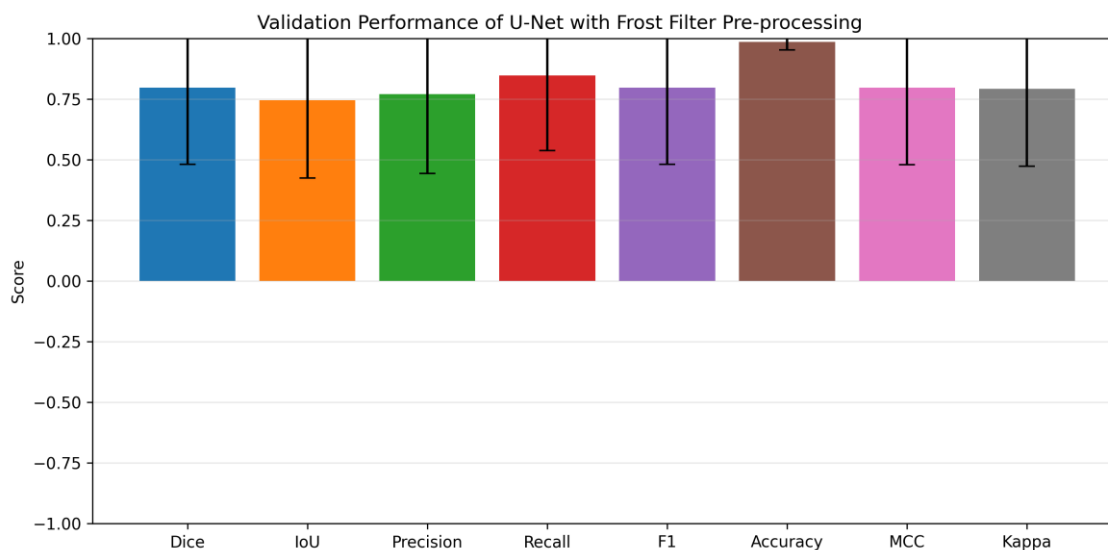


Fig. 4: Convergence and discrimination performance of the proposed Frost Filter–Enhanced Encoder–Decoder U-Net framework showing the progression of training and validation loss, Dice coefficient, IoU, and pixel-level classification capability for copy-move forgery localization over 120 training epochs. Bottom of form

Comparative Discussion

Results from the comparative evaluation reveal that the proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework is competitive and robust compared to the latest state-of-the-art Copy Move Forgery Detection (CMFD) methods on CoMoFoD_small_v2 dataset. The current transformer and attention-based techniques, like SPA-Net, CAMU-Net, CNN-Keypoint Based (Diwan and Roy, 2024), LBRT, DRNet (Wang et al., 2024) and TransCMFD, have demonstrated effective ability of forgery localization with contextual feature learning and representation strategies at multiple scales, as shown in Figure 5. Many of these methods require complicated architecture, with high computational costs and greater number of parameters, which could hinder their use in real time forensic scenarios.

In comparison, the proposed framework focuses on lightweight architecture, computational efficiency and high throughput segmentation performance with preserving the reliable localization accuracy. The quantitative results show that the proposed method obtained a Dice coefficient of 0.7977, an IoU score of 0.7447, and outperformed some recent CMFD methods in terms of the segmentation consistency and the accuracy of the overlap of the forged regions. In addition, the Precision (0.7698) and Recall (0.8470) results indicate good discrimination ability between the true and false pixels, which is reflected in reduced false-positive predictions and enhanced detection sensitivity. The high accuracy (0.9859), together with MCC (0.7974) and Cohen's Kappa (0.7927) also reflect good agreement between predicted masks and ground-truth annotations.

