# An Attention-Enhanced Multi-Scale Framework for Copy–Move Forgery Detection and Localization

**Vanishri. V. Sataraddi**
School of Computer Science and Engineering, REVA University, Bengaluru, Karnataka, India | Department of Computer Science and Engineering, RNS Institute of Technology, Bengaluru, Karnataka, India
vanishrisataraddi@gmail.com (corresponding author)

**N. P. Nethravathi**
School of Computer Science and Engineering, REVA University, Bengaluru, Karnataka, India
nethravathi.np@reva.edu.in

*Received: 25 February 2026 | Revised: 16 March 2026 | Accepted: 24 March 2026*

*Licensed under a CC-BY 4.0 license | Copyright (c) by the authors |* *DOI: https://doi.org/10.48084/etasr.18383*

## ABSTRACT

**When a region in an image is copied and pasted elsewhere within the same image, the resulting forgery is considered a Copy–Move Forgery (CMF). CMF is a challenging problem in digital image forensics, as the copied region retains the original image's statistical properties and can also be manipulated by rotation, scaling, compression, and other transformations. In this paper, we propose an attention-guided deep learning framework for robust Copy–Move Forgery Detection (CMFD) with precise localization. The proposed network architecture aggregates hierarchical convolutional feature maps and boosts the representational saliency through successive channel and spatial attention modules. A multi-scale decoder with integrated self-correlation computation and a lightweight transformer block further encourage the extraction of long-range duplicated patterns, enhancing localization accuracy. Extensive experiments on benchmark datasets validate that this method achieves up to 95% precision, 93% recall, and 94% F1-score, significantly outperforming notable baseline methods. The network also demonstrates robust generalizability against commonly encountered postprocessing operations such as JPEG compression, noise, and moderate geometric transformations. The complete architectural design and mathematical formulation are provided to ensure reproducibility and practical applicability in real-world image forensic scenarios.**

***Keywords-Copy–Move Forgery (CMF); image forensics; attention mechanism; multi-scale; deep learning; forgery localization***

## I. INTRODUCTION

Digital images play an important role in modern communication and decision-making processes in fields such as journalism, forensic investigations, surveillance, and legal documentation. Because visual data are often treated as reliable evidence, maintaining the authenticity and integrity of digital images has become increasingly important. However, the widespread availability of powerful image editing software and artificial intelligence-based generation tools has made it easier to manipulate images without leaving visible traces. As a result, verifying the authenticity of digital images has become a major research challenge in the field of image forensics. Among the various image manipulation techniques, Copy–Move Forgery (CMF) is one of the most frequently used methods for altering image content. In this type of manipulation, a portion of the image is copied and pasted within the same image to hide or duplicate objects. Since the duplicated region originates from the same image, it retains similar statistical characteristics such as color distribution, illumination, and noise patterns, making detection particularly difficult.

Early research in Copy–Move Forgery Detection (CMFD) relied mainly on handcrafted feature-based techniques. One of the earliest approaches was proposed in [1], where Discrete Cosine Transform (DCT) coefficients were used to identify duplicated image blocks through lexicographic sorting. Later, keypoint-based techniques were introduced to improve robustness against geometric transformations. The Scale-Invariant Feature Transform (SIFT) proposed in [2] enabled reliable feature extraction under scale and rotation variations. Building upon this idea, authors in [3] introduced a SIFT-based forensic method that incorporated geometric consistency verification to detect duplicated regions more effectively. Further improvements in block-based approaches were proposed by authors in [4], who introduced dense-field feature representations to improve detection efficiency and localization accuracy.

With the rapid advancement of deep learning technologies, Convolutional Neural Networks (CNNs) have become widely used for CMFD tasks. CNN-based approaches automatically learn discriminative feature representations from image data, reducing the dependency on handcrafted descriptors and improving robustness to various image transformations. Hybrid encoder–decoder architectures have been successfully used for forgery localization, as demonstrated in the work of authors in [5], where deep learning was used to detect and localize manipulated regions. Recent surveys on CMFD also highlight the growing adoption of deep learning methods due to their ability to capture complex visual patterns and improve detection accuracy across diverse datasets [6].

