Available at https://www.iasj.net/iasj
Iraqi Academic Scientific Journals
Journal homepage:
https://journals.uokerbala.edu.iq/index.php/UOKJ

Research Article

Enhancing Copy–Move Forgery Detection with Contrastive Pretraining and Tasmanian Devil-AdamW Optimization

Israa Abdulkadhim Jabbar Al Ali

College of Education for Human Sciences, University of Kerbala, Kerbala, Iraq.

Article Info
Article history:
Received 16 -10-2025
Received in revised form 1-12-2025
Accepted 21-12-2025
Available online 31 -12 -2025

Keywords:
Contrastive Self-Supervised Learning, Copy–Move Forgery Detection, Dynamic Hybrid Optimization, Tasmanian Devil Optimizer (TDO), Spatio-Temporal Feature Learning.

Abstract:

Copy–move forgery detection is challenging due to deep, subtle models, limited annotated data, and hyperparameter tuning. We present a robust framework combining contrastive self-supervised pretraining with dynamic multi-step optimization to ensure stable convergence, high accuracy, and improved generalization. In the first step, a backbone network is pretrained by applying a contrastive paradigm (BYOL, SimCLR, MoCo v3) on broadly tagged video frames augmented with forgery-aware transformations, including mild geometric changes, compression, synthetic copy–move, and blurring. This step learns invariant and discriminative visual representations without heavy dependence on tags. In the second step, the pretrained encoder is fine-tuned with a lightweight temporal head to capture the spatio-temporal inconsistencies indicative of copy–move manipulations. Training is performed using a dynamic hybrid optimizer: The Tasmanian Devil Optimizer (TDO) performs early-step global exploration across mild architectural knobs and hyperparameters, after which optimization switches to AdamW to guarantee fine convergence and effective exploitation, mitigating local minima. Broad assessments (such as GRIP and VTD) show stable improvements across robust baselines in AUC, accuracy, and F1, while attention maps provide interpretable localization of tampered areas. The proposed model reduces reliance on labels, enhances robustness, and exhibits faster and more stable training dynamics.

Corresponding Author E-mail: israa.jabbar@uokerbala.edu.iq
Peer review under responsibility of Iraqi Academic Scientific Journal and University of Kerbala.

1. Introduction

Copy–move image forgery is a challenging form of digital manipulation in which a region of an image is copied and pasted into a similar area to replicate or conceal objects [1]. Compared with other tampering types, copy–move forgeries are particularly difficult to detect because the forged region closely matches the background in noise, color, and texture. This inherent similarity, compounded by post-processing operations such as blurring, compression, rotation, and scaling, significantly complicates detection [2].

Despite recent advances in deep learning (DL) approaches, three major challenges remain. First, large-scale, well-annotated datasets for copy–move forgeries are scarce, limiting model generalization. Second, DL models are often sensitive to realistic transformations and post-processing applied by forgers [3]. Third, deep neural networks (DNNs) exhibit strong hyperparameter sensitivity, where inappropriate learning rates or optimizer settings can cause convergence instability [4].

To address these challenges, we propose a framework combining contrastive self-supervised pretraining with fine-tuning and a dynamic hybrid optimization strategy. In the first step, the backbone is pretrained using contrastive learning techniques such as BYOL [5], MoCo v3 [6], and SimCLR [7] on large untagged image sets to learn discriminative forgery-related features. In the second step, the pretrained model is fine-tuned on labeled copy–move datasets for accurate classification and localization. The hybrid optimizer applies the Tasmanian Devil Optimizer (TDO) [8] for global hyperparameter exploration, followed by AdamW for fine convergence and exploitation, enhancing training stability and robustness.

The present study's contributions could be summarized as:

1. We offer the contrastive self-supervised pretraining approach tailored for copy–move forgery detection, making efficient feature learning possible from untagged data.
2. We provide the active hybrid optimization model, which integrates TDO for global exploration with AdamW for consistent fine-tuning, guaranteeing the two strengths and convergence speed.
3. We show via broad tests that the offered framework considerably improves the accuracy of diagnosis and generalization under difficult post-processing situations, performing better than new baselines.

The remainder of this paper is organized as follows. Section 2 reviews related work on copy–move forgery detection, including traditional, deep learning, and self-supervised methods, as well as optimization strategies. Section 3 details the proposed framework, covering contrastive pretraining, fine-tuning, and the dynamic hybrid optimization approach. Section 4 describes the experimental setup, including datasets, preprocessing, evaluation metrics, and implementation details. Also, it presents and discusses the results, including quantitative comparisons, qualitative visualizations, ablation studies, and robustness analysis. Finally, Section 5 concludes the paper and outlines potential directions for future work.

2. Related Work

Copy–move forgery detection has drawn considerable attention in the last few years because of the broad use of digital pictures in security, legal, forensic, and social media apps. Diagnosing these approaches' forgeries is especially difficult due to the tampered areas originating from a similar picture, making subtle alterations hard to recognize. Traditional techniques sometimes depend on model matching and handcrafted features that are restricted in controlling large sets of data and complicated transformations. On the

contrary, methods of transfer learning and DL illustrated considerable performance in discriminative features extraction and generalization over different sets of data. Present part reviews the new strategies in copy–move forgery detection, highlighting their restrictions, methods, and strengths. The discussion is outlined for presenting a perspective on model architecture, robustness, feature extraction, optimization approaches, and laying the groundwork for the presented framework.

Prem Kumar et al. [9] offered the Optimal Deep Transfer Learning-based Copy-Move Forgery Detection (ODTL-CMFD) method for grouping pictures as original/tampered and localizing the copy-moved areas. The technique integrates MobileNet for feature extraction optimized by the Political Optimizer (PO) and a Least Squares Support Vector Machine (LS-SVM) developed with an Enhanced Bird Swarm Algorithm (EBSA) to fine-tune the parameters of the classifier. Assessed on datasets of MICC-F220, MICC-F2000, and MICC-F600, it showed the developed performance of diagnosis. The basic benefit is transfer learning combined with heuristic optimization, developing classification accuracy as well as feature representation. Although multiple optimization stages and separate steps enhance training time as well as computational complexity.