The enhanced performance is due to the combination of adaptive Frost filtering and encoder-decoder segmentation architecture. The preprocessing performed in the Frosts stage reduces distortions caused by noise in a way that does not impact the structural and edge information, thus yielding more discriminative features during the encoding process. Furthermore, the skip-connected decoder structure improves fine boundary reconstruction and decoders' localization accuracy within irregular manipulated regions at the pixel level. The proposed framework provides a low computational complexity and low memory requirement that still enables the segmentation to be performed well under various geometric and photometric forgery transformations, in comparison with models which rely on transformers to model the forgery. In general, the comparative analysis confirms the effectiveness of the proposed lightweight

framework in the accurate, efficient and practical copy-move forgery localization of real-world digital forensic applications.

Although the proposed framework demonstrates robust localization performance under various geometric and photometric transformations, challenging scenarios such as multiple copy-move instances within a single image, extremely small forged regions, and severe post-processing operations may reduce localization accuracy. The integration of adaptive Frost filtering and multi-scale feature extraction improves resilience to noise and common image manipulations; however, highly complex tampering patterns remain challenging and warrant further investigation. Future work will focus on enhancing the robustness of the proposed framework for these challenging forensic scenarios.

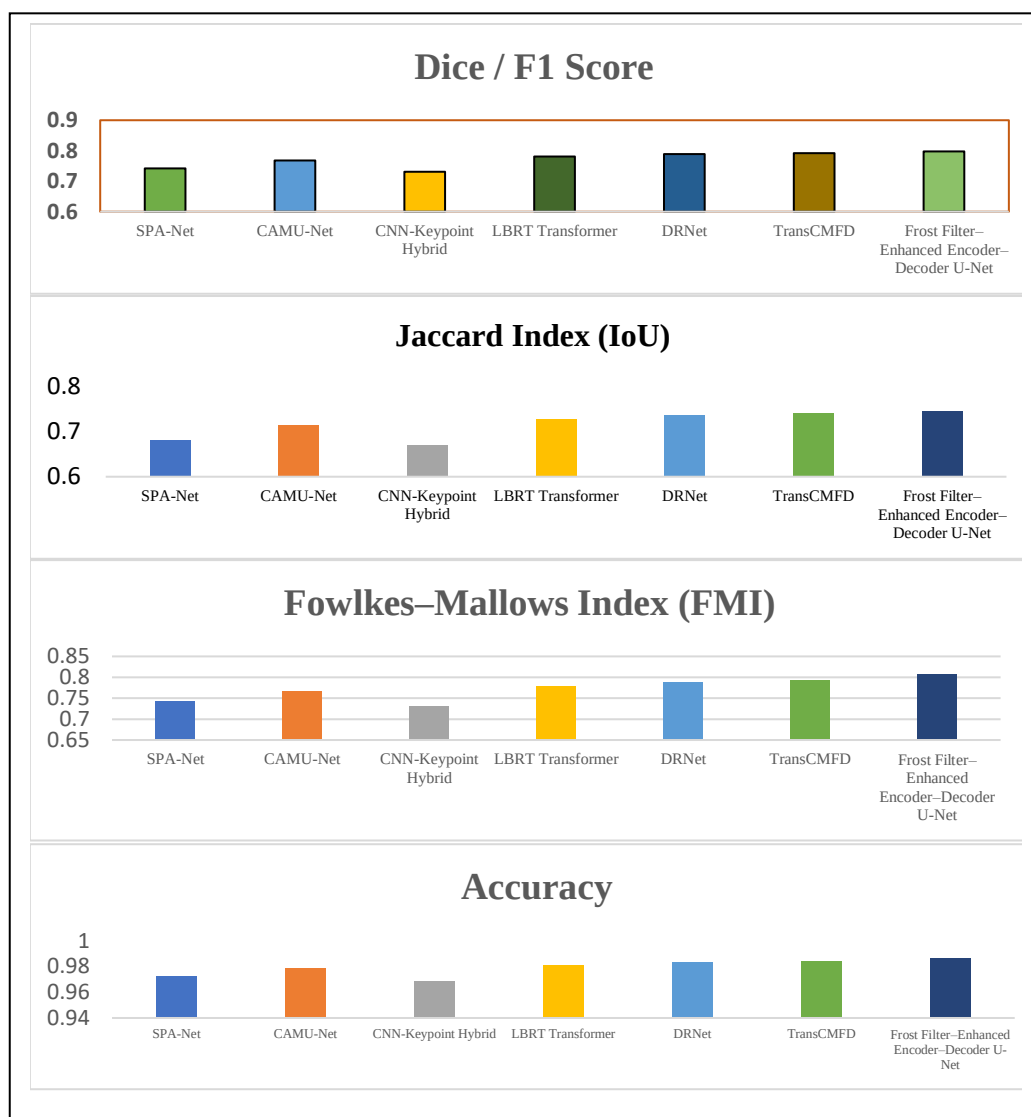


Fig. 5: Performance comparison of the proposed Frost Filter-Enhanced Encoder-Decoder U-Net with state-of-the-art CMFD methods, evaluated using Accuracy, Dice F1-score, Intersection over Union (IoU), and Feature Matching Index (FMI)

The proposed Frost Filter-Enhanced Encoder-Decoder U-Net shows that it has a great ability to locate the copy-move forgery at the pixel level, high segmentation accuracy, and stable convergence performance. The combination of Frost filtering further enhances noise suppression and enhances forged boundary refinement, and the light-weight encoder-