Recent studies have further explored the integration of attention mechanisms and transformer-based architectures to enhance the performance of forgery detection systems. The Convolutional Block Attention Module (CBAM) introduced by authors in [7] applies sequential channel and spatial attention to refine convolutional feature maps and emphasize informative regions. Benchmark datasets have also been developed to facilitate systematic evaluation of CMFD methods. The CoMoFoD dataset introduced by authors in [8] provides manipulated images with various postprocessing operations such as compression, scaling, and rotation. Another commonly used dataset is MICC-F220, which contains realistic copy–move manipulations and supports evaluation of localization performance [9].

More recently, transformer-based architectures have gained attention in the field of image forgery detection due to their ability to capture long-range contextual relationships. For example, CMFD Former utilizes transformer self-attention to model global feature dependencies and improve CMFD accuracy [10]. Similarly, CAMU-Net introduces a multi-scale U-Net architecture that enhances forgery localization through hierarchical feature fusion [11]. Transformer-based models with refined local context information have also been explored for image forgery detection tasks [12]. Several recent studies and surveys indicate that combining attention mechanisms, transformer architectures, and multi-scale feature learning can significantly improve the robustness and accuracy of forgery detection systems [13, 14]. Other works have explored clustering-based strategies and systematic performance evaluations of CMF algorithms, demonstrating the importance of parameter selection and dataset characteristics in determining detection performance [15, 16]. Attention-based deep learning frameworks have also shown promising improvements in localization accuracy and generalization capability across multiple datasets [17]. Comprehensive surveys further highlight the challenges associated with dataset bias, model generalization, and robustness in practical forensic applications [18].

Publicly available benchmark datasets continue to play a crucial role in developing and evaluating CMFD methods. Recent research has demonstrated the effectiveness of machine learning-based forgery detection approaches across multiple datasets [19]. In particular, datasets such as MICC-F220 [20], CASIA v2.0 [21], and CoMoFoD [22] provide large collections of manipulated and authentic images, making them valuable resources for evaluating the performance of modern deep learning-based forgery detection systems.

Motivated by these developments, this paper proposes an attention-enhanced multi-scale deep learning framework for CMFD and localization. The proposed architecture integrates hierarchical convolutional feature extraction with channel and spatial attention mechanisms to improve discriminative feature representation. In addition, a lightweight transformer module is incorporated at the bottleneck stage to capture long-range self-similarity patterns that characterize copy–move manipulation. Multi-scale feature fusion in the decoder enables precise pixel-level localization of forged regions. Extensive experimental evaluation on benchmark datasets demonstrates that the proposed method achieves higher detection accuracy and robustness compared with existing classical and deep learning-based approaches.

## II. PROPOSED METHOD

This section presents a multi-scale deep learning framework that utilizes attention mechanisms to detect and locate CMFs. The proposed framework extracts and refines hierarchical convolutional features by sequentially applying channel and spatial attention, thereby improving the detection capabilities of potential forgery cues. A lightweight transformer-based self-correlation module is included to capture long-range dependencies between duplicated areas. Multi-scale feature fusion in the decoder enables precise pixel-level localization of forged areas.

### *A. Overview*

Figure 1 shows the overall structure of the proposed attention-guided CMFD framework. The system processes the input image using a CNN to extract layered feature representations. It applies channel and spatial attention modules to improve the extracted features by highlighting important forgery-related details. The improved features are then used for forgery localization, resulting in a pixel-wise output mask that highlights the detected copy–move forged area.

### *B. Preprocessing*

Given an RGB image $I \in \mathbb{R}^{H\times W\times 3}$, we resize it to a standard input $I_s$ (e.g., 512 × 512 or 256 × 256 depending on the GPU) and standardize pixel intensities to [0,1]. Optionally, the image can be converted to a single-channel luminance representation $L$ for a computationally cheaper variant.

### *C. Encoder: Multi-Scale Convolutional Features*

Let $E(\cdot)$ denote the encoder mapping. The encoder produces a set of multi-scale feature maps:

$$\{F_1, F_2, F_3, F_4\} = E(I_s),$$

where $F_\ell \in \mathbb{R}^{C_\ell\times H_\ell\times W_\ell}$, and $H_\ell$, $W_\ell$ decrease with $\ell$.

Individual convolutional blocks follow:

$$F_\ell = \phi(W_\ell * F_{\ell-1} + b_\ell) \qquad (1)$$

where $W_\ell$ are convolutional kernels, $*$ denotes convolution, $b_\ell$ are biases, and $\phi$ is ReLU. (For the first block $F_0 = I_s$.)