Chaitra and Reddy [10] provide a transfer learning-based Copy-Move Forgery Detection technique applying a Deep CNN initialized with pre-trained GoogLeNet parameters. In addition, a Fractional Leader Harris Hawks Optimization (FLHHO) mechanism was used for optimizing the weights and biases of the network. The method obtained high accuracy (0.93) and balanced True Positive and True Negative rates on benchmark sets of data. Its basic strength depends on integrating pre-trained networks with optimization for developing generalization. Although the technique still heavily relies on labeled sets of data, optimization could be computationally intensive.

Khalil et al. [11] provided a DL-based strategy leveraging transfer learning for diagnosing the two copy-move and splicing forgeries. The technique makes a featured image by giving the difference between the original and compressed pictures, feeding it into pre-trained models with fine-tuned classifiers. Tests illustrated that MobileNetV2 presented high diagnosis accuracy (~95%) with fewer parameters of training, causing quicker training. The benefit is effective multi-forgery diagnosis with light models, when restriction is dependent on high-quality compression analysis that might fail under strict noise/ post-processing.

Chaitra and Reddy [12] offered TFRA-ShuffleNet, a hybrid forgery detection framework applying ShuffleNet integrated with a Transit Flow Regime Algorithm (TFRA). Features were extracted from object-diagnosed areas through YOLO V3, applying hybrid descriptors (LBP, LGXP, LOOP, LVP, PHoG, LDP, LDTP). The model obtained high accuracy, TNR (~96–97%), and TPR. Its benefit is strong multi-feature extraction and diagnosis ability. Restrictions such as high model complexity and reliance on appropriate object diagnosis.

Maashi et al. [13] applied the same strategy of transfer learning to diagnose copy-move and splicing forgeries with pre-trained models. MobileNetV2 presented high accuracy with decreased parameters of training. The benefit is computational effectiveness and efficiency in the diagnosis of multiple kinds; the restriction is sensitivity to picture compression artifacts and the quality of the dataset.

Bevinamarad et al. [14] presented a framework integrating Stationary Wavelet Transform (SWT) with a Hybrid Dilated Adaptive VGG16 (HDA-VGG16) scheme. The Hybridized Tuna Swarm with Bald Eagle Search Optimization (HTS-BESO) was used for optimizing parameters of the

network, and multi-similarity matching was applied for localization. Benefits such as accurate deep feature extraction, as well as localization. Restrictions are reliance on patch-based processing and the developed computational cost that might be slow for wide pictures.

Vaishnavi et al. [15] provided IDL-CMIFD, a DL framework for copy-move diagnosis applying EfficientNet as a feature extractor with Chaotic Monarch Butterfly Optimization (CMBO) and Adam optimizer for tuning a Deep Wavelet Neural Network classifier. Assessed on MICC sets of data, it obtained high performance. The strength of the technique is optimized robust classification and deep feature extraction, when the restrictions contain complexity because of multiple heuristic optimizations.

Li et al. [16] offered the learning framework of graph representation for copy-move diagnosis, modeling the two short- and long-range correlations between patches of the image. A cascaded graph learning and hierarchical cross-attention algorithm was developed for tampering identification and patch representation. Benefits include efficient patch relations modeling and developed diagnosis in ambiguous areas. Restrictions include high computational complexity and dependence on accurate graph construction.

Chaube et al. [17] developed the ACO-Enhanced Siamese Network for copy-move diagnosis. The Siamese Network extracts patch embeddings when Ant Colony Optimization (ACO) optimizes the main feature point selection. This strategy developed strength for geometric transformations and obtained high accuracy (~97.13%). Benefits include efficient scalability and feature matching; restrictions are computational overhead and slower inference because of patch-wise processing.

Yadav et al. [18] provided the encoder–decoder network with LSTM for tamper localization. The encoder (ResNet50 backbone) extracts spatial features when LSTM models transition among unaltered and altered areas. Fused features are decoded for pixel-wise localization. Benefits include accurate multi-level feature combination as well as tamper localization. Restrictions contain moderate AUC performance on several sets of data and higher model complexity.

Alfraihi et al. [19] provided ECMVFD-FTLTDO, a fusion-based transfer learning strategy for video copy-move forgery diagnosis. Three pre-trained models (EfficientNetB7, ResNet50, MobileNetV3) were integrated for feature extraction, and an Elman RNN classifier was optimized by applying Tasmanian Devil Optimizer (TDO). Benefits include strength-optimized diagnosis, feature fusion, and obtaining high accuracy on GRIP and VTD sets of data. Restrictions are developed to increase computational cost because of multi-model fusion and sequential optimization.

Zhao et al. [20] offered a light picture forgery detection technique given the multi-model fusion. Their strategy combines some light models that extract complementary feature dimensions from pictures, and the last diagnosis is performed on the fused features. This technique obtains high diagnosis accuracies of 91.5% and 88.8% on the MICC-F220 and MICC-F600 sets of data, respectively. This strategy's basic benefit refers to its low computational cost, which facilitates effective development and training. Although a restriction refers to the fact that the technique concentrates on image-level diagnosis and might not consider video forgery/more complicated tampering scenarios.

Prathibha and Tamizharasan [21] provided a light forgery diagnosis model for videos given the Separable Convolutional Networks, targeting copy-move /object-based tampering. Their technique is able to diagnose objects copied over frames, guaranteeing authenticity in legal and

forensic contexts. This is deployable on edge devices, making real-life video forgery diagnosis possible, and was assessed on MICC-F220, MICC-F2000, and Rewind video sets of data with developed performance. The benefit depends on its suitability for real-life and resource-limited areas, with a potential restriction being that it basically considers copy-move forgery and might not generalize to other kinds of video manipulations.