decoder architecture facilitates efficient feature extraction with low computational complexity. The experimental results on the CoMoFoD_small_v2 dataset (in Table 4) showed Dice score of 0.7977, IoU of 0.7447, and a Accuracy of 0.9859, which demonstrates the reliable and high throughput forgery detection capability.

Table 4: Qualitative comparison of the proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework with recent CMFD approaches on the various dataset

Method	Year	Dice / F1	IoU	Precision	Recall	Accuracy
SPA-Net	2023	0.742	0.681	0.731	0.756	0.972
CAMU-Net	2024	0.768	0.714	0.751	0.782	0.978
CNN-Keypoint Hybrid	2024	0.731	0.669	0.718	0.744	0.968
LBRT Transformer	2024	0.781	0.728	0.764	0.793	0.981
DRNet	2024	0.789	0.736	0.772	0.804	0.983
TransCMFD	2025	0.792	0.741	0.776	0.811	0.984
Frost Filter-Enhanced Encoder-Decoder U-Net	2026	0.7977	0.7447	0.7698	0.8470	0.9859

Limitations

Although the proposed Frost Filter-Enhanced Encoder-Decoder U-Net framework demonstrates promising performance on the CoMoFoD_small_v2 benchmark dataset, certain limitations remain. The localization accuracy may decrease when forged regions are very small or when images undergo severe post-processing operations, such as strong compression, large geometric transformations, or significant noise, which can reduce the discriminative information available for segmentation. In addition, the current study is limited to evaluation on the CoMoFoD_small_v2 dataset. Consequently, further validation on diverse publicly available benchmark datasets and real-world forensic images is necessary to comprehensively assess the robustness and cross-dataset generalization capability of the proposed framework. Future work will therefore focus on extending the evaluation to multiple datasets and improving robustness under more challenging forgery scenarios.

Conclusion

This work proposed a framework of robust pixel-level copy-move forgery localization in digital images based on Frost Filter-Enhanced Encoder-Decoder U-Net. A new method was proposed, which reformulates copy-move forgery detection as a semantic segmentation problem, and incorporates adaptive Frost filtering to enhance noise suppression, feature preservation, and forged boundary refinement, and used a lightweight encoder-decoder architecture. The utilization of the VGG11-BN-based encoder, combined with skip-connected decoder layers, helped to obtain effective multi-scale contextual feature extraction and achieve accurate reconstruction of manipulated regions.