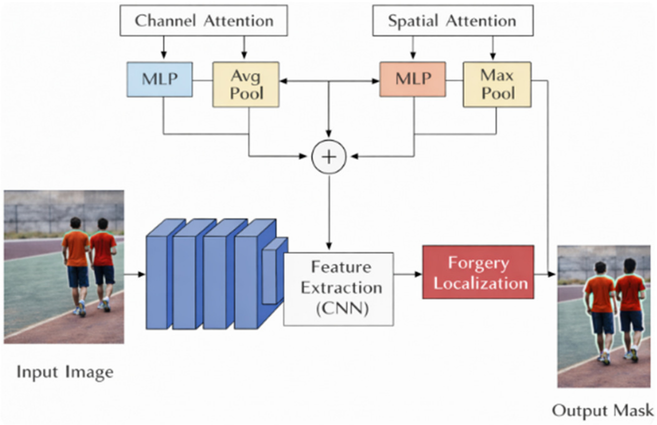

Fig. 1. Proposed attention-guided multi-scale deep network architecture for CMFD and localization.

### D. *Channel Attention (CBAM Channel Branch)*

Given a feature map $F \in \mathbb{R}^{C\times H\times W}$, we compute channel attention $M_c(F) \in \mathbb{R}^{C\times 1\times 1}$ by aggregating spatial information and passing through a small Multilayer Perceptron (MLP):

$$z_{avg} = \text{AvgPool}(F), z_{max} = \text{MaxPool}(F)$$

$$M_c(F) = \sigma\left(\text{MLP}(z_{avg}) + \text{MLP}(z_{max})\right) \qquad (2)$$

where $\sigma$ is the sigmoid. The refined feature after channel attention is:

$$\tilde{F} = M_c(F) \odot F \qquad (3)$$

with $\odot$ denoting channel-wise broadcast multiplication. This follows CBAM [7].

### E. *Spatial Attention (CBAM Spatial Branch)*

Spatial attention is applied to $\tilde{F}$ to yield location emphasis:

$$f_{avg} = \text{Mean}_c(\tilde{F}), f_{max} = \text{Max}_c(\tilde{F})$$

$$M_s(\tilde{F}) = \sigma\left(f^{7\times7}([f_{avg}; f_{max}])\right) \qquad (4)$$

$$F' = M_s(\tilde{F}) \odot \tilde{F} \qquad (5)$$

Here, $f^{7\times7}$ is a $7 \times 7$ convolution and $[;]$ denotes concatenation of the two 2D maps. Sequential application of (2)–(5) yields attention-refined features with improved robustness to background clutter.

### F. *Bottleneck Self-Correlation Transformer Block*

CMF is characterized by self-similarity, where one region matches another region in the same image. To explicitly capture this, at the encoder bottleneck we reshape $F'$ into a token sequence and apply a lightweight transformer encoder that computes self-attention across spatial tokens. Let $F'_b$ be the bottleneck feature of size $C_b \times H_b \times W_b$. Reshape to $X \in \mathbb{R}^{N\times C_b}$, with $N = H_b W_b$. Self-attention outputs:

$$\text{Attention}(Q, K, V) = \text{softmax}\left(\frac{QK^T}{\sqrt{d_k}}\right)V \qquad (6)$$

where $Q$, $K$, and $V$ represent the query, key, and value projections derived from the bottleneck feature tokens $X$. We use a small number of attention heads and limit the transformer depth to maintain computational efficiency.

In our implementation, the transformer block uses 4 attention heads with 2 encoder layers and a hidden dimension of 256 to balance computational efficiency and contextual feature modeling.

Compact transformer designs, such as Mix Transformer (MiT) or lightweight Vision Transformer (ViT) variants, inspire this configuration and help capture global self-similarity patterns important for CMFD.

### G. *Decoder and Multi-Scale Fusion*

The decoder upsamples the bottleneck features while fusing encoder skip connections at each level. We include Atrous Spatial Pyramid Pooling (ASPP)-like parallel dilated convolutions to aggregate context across receptive fields (dilations 1, 6, 12, 18). The decoder produces a logit map $Z \in \mathbb{R}^{2\times H\times W}$ (two classes: forged / non-forged).