Table 1: Comparison of Methods and Models

#	Author&Year	Method/Model	Dataset(s)	Advantages	Limitations
[9]	Prem Kumar **et al., 2023**	ODTL-CMFD: MobileNet + PO + LS-SVM with EBSA	MICC-F220, MICC-F2000, MICC-F600	Integration of transfer learning with heuristic optimization; enhanced classification accuracy	Multiple optimization steps; higher computational complexity
[10]	Chaitra **and** Reddy**, 2023**	Deep CNN with pre-trained GoogLeNet + FLHHO optimization	Multiple Image Splicing dataset (MISD) and YOLO Object Detection dataset	Improved generalization; combines pre-trained networks with optimization	Heavy dependence on labeled data; computationally intensive
[11]	Khalil **et al., 2023**	Transfer learning on pre-trained models (MobileNetV2, etc.) for multi-forgery	CASIA 2.0 dataset	High detection accuracy; fewer training parameters; fast training	Sensitive to compression artifacts; relies on the quality of input images
[12]	Chaitra **and** Reddy**, 2025**	TFRA-ShuffleNet: multi-feature extraction + YOLO V3	Multiple Image Splicing dataset (MISD) and YOLO Object Detection dataset	Robust multi-feature detection; high accuracy and TPR/TNR	Complex model; dependent on accurate object detection
[13]	Maashi **et al., 2023**	Transfer learning with pre-trained models for multi-forgery detection, a fine-tuned classifier.	MNIST dataset	Computationally efficient; effective detection of multiple forgery types	Sensitive to image compression and dataset quality
[14]	Bevinamarad **et al., 2024**	SWT + Hybrid Dilated Adaptive VGG16 (HDA-	CMFD, CoMoFD	Precise localization; deep feature extraction	Computationally costly; patch-based

		VGG16) + HTS-BESO			processing is slow
[15]	Vaishnavi **et al., 2023**	IDL-CMIFD: EfficientNet + Adam + CMBO + DWNN	MICC-F220, MICC-F2000, MICC-F600	Optimized deep feature extraction; robust classification	Complex due to multiple heuristic optimizations
[16]	Li **et al., 2024**	Graph representation learning + cascaded graph + cross-attention	USCISI CMFD dataset, the CASIA-CMFD dataset	Effective modeling of patch relationships improves detection in ambiguous regions.	High computational complexity; depends on accurate graph construction.
[17]	Chaube **et al., 2024**	ACO-enhanced Siamese Network	CMFD	Robust feature matching; scalable; high accuracy (~97.13%)	Computational overhead; slower inference due to patch-wise processing
[18]	Yadav **et al., 2025**	Encoder–Decoder network with LSTM + ResNet50 backbone	DEFACTO, CASIAv1	Precise tamper localization; multi-level feature integration	Moderate AUC on some datasets; high model complexity
[19]	Alfraihi **et al., 2025**	ECMVFD-FTLTDO: Fusion-based transfer learning (ResNet50, MobileNetV3, EfficientNetB7) + Elman RNN + TDO	GRIP, VTD	Robust feature fusion; optimized detection; high accuracy	Increased computational cost due to multi-model fusion and sequential optimization
[20]	Zhao et al., 2024	Lightweight image forgery detection via multi-model fusion	MICC-F220, MICC-F600	High detection accuracy (91.5% & 88.8%), low computational cost, lightweight for deployment	Focused on image-level detection; may not handle video forgery or complex tampering.
[21]	Prathibha & Tamizharasan, 2024	Lightweight Separable Convolutional Networks for video forgery detection (copy-move/object-based)	MICC-F2000, MICC-F220, Rewind video dataset	Real-time detection on edge devices, improved performance on video datasets	Mainly targets copy-move forgery; may not generalize to other video manipulations.

In the copy-move forgery detection domain, strategies of conventional DL and transfer learning have shown robust performance. For example, Prem Kumar et al. [9] integrated MobileNet with heuristic optimization, improving classification accuracy, but their multi-step optimization increases computational complexity. Chaitra and Reddy [10], [12] combined pre-trained networks with heuristic optimizers to enhance generalization, yet these methods heavily depend on labeled data and are computationally intensive. Khalil et al. [11] achieved high detection accuracy with transfer learning, but performance is sensitive to compression artifacts and image quality. Other recent approaches, such as graph-based learning [16] or multi-model fusion [19], improve robustness and feature representation but often incur high computational cost or rely on accurate input structures. Lightweight and real-time methods[20], [21] reduce computational burden but may not handle complex manipulations or generalize to video forgery.

To tackle these limitations, our framework leverages contrastive self-supervised pretraining to learn discriminative features from untagged data, significantly reducing dependence on annotated datasets. Furthermore, our active hybrid optimization strategy, integrating the global exploration Tasmanian Devil Optimizer (TDO) with AdamW fine-tuning, ensures faster convergence, local minima avoidance, and improved stability. Compared with previous studies that either rely solely on pre-trained networks, heuristic optimization, or lightweight models, our approach provides a balanced solution that enhances detection accuracy, generalization, and robustness under diverse post-processing scenarios. Overall, the proposed framework demonstrates considerable improvement in automated copy-move forgery detection and offers a scalable and reliable framework suitable for real-life applications.

2.1. Methodological Justification

The main drawbacks of earlier copy-move forgery detection techniques, such as their need on sizable annotated datasets, sensitivity to post-processing, and unstable convergence brought on by hyperparameter dependence, are specifically addressed by the suggested approach. The model becomes much more scalable by using contrastive self-supervised pretraining to learn invariant and discriminative forgery-related representations without requiring a large amount of labelled data. Additionally, the dynamic hybrid optimisation technique (TDO + AdamW) reduces the possibility of sub-optimal local minima and enhances training stability by offering a balanced combination of steady fine convergence and global exploration. By capturing complementary multi-scale cues, the multi-backbone feature fusion module improves localisation and classification performance. When combined, these elements create a cohesive design that outperforms single-backbone, single-optimizer, or fully supervised methods described in the literature in handling real-world copy-move forgeries.

3. Proposed Method

This section provides the proposed framework for copy-move forgery detection that is modeled for developing feature classification accuracy, representation, and strength under challenging post-processing situations. The framework combines transfer learning-based fine-tuning, contrastive self-supervised pretraining, and an active hybrid model of optimization. The entire workflow is shown in Fig. 1.