The framework was tested on the CoMoFoD_small_v2 dataset, with results showing that

the proposed framework is robust against different geometric and photometric forgery operations, including scaling, rotation, compression, noise addition, and blurring. The performance of the framework achieved Dice coefficient of 0.7977, IoU of 0.7447, Precision of 0.7698, Recall of 0.8470, F1-score of 0.7977 and overall Accuracy of 0.9859, indicating a good capability of localization and a good discrimination performance. Moreover, the obtained MCC and Cohen's Kappa scores indicated the robustness and consistency of the segmentation masks predicted.

The proposed framework achieves competitive localization accuracy, lower computational complexity, and high throughput, compared to the recent transformer-based and attention-based CMFD approaches. Noise sensitive feature enhancement and false positive segmentation responses reducing were important parts of the overall enhancement that were achieved through the integration of Frost filtering. Generally, the proposed framework exhibits good practical potential in digital forensic applications involving the accurate, lightweight and computationally efficient copy-move forgery detection. Future studies could involve the development of the framework for real-time localization, generalization across datasets, and accurate identification of complex manipulated areas in unconstrained forensic settings.

Funding Information

The authors declare that no specific funding was received for this study.

Author's Contributions

Ashutosh Pandey: Contributed to conceptualization, methodology formulation, experimental analysis, visualization, and manuscript preparation.

Niranjan Lal and Ashish Kumar Chakraverti: Provided technical guidance, validation of the analytical results, and critical manuscript edited. All authors reviewed and approved the final version of the manuscript.

Ethics

The datasets used in this study are publicly available for research purposes. The CoMoFoD_small_v2 dataset can be accessed through the CoMoFoD repository (Chauhan, 2026): These datasets were used solely for academic and research purposes.

Acknowledgment

The authors sincerely acknowledge SRM Institute of Science and Technology (SRMIST), Delhi-NCR Campus, Modinagar, Ghaziabad, Uttar Pradesh, for providing the research facilities and academic support necessary to carry out this work.