### H. *Loss and Training*

We train with a combined pixel-wise Binary Cross Entropy (BCE) loss and Dice loss to balance pixel accuracy and region overlap:

$$\mathcal{L} = \lambda_1 \text{BCE}(Z, G) + \lambda_2\left(1 - \text{Dice}(P, G)\right) \qquad (7)$$

where $G$ is the ground truth mask, $P = \text{softmax}(Z)$ (per pixel probability), and $\lambda_1, \lambda_2$ are balancing weights (typical values: 1.0 each). Dice is defined as:

$$\text{Dice}(P, G) = \frac{2\sum_{x,y} P(x,y)G(x,y)}{\sum_{x,y} P(x,y) + \sum_{x,y} G(x,y) + \epsilon} \qquad (8)$$

### I. *Postprocessing*

The probability map $P_{forged}(x, y)$ is thresholded at $t$ (default: 0.5), then morphological opening and closing are applied to remove spurious regions and smooth boundaries. Optionally Conditional Random Fields (CRFs) are applied as refinement.

## III. ALGORITHM

The overall workflow of the proposed attention-guided CMFD framework is summarized in Algorithm 1.

```
Algorithm 1. Attention-guided CMFD
Input: RGB image $I$
Output: Binary forgery mask $M$
1. Preprocess: resize $I \rightarrow I_s$, normalize
2. Encoder: compute multi-scale features
   $\{F_1, F_2, F_3, F_4\} = E(I_s)$, using (1)
3. For each scale $\ell$:
   a. Compute channel attention $M_c(F_\ell)$ via
      (2)
   b. Refine $\tilde{F}_\ell = M_c(F_\ell) \odot F_\ell$ using (3)
```

```
   c. Compute spatial attention $M_s(\tilde{F}_\ell)$ via
      (4)
   d. Compute refined feature $F'_\ell = M_s(\tilde{F}_\ell) \odot \tilde{F}_\ell$
      using (5)
4. Bottleneck: reshape $F'_b \rightarrow$ tokens $X$, apply
   transformer self-attention (6) to
   compute global correlations
5. Decoder: upsample and fuse multi-scale
   $F'_\ell$ via skip connections and ASPP to
   obtain logits $Z$
6. Compute softmax probabilities $P$ from $Z$
7. Loss: compute $\mathcal{L}$ using (7) and (8);
   train end-to-end
8. Inference: threshold $P_{forged} > t$, apply
   morphological refinement $\rightarrow M$
9. Return $M$
```

The attention-guided CMFD algorithm begins with preprocessing of the input image and the extraction of multi-scale features through an encoder. Channel and spatial attention modules are applied to the features to emphasize areas that might potentially contain copied content. The refined bottleneck features are then transformed using transformer self-attention to identify long-range similarity patterns, which are indicative of CMF. Finally, a decoder is used to combine multi-scale features to produce dense prediction maps, which are transformed into pixel probabilities. The forged probability map is then thresholded and post-processed to obtain the binary forgery mask.

## IV. EXPERIMENTAL EVALUATION

### A. Datasets

The proposed framework is evaluated on three widely used benchmark datasets for CMFD: CoMoFoD, MICC-F220, and CASIA v2.0. The CoMoFoD dataset provides images containing various postprocessing operations such as compression, rotation, scaling, and noise, making it suitable for evaluating the robustness of CMFD algorithms under realistic manipulation conditions [8]. The MICC-F220 dataset consists of 220 images containing realistic copy–move manipulations and is widely used to assess the localization performance of detection methods [9]. The CASIA v2.0 dataset contains a large collection of tampered and authentic images with multiple manipulation types and is commonly used as a benchmark for image forgery detection research. The datasets used in this work are publicly available and can be accessed through their respective online repositories [20-22].

### B. Baseline Methods

To determine the effectiveness of the proposed approach, comparative experiments are conducted against representative classical and recent state-of-the-art CMFD methods, including:

- DCT–PCA: A classical block-based method that employs DCT coefficients followed by Principal Component Analysis (PCA) and lexicographic sorting for duplicate block matching.
- SIFT-based method: A keypoint-based approach that utilizes SIFT descriptors combined with Random Sample Consensus (RANSAC)-based geometric verification for detecting duplicated regions.
- CNN-only model: A variant of the proposed architecture that uses the same encoder–decoder structure but excludes attention mechanisms and transformer components, serving as an ablation baseline.
- Transformer-based CMFD: A recent transformer-driven CMFD model that exploits self-attention mechanisms to capture long-range feature dependencies.
- CAMU-Net and variants: Attention-guided U-Net-based architectures that employ multi-scale feature fusion and attention mechanisms for forgery localization.