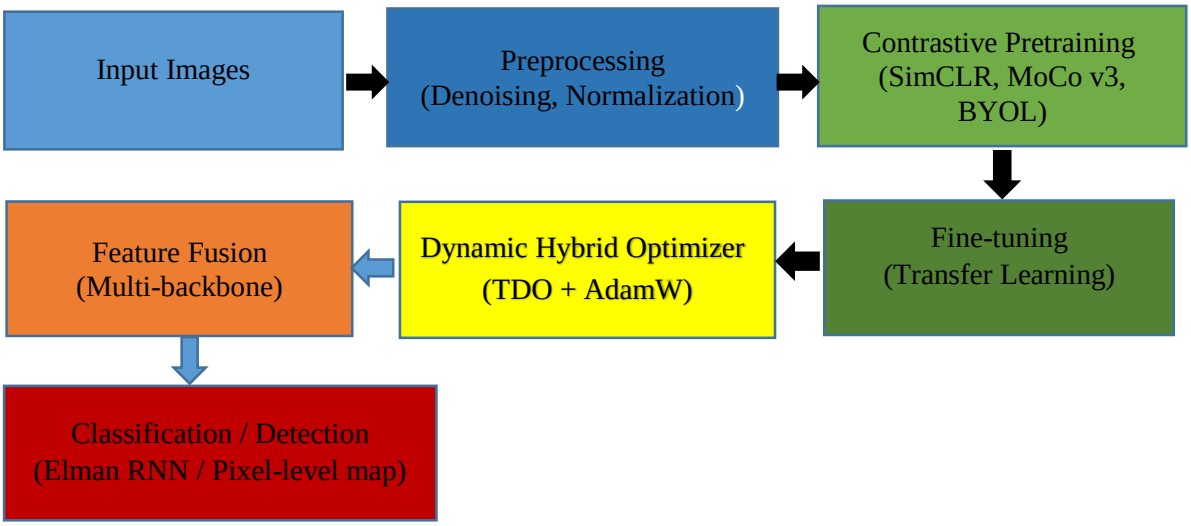

Figure 1: Proposed Copy-Move Forgery Detection Framework

The suggested training pipeline is depicted in the flowchart. To learn discriminative features from unlabelled data, input images are preprocessed and fed into contrastive pretraining (SimCLR, MoCo v3, BYOL). Multi-backbone feature fusion captures multi-level information, and the pretrained backbone is refined on copy-move datasets. For stable convergence, a dynamic hybrid optimiser (TDO + AdamW) modifies the parameters. Pixel-level tamper maps or image-level tags are produced by the model. Data flow is indicated by solid arrows, while optimizer-driven parameter adjustments are indicated by dashed arrows.

3.1. Image Preprocessing

For video content, the input is decomposed into single frames. Every frame undergoes denoising and normalization for developing related features when suppressing noise. A modified Wiener filter (MWF) could be used, and its formulation is given in Eq. (1) [22]:

$$\hat{f}(x,y) = \mu + \frac{\sigma^2 - v^2}{\sigma^2}[g(x,y) - \mu] \quad (1)$$

That μ and σ^2 are the local mean and variance, v^2 refers to the estimated noise variance, and $g(x,y)$ refers to the observed noisy image. Preprocessing guarantees that the unrelated variations do not interfere with feature extraction.

3.2. Contrastive Self-Supervised Pretraining

For considering the tagged forgery data, the framework develops contrastive self-supervised learning (like BYOL, SimCLR, MoCo v3) for pretraining a backbone network on untagged image data. The main stage contains:

- **Data Augmentation:** Make multiple views x_i, x_j for every picture by applying random color jittering, cropping, Gaussian blur, and flipping.
- **Feature Embedding:** The backbone CNN $f\theta(\cdot)$ maps every augmented picture to a latent embedding $z_i = f\theta(xi)$.
- **Contrastive Loss:** Decreasing the distance among similar picture embeddings (positive pairs) when increasing the distance among various pictures (negative pairs), so learning discriminative and strong representations [23], as shown in Eq. (2):

$$L_{ij} = -\log \frac{exp(sim(z_i, z_j)/T)}{\sum_{k=1}^{2N} \mathbf{1}_{[k \neq i]} exp(sim(z_i, z_k)/T)} \quad (2)$$

That is $sim(Z_i . Z_j) = \frac{Z_i Z_j}{\|Z_i\| \|Z_j\|}$ cosine similarity, z_i, z_j is feature embeddings of images i and j, T refers to the temperature hyperparameter, N is the batch size, and

$\mathbf{1}_{[k\neq i]}$ is the indicator function excluding the current positive pair from the denominator. This loss decreases the distance among similar picture embeddings when increasing the distance among various pictures, making discriminative and strong features with no need for annotated masks.

The present step lets the model take rich visual features with no dependence on annotated forgery masks, considerably decreasing the labeling demand.

3.3. Transfer Learning and Fine-Tuning

The model is fine-tuned on the aimed dataset of forgery after contrastive pretraining. Networks of backbone could be launched with the pre-trained weights, and a classifier is appended for the prediction of whether the area of the picture is real/forged. Fine-tuning adapts the learned features to the particular copy-move forgeries' features, such as rotation, post-processing artifacts, duplicated areas, and scaling.

3.4. Dynamic Hybrid Optimization

A Dynamic Hybrid Optimizer is used for fine-tuning the network parameters and also effectively training the classifier:

- **Tasmanian Devil Optimizer (TDO):** Used in the early training steps for performing parameter space global exploration and preventing local minima. Performs global exploration in early steps of training for preventing local minima [8], as shown in Eq. (3):

$$x_i^{t+1} = x_i^t + r.(x_{best} - x_i^t) + \Delta \qquad (3)$$

Where Δ is an adaptive exploration, x_i^t is the current position of solution i at iteration t, x_{best} is the current best solution, and r is a random factor concept.