References

- Amerini, I., Ballan, L., Caldelli, R., Bimbo, A. D., & Serra, G. (2011). A SIFT-Based Forensic Method for Copy-Move Attack Detection and Transformation Recovery. *IEEE Transactions on Information Forensics and Security*, 6(3), 1099–1110. <https://doi.org/10.1109/tifs.2011.2129512>
- Barni, M., Phan, Q.-T., & Tondi, B. (2021). Copy-Move Source-Target Disambiguation Through Multi-Branch CNNs. *IEEE Transactions on Information Forensics and Security*, 16, 1825–1840. <https://doi.org/10.1109/tifs.2020.3045903>
- Bay, H., Tuytelaars, T., & Van Gool, L. (2006). SURF: Speeded Up Robust Features. *Computer Vision – ECCV*, 3951, 404–417. https://doi.org/10.1007/11744023_32
- Chauhan, T. (2026). CoMoFoD Small v2 [Data set]. Kaggle. [Kaggle.](https://doi.org/10.1016/j.knosys.2023.110508)
- Chen, B., Tan, W., Coatrieux, G., Zheng, Y., & Shi, Y.-Q. (2021). A Serial Image Copy-Move Forgery Localization Scheme With Source/Target Distinguishment. *IEEE Transactions on Multimedia*, 23, 3506–3517. <https://doi.org/10.1109/tmm.2020.3026868>
- Diwan, A., & Roy, A. K. (2024). CNN-Keypoint Based Two-Stage Hybrid Approach for Copy-Move Forgery Detection. *IEEE Access*, 12, 43809–43826. <https://doi.org/10.1109/access.2024.3380460>
- Fridrich, J., Soukal, D., & Lukas, J. (2003). Detection of copy-move forgery in digital images. *2003*, 663–652.
- Khan, M. A., Kadry, S., Nam, Y., & Zhang, Y.-D. (2023). An approach for copy-move image multiple forgery detection based on an optimized pre-trained deep learning model. In *Knowledge-Based Systems* (Vol. 269, p. 110508). <https://doi.org/10.1016/j.knosys.2023.110508>
- Liang, E., Zhang, K., Hua, Z., Li, Y., & Jia, X. (2025). TransCMFD: An adaptive transformer for copy-move forgery detection. *Neurocomputing*, 638, 130110. <https://doi.org/10.1016/j.neucom.2025.130110>
- Liang, P., Li, Z., Tu, H., & Zhao, H. (2024). LBRT: Local-Information-Refined Transformer for Image Copy-Move Forgery Detection. *Sensors*, 24(13), 4143. <https://doi.org/10.3390/s24134143>
- Liu, Y., Xia, C., Zhu, X., & Xu, S. (2022). Two-Stage Copy-Move Forgery Detection With Self Deep Matching and Proposal SuperGlue. *IEEE Transactions on Image Processing*, 31, 541–555. <https://doi.org/10.1109/tip.2021.3132828>
- Lowe, D. G. (2004). Distinctive Image Features from Scale-Invariant Keypoints. *International Journal of Computer Vision*, 60(2), 91–110. <https://doi.org/10.1023/b:visi.0000029664.99615.94>
- Mahdian, B., & Saic, S. (2007). Detection of copy-move forgery using a method based on blur moment invariants. *Forensic Science International*, 171(2–3), 180–189. <https://doi.org/10.1016/j.forsciint.2006.11.002>
- Pan, X., & Lyu, S. (2010). Region Duplication Detection Using Image Feature Matching. *IEEE Transactions on Information Forensics and Security*, 5(4), 857–867. <https://doi.org/10.1109/tifs.2010.2078506>
- Pandey, A., Lal, N., & Chakraverti, A. K. (2025). Correction: Hybrid dense net 201 with CBPN based image pixel enhancement and optimized ADBGAN for image copy-move forgery detection. *Signal, Image and Video Processing*, 19(10). <https://doi.org/10.1007/s11760-025-04502-z>
- Rao, Y., & Ni, J. (2016). A deep learning approach to detection of splicing and copy-move forgeries in images. *2016 IEEE International Workshop on Information Forensics and Security (WIFS)*, 1–6. <https://doi.org/10.1109/wifs.2016.7823911>
- Ronneberger, O., Fischer, P., & Brox, T. (2015). U-Net: Convolutional Networks for Biomedical Image Segmentation. In *Medical Image Computing and Computer-Assisted Intervention – MICCAI* (Vol. 9351, pp. 234–241). Springer International Publishing. https://doi.org/10.1007/978-3-319-24574-4_28
- Wang, J., Gao, X., Nie, J., Wang, X., Huang, L., Nie, W., Jiang, M., & Wei, Z. (2024). Strong robust copy-move forgery detection network based on layer-by-layer decoupling refinement. *Information Processing & Management*, 61(3), 103685. <https://doi.org/10.1016/j.ipm.2024.103685>
- Zhang, Y., Goh, J., Shen, L. L., & Thing, V. L. L. (2023). CNN-transformer based generative adversarial network for copy-move forgery detection. *IEEE Transactions on Information Forensics and Security*, 18, 1741–1757.

- Zhao, K., Yuan, X., Liu, T., Xiang, Y., Xie, Z., Huang, G., & Feng, L. (2024). CAMU-Net: Copy-move forgery detection utilizing coordinate attention and multi-scale feature fusion-based up-sampling. *Expert Systems with Applications*, 238, 121918. <https://doi.org/10.1016/j.eswa.2023.121918>
- Zhao, K., Yuan, X., Xie, Z., Xiang, Y., Huang, G., & Feng, L. (2023). SPA-Net: A Deep Learning Approach Enhanced Using a Span-Partial Structure and Attention Mechanism for Image Copy-Move Forgery Detection. *Sensors*, 23(14), 6430. <https://doi.org/10.3390/s23146430>
- Zhong, J.-L., & Pun, C.-M. (2020). An End-to-End Dense-Inception Net for Image Copy-Move Forgery Detection. *IEEE Transactions on Information Forensics and Security*, 15, 2134–2146. <https://doi.org/10.1109/tifs.2019.2957693>