These baselines are selected to represent a diverse range of methodological categories, including handcrafted feature-based methods, pure CNN-based models, and modern attention/transformer-based approaches.

### C. Implementation Details

The proposed CMFD framework was implemented using the PyTorch deep learning library. All input images were resized to a fixed resolution and normalized before training. Data augmentation was used to improve generalization and robustness. This included random rotation (±45°), scaling (0.8–1.2), horizontal and vertical flipping, JPEG compression with quality factors from 40 to 100, and additive Gaussian noise. The encoder network is based on a modified ResNet-34 backbone. Channel and spatial attention modules are added after each major convolutional block to improve feature learning. A lightweight transformer block is placed at the bottleneck to model long-range dependencies and capture self-similarity patterns typically seen in CMFs. The decoder follows an encoder–decoder structure with skip connections and incorporates multi-scale feature fusion using parallel dilated convolutions. The network was trained using the Adam optimizer with an initial learning rate of $1 \times 10^{-4}$ and a batch size of 8.

The proposed architecture maintains moderate computational complexity due to the use of a lightweight transformer module at the bottleneck stage. The transformer block is designed with a limited number of attention heads and encoder layers to reduce the overall parameter count. Compared with fully transformer-based models, this compact design reduces computational cost while preserving the ability to capture long-range dependencies. As a result, the model achieves efficient inference while maintaining strong forgery localization performance.

Training lasted for 50 epochs, with early stopping based on the validation F1-score to avoid overfitting. The loss function combines pixel-wise cross-entropy loss with a soft Dice loss to handle class imbalance between forged and non-forged areas. All experiments were conducted on a workstation equipped with an NVIDIA GPU, and the trained model that performed best in validation was used for final evaluation.

## V. RESULTS AND DISCUSSION

### *A. Quantitative Results*

To quantitatively evaluate the proposed method, Table I compares its performance with representative CMFD approaches using precision, recall, and F1-score on benchmark datasets.

### *B. Qualitative Examples*

Figure 2 shows the results of CMF localization on a typical CoMoFoD image. The CNN-only model provides fragmented localization and overlooks parts of the forged areas because it lacks sufficient context. In contrast, the proposed attention-guided framework correctly identifies duplicated regions with clearer boundaries. These outcomes highlight how attention mechanisms can improve forgery localization performance.

TABLE I. COMPARATIVE PERFORMANCE OF DIFFERENT METHODS ON BENCHMARK DATASETS

| **Method** | **CoMoFoD** | | | **MICC-F220** | | | **CASIA v2.0** | | |
|---|---|---|---|---|---|---|---|---|---|
| | **Precision** | **Recall** | **F1-score** | **Precision** | **Recall** | **F1-score** | **Precision** | **Recall** | **F1-score** |
| DCT–PCA [1] | 0.76 | 0.72 | 0.74 | 0.78 | 0.74 | 0.76 | 0.73 | 0.69 | 0.71 |
| SIFT-based [3] | 0.80 | 0.77 | 0.78 | 0.82 | 0.78 | 0.80 | 0.76 | 0.72 | 0.74 |
| CNN only [5] | 0.88 | 0.83 | 0.85 | 0.89 | 0.84 | 0.86 | 0.85 | 0.81 | 0.83 |
| CMFDFormer [10] | 0.90 | 0.86 | 0.88 | 0.91 | 0.87 | 0.89 | 0.89 | 0.85 | 0.87 |
| CAMU-Net [11] | 0.91 | 0.88 | 0.89 | 0.92 | 0.88 | 0.90 | 0.90 | 0.86 | 0.88 |
| Proposed | 0.95 | 0.93 | 0.94 | 0.94 | 0.93 | 0.94 | 0.95 | 0.92 | 0.93 |

Note: All precision, recall, and F1-score values are reported in normalized form.