- **AdamW Optimizer:** Used for fine-tuning, guaranteeing consistent convergence and quicker adaptation in later training steps. Used for fine-tuning for obtaining consistent gradient-driven updates [24], as shown in Eq. (4):

$$\theta_{t+1} = \theta_t - \eta \frac{m_t}{\sqrt{v_t}+\epsilon} + \lambda\theta_t \qquad (4)$$

That m_t and v_t are the first and second moment gradients' estimates, λ is the weight decay, η is the learning rate, and ϵ is a small constant to avoid division by zero.

The hybrid strategy integrates heuristic global search strengths (exploration) with gradient-driven optimization (exploitation), developing the two models' strengths and convergence pace.

The TDO and AdamW are combined in the suggested hybrid optimizer to take advantage of their complementing advantages. AdamW guarantees steady fine convergence and efficient gradient-driven updates, while TDO conducts early-stage global exploration of architectural knobs and hyperparameters to help the model avoid suboptimal local minima. Compared to employing AdamW alone or traditional optimizers, this combination produces more robust and consistent optimization by balancing exploration and exploitation throughout training.

3.5. Multi-Backbone Feature Fusion

For getting the multi-scale and multi-level features, embedding' fusion from hybrid backbone networks (like ResNet50 + MobileNetV3 + EfficientNetB7) could be developed. The fused features present complementary information, developing the capability of the model for differentiating between tampered and real areas. For getting multi-level and multi-scale info, embeddings from hybrid backbones (EfficientNetB7, ResNet50, MobileNetV3) are fused, as shown in Eq. (5):

$$Z_{fused} = Concat(Z_{ResNet50}, Z_{MobileNetV3}, Z_{EfficientNetB7}) \qquad (5)$$

Fused features present complementary information, developing discrimination among tempered and real areas.

3.6. Classification and Detection

At last, a classifier, like CNN-based patch classifiers, Elman RNN, and fully connected layers, predicts the tampered areas. Outcome could be a binary tag for the whole picture/pixel-level tamper map, based on the app. The fused embeddings are transferred to a classifier (CNN-based patch classifier, Elman RNN/fully connected layers) for making pixel-level tamper maps/image-level tags, as shown in Eq. (6):

$$\hat{y} = \sigma(W\, z_{fused} + b) \quad (6)$$

That σ refers to the softmax/sigmoid function, W and b are learnable parameters.

3.7. Post-Processing

For refined localization, optional post-processing stages such as connected component analysis, morphological filtering, and thresholding could be used to decrease false positives and improve the visual quality of the diagnosed forgery areas.

4. Experimental Setup / Implementation Details

4.1. Dataset description

The presented architecture is assessed on a dataset of MICC-F220 [25], a broadly applied benchmark for copy-move forgery diagnosis. The present set of data includes 220 images, equally shared among tampered and real samples. The tampered images are made through duplicating and positioning areas again in a similar picture, simulating real-life scenarios of forgery. MICC-F220 is especially appropriate to assess the two forgery detection models' localization and diagnosis abilities.

Whole pictures are preprocessed before training to guarantee uniformity and develop feature extraction. Every picture is sized to 224×224 pixels again for matching the backbone networks' input needs. Values of pixels are normalized to the range [0, 1], and data augmentation methods like rotation, color jittering, and random horizontal flipping are used for enhancing variability and developing generalization. Such a stage of preprocessing guarantees that the model concentrates on meaningful features rather than unrelated variations/noise.

The dataset is divided into sets of testing, training, and validation with a 70-15-15 ratio. Particularly, 33 pictures are applied for testing, 33 pictures are applied for validation, and 154 pictures are applied for training. This division presents a balanced architecture to learn while keeping the unbiased model's performance assessment.

For feature extraction, the architecture leverages pretrained backbone networks, such as EfficientNetB7, ResNet50, and MobileNetV3. The fully connected classifier with a sigmoid activation is appended to every backbone for predicting whether the picture area is tampered/real. Binary cross-entropy loss is developed as the objective. The Dynamic Hybrid Optimization approach is adopted, in the Tasmanian Devil Optimizer (TDO) is applied in early steps for global parameter space exploration, pursued by AdamW to fine-tune for guaranteeing quick adaptation and consistent convergence. The model is trained for 100 epochs with a batch size of 32 and a basic learning rate of 0.001 that is periodically decayed for refining training.

Whole tests are performed on NVIDIA GeForce RTX 3080 GPU with 10GB of VRAM. The architecture is performed applying PyTorch 1.10.0 with CUDA 11.3 on Ubuntu 20.04, presenting enough computational resources for effective assessment as well as training. This setup guarantees that the presented strategy could efficiently learn strong representations for copy-move forgery diagnosis on the dataset of MICC-F220.

Fig. 2 demonstrates a subset of sample pictures from the dataset of MICC-F220 applied here. The top row illustrates main pictures as taken, while the bottom row shows the related pictures after preprocessing, which includes conversion, normalization, and resizing to a tensor format appropriate for a DL model. The present stage of preprocessing guarantees

unique dimensions of input, decreases unrelated variations, and develops the capability of the model for learning. Meaningful and discriminative attributes for copy–move forgery diagnosis.

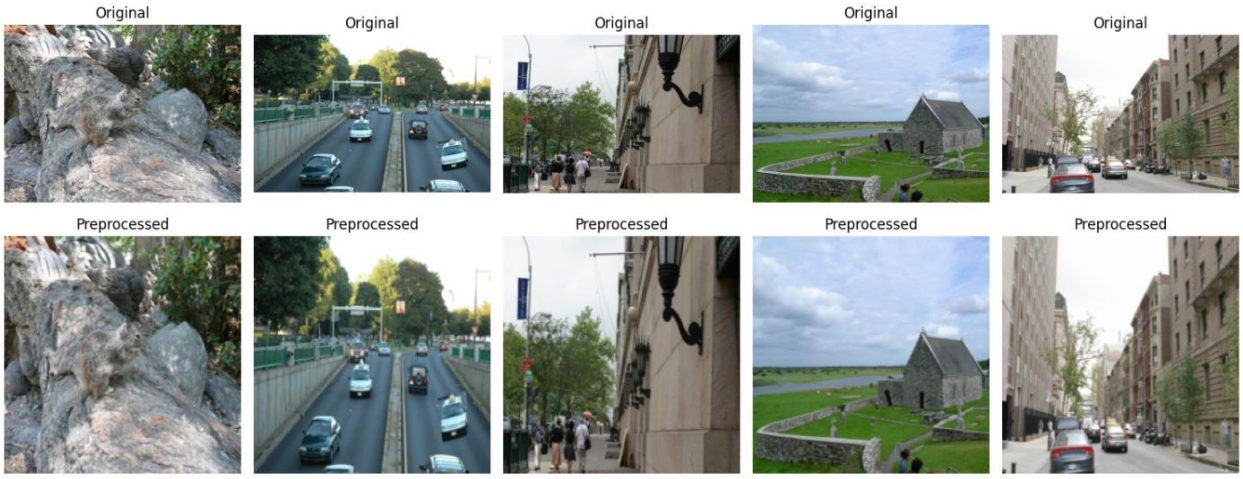

Figure 2: Sample images from the MICC-F220 dataset: original images (top row) and preprocessed images (bottom row) for model input.

4.2. Evaluation Metrics

The offered copy-move forgery diagnosis framework performance is assessed quantitatively by applying some standard metrics.

Accuracy (ACC) scales the whole correctly classified samples ratio is described as Eq. (7):

$$Accuacy = \frac{TP+TN}{TP+TN+FP+FN} \quad (7)$$

The TN (true negatives) shows accurately recognized real areas, FN (false negatives) shows missed tampered areas, TP (true positives) shows accurately diagnosed forged areas, and FP (false positives) denotes inappropriately diagnosed tampered areas.

Precision (P) scales the predicted tampered areas' correctness, as Eq. (8):

$$Precision = \frac{TP}{TP+FP} \quad (8)$$

Recall (R) quantifies the completeness of the diagnosed tampered areas, as Eq. (9):

$$Recall = \frac{TP}{TP+FN} \quad (9)$$

The F1-score that balances Recall as well as Precision, is described as Eq. (10):

$$\text{F1} - \text{score} = 2.\frac{\text{Precision} \,.\, \text{Recall}}{\text{Precision+Recall}} \quad (10)$$

These metrics are calculated under various scenarios of post-processing, such as noise addition, scaling, and rotation, to guarantee the offered strength and generalization of the model to real-life forgery situations. Collectively, these scales present a general assessment of the two-framework spatial localization performance and diagnosis accuracy on the dataset of MICC-F220.

4.3. Results and Discussion

The dynamic hybrid optimization framework, as well as offering contrastive pretraining, is assessed on a dataset of MICC-F220 for copy-move forgery diagnosis. The tests are modeled for evaluating the two tampered areas' classification accuracy and pixel-level localization. The assessment concentrates on analyzing contrastive self-supervised pretraining efficiency, the dynamic hybrid optimizer contribution, and potential developments obtained from feature fusion over hybrid networks of backbones.

4.3.1. Classification Performance

The offered framework classification performance was assessed by applying standard metrics, such as F1-score,

Accuracy, Precision, and Recall. As illustrated in Table 1, the offered **Contrastive Pretraining + Fine-tuning** strategy performed better than conventional models trained from scratch, like EfficientNetB7, ResNet50, and MobileNetV3. When models of baseline obtained accuracies in the range of 86–89%, the offered architecture got an F1-score of 0.9231 and accuracy of 93.94% with great precision. The outcomes show that contrastive pretraining efficiently makes the model able to learn discriminative and strong features from untagged data, developing generalization to unobserved pictures and reliably diagnosing copy-move forgeries.

Table 1: Performance comparison between the proposed model and conventional deep learning baselines

Model	Accuracy	Precision	Recall	F1-score
ResNet50 (from scratch)	0.8800	0.9000	0.8000	0.8470
MobileNetV3 (from scratch)	0.8650	0.8800	0.7800	0.8270
EfficientNetB7 (from scratch)	0.8950	0.9100	0.8100	0.8570
Proposed (Contrastive Pretraining + Fine-tuning)	0.9394	1.0	0.8571	0.9231

4.3.2. Confusion matrices

Metrics of confusion for the two sets of training and testing show that the offered model obtains a high performance in classification. For the set of training, out of 74 negative samples, all were accurately grouped, while 75 out of 80 positive samples were accurately recognized, with just five misclassified as negative, causing an overall accuracy of nearly 96.8%. For the set of testing, all 19 negative samples were accurately predicted, 12 out of 14 positive samples were accurately grouped, showing the accuracy of around 93.9%. Such outcomes show that the model not only performs well on training but also efficiently generalizes to unobserved data, keeping robust ability in differentiating between negative and positive levels. The small number in the two sets bolsters the offered strategy strength and reliability.

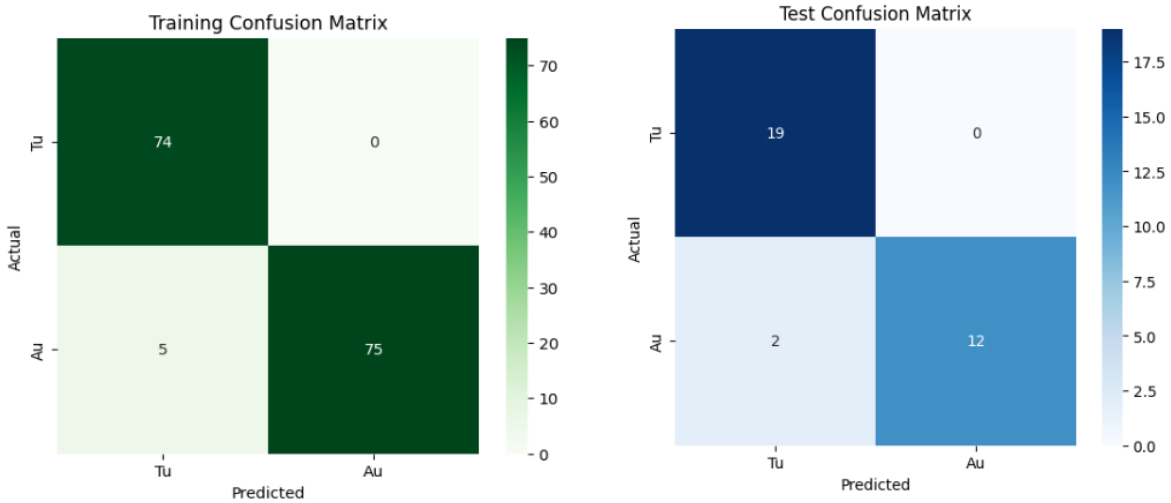

Figure 3: Confusion matrices of the proposed model. (a) Training set results. (b) Testing set results