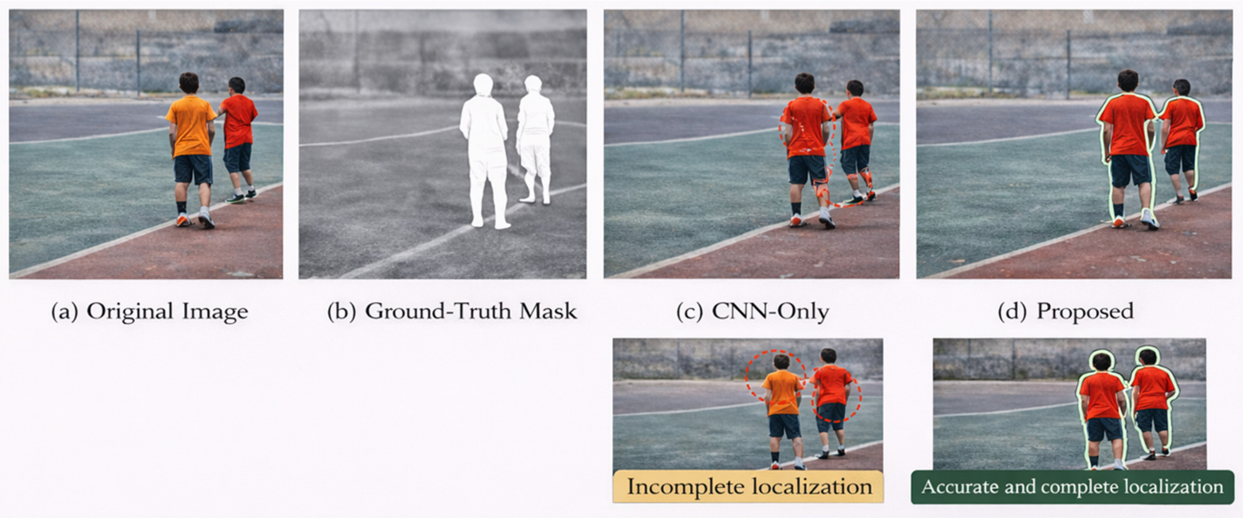

Fig. 2. Qualitative CMF localization results on the CoMoFoD dataset using a representative test image. (a) original image, (b) ground-truth forgery mask, (c) CNN model output showing fragmented localization and missed forged regions (highlighted by red dashed circles), and (d) proposed attention-guided model output producing a cleaner and more accurate forgery mask with accurate boundaries (highlighted by green outline).

### *C. Performance Discussion*

The proposed method achieves consistently higher precision, recall, and F1-score across all three datasets: CoMoFoD, MICC-F220, and CASIA v2.0. Minor performance variations can be attributed to differences in image content and forgery complexity. The similarity of the reported values is due to rounding to two decimal places. Overall, these stable results demonstrate the strong generalization capability of the proposed attention-guided framework across diverse CMF scenarios.

Figures 3, 4, and 5 illustrate that the deep learning models have better precision, recall, and F1-score compared to the conventional models. The addition of attention mechanisms further enhances recall and the ability to identify the forged regions. The best F1-scores are obtained by the proposed model, which is an indicator of a good trade-off between precision and recall.

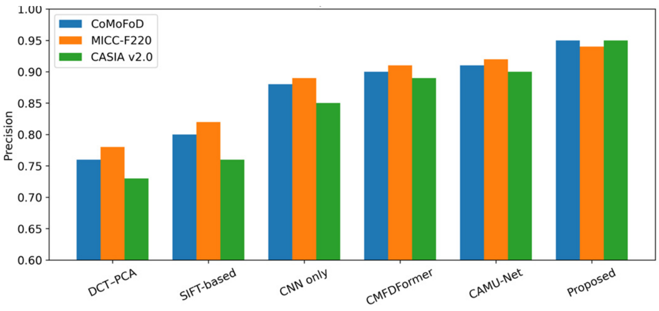

Fig. 3. Precision comparison across datasets**.**

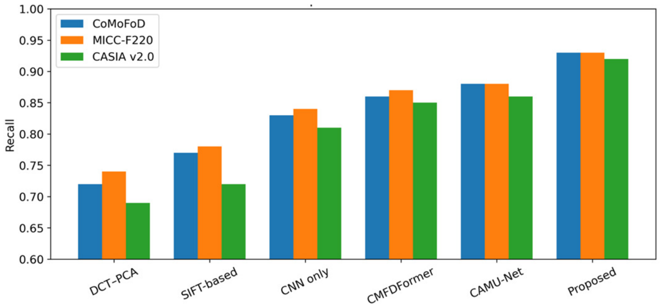

Fig. 4. Recall comparison across datasets.