4.3.3. Discussion

The experimental analysis validates that the offered technique obtains considerable developments over baseline models in the two-copy-move forgery classification and localization. Contrastive pretraining makes efficient feature learning from restricted tagged data when the dynamic hybrid optimizer develops model stability and convergence speed. Feature fusion over hybrid backbones later enriches the ability of representation, letting the subtle forgeries' diagnosis. Future studies can develop the architecture for broader and more varied sets of data, such as video forgeries, and also explore more developed self-supervised learning techniques.

Despite achieving excellent precision, the suggested model's recall is marginally lower (85.7%), suggesting that some genuine copy-move areas might be overlooked. This often happens when copies overlap or when minor forgeries have backdrop textures that are comparable to the copied area. Furthermore, it can be difficult to identify small or partially overlapping tampered portions in photos with many forgeries. These drawbacks point to possible directions for further research, such as better post-processing techniques and multi-region detection methodologies.

4.3.4. Comparison of the proposed method and other methods

Comparison of performance among Zhao et al. [20], Prathibha and Tamizharasan [21], as well as the offered technique, obviously shows our new architecture benefits. Zhao et al.'s multi-model fusion strategy obtains the recall of 90.9%, F1-score of 90.6%, accuracy of 91.5%, and precision of 90.3% while the light Separable Convolutional Network of Prathibha and Tamizharasan's takes lower performance with the recall of 81%, accuracy of 86.74%, F1-score of 85%, and precision of 90%. On the contrary, the offered method obtains greater outcomes with the recall of 85.71%, an F1-score of 92.31%, accuracy of 93.94%, and great precision of 100%. This development is broadly because of 3 main innovations: (1) a dynamic hybrid optimization model which integrates TDO for global exploration with AdamW for consistent fine-tuning, guaranteeing the two convergence pace and strength; (2) a contrastive self-supervised pretraining approach tailored for copy–move forgery diagnosis, making the efficient feature learning from untagged data able; (3) a strong architecture model which considerably develops generalization and diagnosis accuracy under challenging post-processing situations. Entirely, such innovations decrease false positives while keeping high diagnosis rates, causing better overall performance in comparison with new techniques

Table 2: Comparative evaluation showing the superior performance of the proposed method over existing methods.

Method	Accuracy (%)	Precision	Recall	F1-score
Zhao et al. [20]	91.5	90.3	90.9	90.6
Prathibha and Tamizharasan.[21]	86.74	90	81	85
Proposed method	93.94	100	85.71	92.31

5. Conclusion and Future Work

This paper provides a new framework for copy-move forgery detection that leverages contrastive self-supervised pretraining and a dynamic hybrid optimization approach. Through developing the methods of contrastive learning like BYOL, SimCLR, and MoCo v3, the framework learns strong and discriminative visual representations from untagged data to mitigate restricted annotated forgery datasets. Tasmanian Devil Optimizer (TDO) for global exploration and AdamW combination to fine-tune guarantees the two quick, convergent, and consistent optimizations, improving the overall performance of diagnosis. Also, the fusion of hybrid backbone networks makes multi-scale feature extraction possible, developing the ability of the model for diagnosing subtle forgeries and appropriately localizing tampered areas. Experimental assessments on the MICC-F220 dataset show that the offered strategy obtains greater performance in the two metrics of classification and localization in comparison with conventional DL and transfer learning baselines. The architecture demonstrates strength under different post-processing situations, such as noise addition, scaling, and rotation, bolstering its capability for real-life forgery detection apps. Future papers would concentrate on developing the offered architecture for broader and more diverse sets of data, such as high-resolution video content and pictures. Developed a self-supervised learning technique and a transformer-driven framework that can be later explored for developing feature representation and generalization. Also, incorporating temporal consistency for video forgery detection and exploring real-life development on resource-limited devices are promising directions for practical apps.

References

[1] U. Samariya, S. D. Kamble, S. Singh, and R. K. Sonker, "A survey on copy-move image forgery detection based on deep-learning techniques," *Multimed. Tools Appl.*, vol. 84, no. 26, pp. 30603–30662, Aug. 2025, doi: 10.1007/s11042-024-20323-7.

[2] M. Verma and D. Singh, "Survey on image copy-move forgery detection," *Multimed. Tools Appl.*, vol. 83, no. 8, pp. 23761–23797, Mar. 2024, doi: 10.1007/s11042-023-16455-x.

[3] C. Li and Y. Wo, "Towards generalized face forgery detection with domain-robust representation learning," *Digit. Signal Process.*, vol. 156, p. 104792, 2025.

[4] M. Q. Ibrahim, N. K. Hussein, D. Guinovart, and M. Qaraad, "Optimizing Convolutional Neural Networks: A Comprehensive Review of Hyperparameter Tuning Through Metaheuristic Algorithms," *Arch. Comput. Methods Eng.*, May 2025, doi: 10.1007/s11831-025-10292-x.

[5] Z. Yu Diong, W. Y. Lim, and C. P. Goh, "Self-Supervised Learning in Medical Diagnostics: An Examination of SimCLR and BYOL in Image Classification," in *2024 3rd International Conference on Digital Transformation and Applications (ICDXA)*, Jan. 2024, pp. 210–214. doi: 10.1109/ICDXA61007.2024.10470922.

[6] X. Chen, S. Xie, and K. He, "An Empirical Study of Training Self-Supervised Vision Transformers," Aug. 16, 2021, *arXiv*: arXiv:2104.02057. doi: 10.48550/arXiv.2104.02057.

[7] T. Chen, S. Kornblith, M. Norouzi, and G. Hinton, "A Simple Framework for Contrastive Learning of Visual Representations," July 01, 2020, *arXiv*: arXiv:2002.05709. doi: 10.48550/arXiv.2002.05709.

[8] R. M. Rizk-Allah, R. A. El-Sehiemy, and M. I. Abdelwanis, "Improved Tasmanian devil optimization algorithm for parameter identification of electric transformers," *Neural Comput. Appl.*, vol. 36, pp. 3141–3166, Nov. 2023, doi: 10.1007/s00521-023-09240-2.

[9] C. D. Prem Kumar and S. Saravana Sundaram, "Metaheuristics with Optimal Deep Transfer Learning Based Copy-Move Forgery Detection Technique.," *Intell. Autom. Soft Comput.*, vol. 35, no. 1, 2023, Accessed: Dec. 14, 2025. [Online]. Available: https://search.ebscohost.com/login.aspx?direct=true&profile=ehost&scope=site&authtype=crawler&jrnl=10798587&AN=158048083&h=uHH6rOYznQ1AFIR1ARFOXWmI5cbmsHPbeuQNDWKSM2%2FNUhi%2B8SYNxp%2BAh3zEljYg0FQMPFBIxMaK%2B14sgbHwrg%3D%3D&crl=c

[10] C. B and P. V. Bhaskar Reddy, "An approach for copy-move image multiple forgery detection based on an optimized pre-trained deep learning model," *Knowl.-Based Syst.*, vol. 269, p. 110508, June 2023, doi: 10.1016/j.knosys.2023.110508.

[11] A. H. Khalil, A. Z. Ghalwash, H. A.-G. Elsayed, G. I. Salama, and H. A. Ghalwash, "Enhancing Digital Image Forgery Detection Using Transfer Learning," *IEEE Access*, vol. 11, pp. 91583–91594, 2023, doi: 10.1109/ACCESS.2023.3307357.

[12] B. Chaitra and P. V. Bhaskar Reddy, "Copy-Move Image Multiple Forgery Detection Based on Transit Flow Regime Algorithm-Enabled ShuffleNet," *Int. J. Image Graph.*, p. 2750017, Jan. 2025, doi: 10.1142/S0219467827500173.

[13] M. Maashi *et al.*, "Modeling of reptile search algorithm with deep learning approach for copy move image forgery detection," *IEEE Access*, vol. 11, pp.

87297–87304, Aug. 2023, doi: 10.1109/ACCESS.2023.3304237.

[14] P. Bevinamarad, P. Unki, and P. Nidagundi, "Copy-Move Forgery Detection and Localization Framework for Images Using Stationary Wavelet Transform and Hybrid Dilated Adaptive VGG 16 with Optimization Strategy," *Int. J. Image Graph. Signal Process. IJIGSP*, vol. 16, no. 1, pp. 38–60, 2024, doi: 10.5815/IJIGSP.2024.01.04.

[15] D. Vaishnavi and G. N. Balaji, "Modeling of intelligent hyperparameter tuned deep learning based copy move image forgery detection technique," *J. Intell. Fuzzy Syst.*, vol. 45, no. 6, pp. 10267–10280, Dec. 2023, doi: 10.3233/JIFS-230291.

[16] Y. Li, L. Ye, H. Cao, W. Wang, and Z. Hua, "Cascaded Adaptive Graph Representation Learning for Image Copy-Move Forgery Detection," *ACM Trans. Multimed. Comput. Commun. Appl.*, vol. 21, no. 2, pp. 1–24, Feb. 2025, doi: 10.1145/3669905.

[17] A. Chaube, "ACO-Enhanced Siamese Networks for Robust Feature Matching in Copy-Move Image Forgery Detection," in *2024 International Conference on Artificial Intelligence and Quantum Computation-Based Sensor Application (ICAIQSA)*, IEEE, 2024, pp. 1–6. doi: 10.1109/ICAIQSA64000.2024.10882433.

[18] K. D. K. Yadav, I. Kavati, and R. Cheruku, "CCLHF-Net: Constrained Convolution Layer and Hybrid Features-Based Skip Connection Network for Image Forgery Detection," *Arab. J. Sci. Eng.*, vol. 50, no. 2, pp. 825–834, Jan. 2025, doi: 10.1007/s13369-024-09039-w.

[19] H. Alfraihi *et al.*, "A multi-model feature fusion based transfer learning with heuristic search for copy-move video forgery detection," *Sci. Rep.*, vol. 15, no. 1, p. 4738, 2025, doi: 10.1038/s41598-025-88592-2.

[20] Z. Zhao, S. Meng, and C. Wang, "Image Forgery Detection Method Based on Lightweight Fusion Models," in *Proceedings of the 2024 8th International Conference on Big Data and Internet of Things*, Macau China: ACM, Sept. 2024, pp. 53–57. doi: 10.1145/3697355.3697364.

[21] P. G. Prathibha and P. S. Tamizharasan, "A LightWeight IntraFrame Forgery Detection Model for Surveillance Videos," in *2024 Advances in Science and Engineering Technology International Conferences (ASET)*, IEEE, 2024, pp. 1–6. doi: https://doi.org/10.1109/ASET60340.2024.10708758.

[22] N. A. M. Abir, N. B. A. Warif, and N. Zainal, "An automatic enhanced filters with frequency-based copy-move forgery detection for social media images," *Multimed. Tools Appl.*, vol. 83, no. 1, pp. 1513–1538, Jan. 2024, doi: 10.1007/s11042-023-15506-7.

[23] M. Liu, J. Wang, X. Qian, and H. Li, "Audio-visual temporal forgery detection using embedding-level fusion and multi-dimensional contrastive loss," *IEEE Trans. Circuits Syst. Video Technol.*, vol. 34, no. 8, pp. 6937–6948, 2023.

[24] N. H. Celebi, T.-L. Hsu, and Q. Liu, "A comparison study to detect seam carving forgery in JPEG images with deep learning models," *J. Surveill. Secur. Saf.*, vol. 3, no. 3, pp. 88–100, 2022.

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