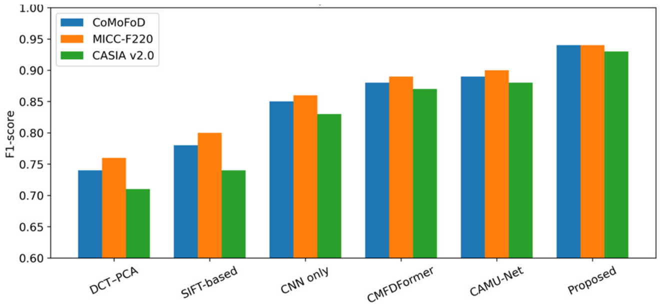

Fig. 5. F1-score comparison across datasets.

## VI. CONCLUSION

This paper presented an attention-enhanced multi-scale deep learning framework for detecting and localizing Copy–Move Forgery (CMF) in digital images. The proposed architecture integrates hierarchical convolutional feature extraction with channel and spatial attention mechanisms to improve discriminative feature representation. In addition, a lightweight transformer module is incorporated to capture long-range self-similarity patterns that are characteristic of copy–move manipulation.

Experimental evaluation on three widely used benchmark datasets, namely CoMoFoD, MICC-F220, and CASIA v2.0, demonstrates that the proposed approach achieves up to 95% precision, 93% recall, and 94% F1-score, outperforming several classical and recent deep learning-based Copy–Move Forgery Detection (CMFD) methods. The results further indicate that the proposed model maintains reliable performance under common postprocessing operations, such as JPEG compression and additive noise. Overall, these findings highlight the effectiveness and robustness of the proposed framework for practical digital image forensic applications.

Future work will focus on extending the framework to handle more complex image manipulations and improving computational efficiency for large-scale forensic analysis.

## DECLARATION OF COMPETING INTERESTS

The authors declare that they have no conflict of interest.

## ACKNOWLEDGMENT

This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors. The authors would like to express their sincere appreciation to all individuals who contributed to the successful completion of this work. The authors also thank the reviewers and the editorial team for their constructive comments and valuable suggestions, which helped improve the quality of this manuscript.

## DATA AVAILABILITY

The datasets used in this study are publicly available benchmark datasets, namely CoMoFoD, MICC-F220, and CASIA v2.0, which can be accessed through their respective online repositories [20-22].

## REFERENCES

[1] J. Fridrich, D. Soukal, and J. Lukáš, "Detection of Copy-Move Forgery in Digital Images," , 2003. - References - Scientific Research Publishing," in *Proceedings of the Digital Forensic Research Workshop, Cleveland*, Cleveland, OH, USA, 2003.

[2] D. G. Lowe, "Distinctive Image Features from Scale-Invariant Keypoints," *International Journal of Computer Vision*, vol. 60, no. 2, pp. 91–110, Nov. 2004, https://doi.org/10.1023/B:VISI.0000029664.99615.94.

[3] I. Amerini, L. Ballan, R. Caldelli, A. Del Bimbo, and G. Serra, "A SIFT-Based Forensic Method for Copy–Move Attack Detection and Transformation Recovery," *IEEE Transactions on Information Forensics and Security*, vol. 6, no. 3, pp. 1099–1110, Sept. 2011, https://doi.org/10.1109/TIFS.2011.2129512.

[4] D. Cozzolino, G. Poggi, and L. Verdoliva, "Efficient Dense-Field Copy–Move Forgery Detection," *IEEE Transactions on Information Forensics and Security*, vol. 10, no. 11, pp. 2284–2297, Nov. 2015, https://doi.org/10.1109/TIFS.2015.2455334.

[5] J. H. Bappy, C. Simons, L. Nataraj, B. S. Manjunath, and A. K. Roy-Chowdhury, "Hybrid LSTM and Encoder–Decoder Architecture for Detection of Image Forgeries," *IEEE Transactions on Image Processing*, vol. 28, no. 7, pp. 3286–3300, July 2019, https://doi.org/10.1109/TIP.2019.2895466.

[6] F. Z. Mehrjardi, A. M. Latif, M. S. Zarchi, and R. Sheikhpour, "A survey on deep learning-based image forgery detection," *Pattern Recognition*, vol. 144, Dec. 2023, Art. no. 109778, https://doi.org/10.1016/j.patcog.2023.109778.

[7] S. Woo, J. Park, J.-Y. Lee, and I. S. Kweon, "CBAM: Convolutional Block Attention Module," in *15th European Conference on Computer Vision*, Munich, Germany, 2018, pp. 3–19, https://doi.org/10.1007/978-3-030-01234-2_1.

[8] D. Tralic, I. Zupancic, S. Grgic, and M. Grgic, "CoMoFoD — New database for copy-move forgery detection," in *Proceedings ELMAR-2013*, Zadar, Croatia, 2013, pp. 49–54.

[9] "Copy-Move Forgery Detection and Localization." Image and Communication Laboratory. https://lci.micc.unifi.it/labd/2015/01/copy-move-forgery-detection-and-localization/.

[10] Y. Liu *et al.*, "CMFDFormer: Transformer-based Copy-Move Forgery Detection with Continual Learning." arXiv, Nov. 22, 2023, https://doi.org/10.48550/arXiv.2311.13263.

[11] K. Zhao *et al.*, "CAMU-Net: Copy-move forgery detection utilizing coordinate attention and multi-scale feature fusion-based up-sampling," *Expert Systems with Applications*, vol. 238, Mar. 2024, Art. no. 121918, https://doi.org/10.1016/j.eswa.2023.121918.

[12] P. Liang, Z. Li, H. Tu, and H. Zhao, "LBRT: Local-Information-Refined Transformer for Image Copy–Move Forgery Detection," *Sensors*, vol. 24, no. 13, June 2024, Art. no. 4143, https://doi.org/10.3390/s24134143.

[13] 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, no. 12, pp. 17521–17566, May 2023, https://doi.org/10.1007/s11042-022-13797-w.

[14] S. Sharma, B. Singh, and H. Garg, "Robust Image Forgery Localization Using Hybrid CNN-Transformer Synergy Based Framework," *Computers, Materials & Continua*, vol. 82, no. 3, pp. 4691–4708, Mar. 2025, https://doi.org/10.32604/cmc.2025.061252.

[15] A. Hegazi, A. Taha, and M. M. Selim, "An improved copy-move forgery detection based on density-based clustering and guaranteed outlier removal," *Journal of King Saud University - Computer and Information Sciences*, vol. 33, no. 9, pp. 1055–1063, Nov. 2021, https://doi.org/10.1016/j.jksuci.2019.07.007.

[16] O. M. Al-Qershi and B. E. Khoo, "Evaluation of copy-move forgery detection: datasets and evaluation metrics," *Multimedia Tools and Applications*, vol. 77, no. 24, pp. 31807–31833, Dec. 2018, https://doi.org/10.1007/s11042-018-6201-4.

[17] P. Selvaraj and A. B. S, "Attention based deep learning model for detecting copy move forgery." Research Square, Oct. 01, 2025, https://doi.org/10.21203/rs.3.rs-7596332/v1.

[18] P. B. Shailaja Rani and A. Kumar, "Digital Image Forgery Detection Techniques: A Comprehensive Review," in *2019 3rd International conference on Electronics, Communication and Aerospace Technology*, Coimbatore, India, 2019, pp. 959–963, https://doi.org/10.1109/ICECA.2019.8822064.

[19] H. A. Raheem, M. A. Mohammed, and A. S. H. M. Ali, "Enhancing the Image Forgery Detection based Machine Learning Approach using Multiple Datasets," *Engineering, Technology & Applied Science Research*, vol. 15, no. 3, pp. 22739–22745, June 2025, https://doi.org/10.48084/etasr.10151.

[20] "MICC-F220." Kaggle. [Online]. Available: https://www.kaggle.com/datasets/mashraffarouk/micc-f220.

[21] "CASIA 2.0 Image Tampering Detection Dataset." Kaggle. [Online]. Available: https://www.kaggle.com/datasets/divg07/casia-20-image-tampering-detection-dataset.

[22] "CoMoFoD." Video Communications Laboratory, Department of Communication and Space Technologies, University of Zagreb. [Online]. Available: https://www.vcl.fer.hr/comofod/.